Response Adaptive Randomization Using Biomarkers with Exponentially Decreasing Probability Sequence

T. Palanisamy, J. Ravichandran* and Midhuna Ramesh

Department of Mathematics, Amrita School of Physical Sciences, Coimbatore Amrita Vishwa Vidyapeetham, India

Abstract: In this article, it is proposed to study the application of Response Adaptive Randomization (RAR) design in clinical trials. The approach involves the prediction of treatment outcomes based on the biomarker of patients using a regression model. The focus is on rare diseases to efficiently allot the patients among various treatments so as to ensure not only the clinical rights but also the maximum possible benefits to the patients even when they are in clinical trials. Initially, the method uses conventional equal randomization to understand how well every treatment works in patients and this initial duration is known as burn-in period. The proposed work allocates patients to treatments by using an exponentially decreasing probability sequence instead of the existing linearly decreasing sequence to have higher allocation probability to the efficient treatment. In the case of rare disease, it is observed from simulation study that the use of exponentially decreasing probability sequence in RAR design increases the benefit to the patients in the clinical trials when compared to the existing method that uses linearly decreasing sequence. The study also investigates the performance of the proposed RAR design when used with different regression methods under various scenarios. The performance of the proposed design is measured by the proportion of patients assigned to the best treatment in addition to Type I error and power. From the impressive results, it is suggested that the proposed RAR design can be implemented practically in clinical trials of rare diseases without any apprehension.

Keywords: Biomarker, Clinical trials, Exponentially decreasing probability sequence, Rare diseases, Response adaptive randomization.

1. INTRODUCTION

Clinical trials play a significant role in modern healthcare system that always tries to provide right treatment for various newly rising ailments and in turn will help to improve upon the existing healthcare provisions. Applications of statistical methods in such clinical trials provide valuable decisions, particularly in the allocation of patients to different treatments so as to assure maximum advantage to every individual involved. Statistical methods and their applications keep evolving and are effectively used to analyze clinical data to draw evidence-based conclusions in the healthcare sector [1- 3]. It may be noted that while randomized controlled trials (RCT) with equal probability of allotment to each treatment arm is the approach that is most often used, the RAR design which emphasizes on the benefit to the trial population [4] is, of course, more suitable in case of rare diseases. In fact, the main focus of RAR is to ensure that the patients within a trial are able to avail the best treatment with high probability on the basis of accumulating data [5] which is very slow for rare diseases [6]. Some important studies on the development of optimal RAR algorithms include, the one by Jennison and Turnbull [7] who discussed group sequential comparisons of two treatments where treatment assignments can be based on previously observed responses. Then, Rosenberger *et al*. [8] suggested a sequential design that asymptotically achieves the optimal allocation and reveals that the sequential procedure consistently results in fewer treatment failures in case of treatments with lower success rates. Hu *et al*. [9] considered a novel doubly adaptive biased coin design for making allocations to multiple treatments. Some of the drawbacks of these existing approaches include the low power of the statistical testing and not handling the time trend. However, these drawbacks cannot be generalized among the vast subclasses of RAR designs.

In spite of illustrated advantages, the use of RAR approach in clinical trials is found to be relatively low compared to the theoretical interest [10] except some notable implementation of RAR in clinical trials. It is noted that there are few improved RAR designs that use information obtained from previous patients to decide the treatment for the next before the completion of the trial. This strategy of RAR obviously is able to increase the number of successful outcomes in patients and is, of course, analogous to earning while learning [11]. It was observed from various research studies that the use of RAR design is more beneficial to patients even when the clinical practice of this design has an unbalanced sample sizes for treatments. However, this suggests that introducing a burn-in period of conventional equal randomization helps in

*Address correspondence to this author at the Department of Mathematics, Amrita School of Physical Sciences, Coimbatore Amrita Vishwa Vidyapeetham, India; E-mails: aishwar2@rediffmail.com and j_ravichandran@cb.amrita.edu

understanding how well every treatment works in patients [12]. Also, there are attempts to modify the assumptions followed in many RAR trials in which each patient will react the same way for each treatment.

Taking into consideration of these aspects, development of different models, such as Covariate Adjusted Response Adaptive (CARA) design, that focused on personalization of treatment integrating covariates to treatment responses have been tried. Also there are few more significant works in CARA designs, e.g., [13, 14], that utilized a modified Gittins index rule to solve this kind of problems to determine the allocation probabilities. Thall and Wathen [15] considered Bayesian design for a multicenter and randomized clinical trial of two chemotherapy regimens using covariates. Amidst such developments, the decision on treatments in the case of cancer patients and the use of respective medicines could be personalized [16,17] based on clinical biomarker tests considering them as covariates. Also, a study on the estimation of parameters related to the success rates in clinical trials [18] reveals the fact that the use of biomarkers has higher overall success probabilities than trials without biomarkers.

Jackson *et al*. [4] in their research study incorporated the biomarkers of patients with RAR trials for general diseases with an aim to adjust treatment allocation probabilities. Such allocations are done by using regression analysis relating biomarker value of an individual to the treatment outcome. These studies explore the application of different regression techniques to predict the optimal treatment for subsequent patients. It may be noted that the studies based on RAR design and burn-in period are obviously due to the influence of randomized allocation with nonparametric estimation for a multi-armed bandit problem with covariates [19]. Park [20] made an elaborate statistical analysis to study the challenges and opportunities in biomarker-driven RAR trials. Jackson *et al*. [4] has conducted statistical analysis for such scenarios and highlighted impressive advantages of using biomarkers in RAR design for the patients involved in clinical trial of rare disease. Motivated by such advantages, in this work, the authors have proposed to develop a RAR design in clinical trials by using exponentially decreasing probability sequence of allocation of patients to different treatment arms instead of the existing linearly decreasing sequence.

The remainder of the paper is organized as follows: Section 2 presents the proposed work in which the RAR design is developed using exponentially decreasing probability sequence (Section 2.1) of allocation of patients to different treatments. A complete algorithm for the implementation of the proposed biomarker adjusted RAR procedure is in Section 2.2. In Section 3, a simulation study is conducted to evaluate the proposed RAR design using biomarkers and various regression techniques by making comparison among different techniques considered. In addition, comparison with RAR design using linearly decreasing sequence and equal fixed randomization (FR) method are also made. The simulation results of various performance measures considered are presented in detail in Section 3. Finally, in Section 4, the outcomes of the study are discussed and conclusions are drawn.

2. RAR DESIGN WITH EXPONENTIALLY DECREASING PROBABILITY SEQUENCE

As presented in the previous section, it is observed from the literature that there has been extensive research conducted by various authors in RAR design of clinical trials where biomarkers are used as covariates to improve treatment assignments, particularly Jackson *et al*. [4]. In tune with the work of Jackson *et al*. [4], in our work it is proposed to study the relationship between the biomarker and the outcome of the treatment using regression models for every possible treatment under consideration. It may be noted that unlike Jackson *et al*. [4], the allocation of patients to treatments is based on the exponentially decreasing probability in the proposed RAR design. Accordingly, our work starts by considering the biomarker and the treatment outcome of each patient from initial burn-in period to fit a regression models for every treatment of interest. The purpose of this step is to identify the suitable treatment for the patients after burn-in period by estimating the response of the model depending on the biomarker of the individual. This regression approach is applied further to the data from all the previous patients to estimate the best treatment for the consecutive patients using their biomarkers.

In fact, the proposed work considers the following regression models: (i) Nearest Neighborhood (NN) model which is a supervised ML algorithm that incorporates the information from a certain number of neighborhood data, (ii) Polynomial Regression (PR) model which considers a 3^{rd} degree polynomial for whole of the data after four patients. For cases where there are less than four patients, polynomials of degree one less than the number of patients are considered, (iii) Spline Regression (SR) model is used for fitting a

smooth curve by dividing independent variables into segments in which for each segment a cubic polynomial is fitted, (iv) Gaussian Process (GP) model which is a nonparametric supervised method that uses stochastic process and (v) Random Forest (RF) model which is a robust and accurate ensemble ML algorithm.

It may be noted that, while these regression models are used for predicting the treatment efficiency, the allocations are made using the linear probability sequence. Assuming that use of exponentially decreasing sequence of probability will result in better allocation, in this work along with these regression models we have applied the exponentially decreasing sequence instead of linear probability sequence. Such an assumption is made due to the fact that the increase in probability for allocating the next patient to the best treatment using exponential sequence is naturally higher when compared to linearly decreasing sequence. We have performed simulations and the results are compared under six different scenarios reflecting real time situations taking into account the relationship between the biomarker and the outcome of the treatment. As expected, we noticed that the use of exponentially decreasing sequence outperforms the linearly decreasing sequence.

Let us consider a clinical trial situation in which there are $K \geq 2$ number of treatment possibilities and a total of N patients before burn-in period and n_t patients after burn-in period. The task is to administer an efficient treatment to every patient in the clinical trial. It may be noted that the proposed design, in specific, takes into consideration the case of clinical trials with rare diseases where the outcome of the treatment for previous patient is available before the entry of every next patient. It is assumed that the biomarker value x_n of n^{th} patient is assumed to be available for all the patients. In the trial, patients are divided and allotted equally to all the treatments till a predetermined burn-in period. Let us assume that $L = N/K$ patients are allotted for each treatment during the burn-in period. Here, the specific treatment for the first patient after burn-in period is decided based on the outcomes $Y(0,k)$ of the regression models $Y(0,k) = f_k(x) + \epsilon$ fitted for k^{th} treatment using the relationship between the biomarker values and the treatment outcomes of all those patients completed trial during the burn-in period. It is felt essential to stress that the regression model $Y(i,k)$ is different from the regression model $Y(j,k)$, though both the models are for the k^{th} treatment. They are regression models fitted considering all the patients in burn-in period along with first i and j patients after burn-in period respectively and hence they are not necessarily the same. It is important to note that the outcome of the regression model may be binary, integer valued or continuous value depending on the type of disease. At the end of the burn-in period, the regression models are updated for every treatment after the completion of the trial for every consecutive patient. It is essential that the allocation must be made to balance the tendency of trusting the currently most promising treatment with the exploration of all the available new treatments [19]. Since this kind of allocation can be accomplished only by means of some randomization procedure, we have considered allotting the patients to the more successful treatment with a probability $1 - P_n$ given by an exponentially decreasing sequence P_n.

2.1. Exponentially Decreasing Probability Sequence

The exponentially decreasing probability sequence is employed to allocate the patients probabilistically to the treatment which is estimated to be the best from the available small data. It is obvious that more the patients, better the efficiency of prediction of treatment outcome. Due to this fact and the availability of little amount of data right after the burn-in phase, there is no discernible confidence in the estimated best treatment using the regression model linking the biomarker by the suggested design. However, when the number of patients increases, its apparent that the confidence on the best treatment raises. Thus, allotting the patient to the estimated best treatment using $1 - P_n$, where P_n, an exponentially decreasing probability sequence, is more justifiable rather than allotting them directly to the estimated best treatment or by using linearly decreasing probability sequence.

In fact, during the burn-in period the first L patients are randomized to the two treatments in a $1:1$ ratio. From the $(N+1)$ st patient onwards, the n^{th} patient after the burn-in period will be assigned to their estimated best treatment with probability $1 - P_n$. Let us assume that there will be totally n_t patients in the clinical trial after burn-in period. Due to the fact that the patients are allotted in equal randomization during the burn-in period, the probability for those patients to be allotted to any one of the K treatments is $1/K$. We denote this consistently with the notation $1 - P_n$ as $1 - P_0 = \frac{1}{K}$. Let us assume, $P_1 = P_0 e^{-\tau}$ and $P_{i+1} = P_i e^{-\tau}$. Therefore, $P_2 = P_1 e^{-\tau}$, which is otherwise $P_2 = P_0 e^{-2\tau}$. This leads to the equation $P_n = P_0 e^{-n\tau}$ which is $P_n = \frac{K-1}{K} e^{-n\tau}$ for any n. Here, τ is a parameter to be estimated using P_0 and P_n. It is to be noted that P_n denotes the probability for n^{th}patient to be allotted to

Table 1: Exponentially Decreasing Sequence of Probabilities for 2 Treatments and 10 Patients

n	0	1	2	3	4	5	6	7	8	9	10
P_n	0.5	0.426	0.362	0.309	0.263	0.224	0.19	0.162	0.138	0.117	0.1
$1-P_n$	0.5	0.574	0.638	0.691	0.737	0.776	0.81	0.838	0.862	0.883	0.9

any other treatment but not the estimated best treatment. Assuming that there will always be at least a 0.1 probability of allocating even the last patient in the clinical trial to any other treatment, we will have $P_{n_t} = 0.1 = \frac{K-1}{K} e^{-n_t \tau}$. It can easily be computed as $\tau = -\frac{1}{n_t} \ln\left(\frac{K}{10K-10}\right)$. Thus, n^{th} patient will be allotted to the estimated best treatment with the probability given by the exponentially decreasing sequence $1 - \frac{K-1}{K} e^{-\frac{n}{n_t} \ln\left(\frac{K}{10K-10}\right)}$ for any n. Obviously, this probability sequence continues to increase from $\frac{1}{K}$ for first patient after burn-in period to 0.9 for the last n_t th patient in the trail. For example, in the case of two treatments and 10 patients in the trail probability sequence, the probabilities are computed as given in Table **1**.

2.2. Development of the Proposed Algorithm

In this section we develop a step-by-step algorithm for the proposed work which can dynamically distribute treatments based on the data of the previous patients to improve efficiency of trials under consideration. First we summarize important notations used in the algorithm and their descriptions as given in Table **2** follows:

The complete step-by-step algorithm for the proposed biomarker adjusted RAR procedure is given as follows:

Step1: Initial burn-in period: Consider K treatments and allot L patients to each of the K treatments so that there are $L \times K = N$ number of patients. Here, the patients are allocated to the K treatments by equal randomization.

Step 2: Based on the initial burn-in period data, fit a regression model to the biomarker (x) and the treatment outcome (y) for each treatment separately

Step 3: Consider the biomarker of the next patient.

Step 4: Use the regression models obtained to estimate the outcomes for each of the K treatments.

Step 5: Arrange the outcomes in an ascending order and assign the probability to the treatment, say, b, corresponding to the best estimated outcome using an exponentially decreasing probability sequence as detailed in Section 2.1.

Step 6: The observed outcome of the patient given in Step 3, will be used to update the regression model for the treatment b.

Steps 3-5 are repeated for the next patients consecutively until the last patient $N + n_t$ in the clinical trial is considered.

It is interesting to note that the proposed design with the exponential sequence of probability can maximize the efficiency even for the patients in the clinical trial.

3. SIMULATION STUDY

In this section, a simulation study is presented to evaluate the proposed RAR design using biomarkers with exponentially decreasing probability sequence. It

Table 2: Notations and Their Descriptions

Notation	Description
K	Number of possible treatments
N	Number of patients available for treatment during burn-in period
L	Number of patients considered per treatment ($L = N/K$)
n_t	Total Number of patients in the trial after burn-in period
x	Biomarker variable (independent variable)
y	Outcome of the treatment (dependent variable)
$x(i,k)$	The biomarker of i^{th} patient considered for treatment k of K, $k = 1, 2, \dots, K$ and $i = 1,2, \dots, L, L+1, \dots, L+n_t$
$y(i,k)$	The outcome of i^{th} patient considered for treatment k of K, $k = 1, 2, \dots, K$ and $i = 1,2, \dots, L, L+1, \dots, L+n_t$
$Y(n,k)$	Regression model relating x and y fitted by considering first n patients after burn-in period along with the L patients in the burn-in period for treatment k , $k = 1, 2, \dots, K$ and $n = 1,2, \dots, n_t$
P_n	The probability associated with n^{th} patient.

is also compared with the RAR design using biomarker with linearly decreasing probability sequence and with the equal fixed randomization (FR) design. To make our study simpler, we consider two treatment trials and only one biomarker say x for all our simulation. The biomarker values are assumed to follow continuous uniform distribution in $[-100,100]$. The purpose of the proposed work is to improve the efficiency of the patient allocation in clinical trials of a rare disease using regression models relating the biomarker values to the trial outcomes. Accordingly, we consider the treatment outcome of a patient say y as a function of the patient's biomarker value x, added with an error term say, ϵ and is $Y = f(x) + \epsilon$. Therefore for two treatments, we have the regression models $Y(n,1) = f_1(x) + \epsilon_1$ and $Y(n,2) = f_2(x) + \epsilon_2$ for the $(n+1)th$ patient. Clearly, the functions $f_1(x)$ and $f_2(x)$ are different and they represent two different treatments 1 and 2 respectively. We have also considered the comparison of the performances of various regression techniques such as nearest neighbor (NN), polynomial regression (PR), spline regression (SR), Gaussian process (GP) and random forests (RF) used in our simulation study.

The evaluation study is assumed to involve 80 patients to match with the real time clinical trials of a rare disease. Number of patients of different sizes are also considered in the simulation without deviating from the clinical trial requirements of a rare disease. We have discussed the effects of all such cases in the section dedicated for discussion of results towards the end. The simulation is done using 5000 number of iterations except for random forest which involves complex computations. Here, the first 10 patients are considered for the burn-in period with equal randomized allotments for both the treatments of considerations. The allotment after burn-in period starting with eleventh patient is made to the estimated best treatment by the regression method but with a probability of $1-P_n$. This probability P_n is an exponentially decreasing sequence starting with $P_{10} = 0.5$ for the 10^{th} patient and ending with $P_{n_t} = 0.1$ for the ${n_t}^{th}$ patient.

In fact, we have considered six different scenarios for our simulations, mapping real-time situations that may relate the biomarker with treatment outcome of the patients. Scenario 1 is considered such that there is no influence of the biomarker value on the treatment outcomes and hence the functions become $f_1(x) = f_2(x) = 0$. Scenario 2 is about the prognostic markers [21] where the outcomes of both the treatments increase with a similar amount when the biomarker changes. They are suitably modelled by $f_1(x) = \frac{20}{e^{0.002x+1}} - 10$ and $f_2(x) = \frac{20}{e^{0.002x+1}} - 4$. Scenario 3 considers predictive markers [21] which improves the outcome of a treatment over the other for specific biomarker values. We model them by $f_1(x) = 0$ and $f_2(x) = \frac{20}{e^{0.02(x+8)+1}} - 10$. Scenario 4 deals with an investigation of prognostic and predictive markers in which the outcomes of both the treatments will improve but the rate of improvement will differ and is modelled by the functions $f_1(x) = \frac{20}{e^{0.02(x+5.2)+1}} - 10$ and $f_2(x) = \frac{20}{e^{0.011x+1}} - 10$. Scenario 5 studies the situation where both the treatments are predictive but the efficiency of one treatment will be better till a specific biomarker value after which the other will become better. Functions representing treatment 1 and treatment 2 of this scenario are well depicted by $f_1(x) = \frac{20}{e^{0.01(x+16)+1}} - 10$ and $f_2(x) = \frac{20}{e^{-0.01x+1}} - 10$. Scenario 6 investigates the distinctive features of the treatments in connection to the biomarker value of the patient. In this case, treatment 1 is independent of the biomarker while treatment 2 is of step behaving and this situation is represented by $f_1(x) = 5$ and $f_2(x) = \begin{cases} 8, if\ x \leq -8 \\ -8, otherwise \end{cases}$. The functions representing the treatments in the considered scenarios are presented in Figure **1**. In fact, Figure **1** displays the relationships between the patient's biomarker $(x-axis)$ and the respective outcomes of the treatments $(y-axis)$ by scatter plots in all different scenarios for both treatments along with the underlying functions by the curves.

The one-sided Type 1 error, two-sided Type 1 error and power are the main performance measures used to evaluate the proposed design. The performance of the proposed RAR design with exponentially decreasing sequence of probability is compared with RAR design with linearly decreasing probability and the fixed randomization design as well. We use the probability of Type I error for two sided test as $\alpha = 0.05$ and accordingly one sided test to assess the power. In addition to these statistical measures, the proportion of patients allotted to the best treatment is also a significant parameter. But, the very aim of the proposed RAR design using biomarker will ensure the maximum proportion unless otherwise the estimation of the best treatment goes incorrect.

Simulation results show the proportion of patients receiving the best treatment in the proposed RAR design with exponentially decreasing sequence outperforms the existing design with linearly decreasing probability sequence in all the considered scenarios excluding Scenario 1. They are presented in Table **3** for exponential probability sequence and Table **4** for linearly decreasing sequence. Obviously, we noticed that this proportion is higher compared to the conventional fixed randomization (FR) as well**.** Among

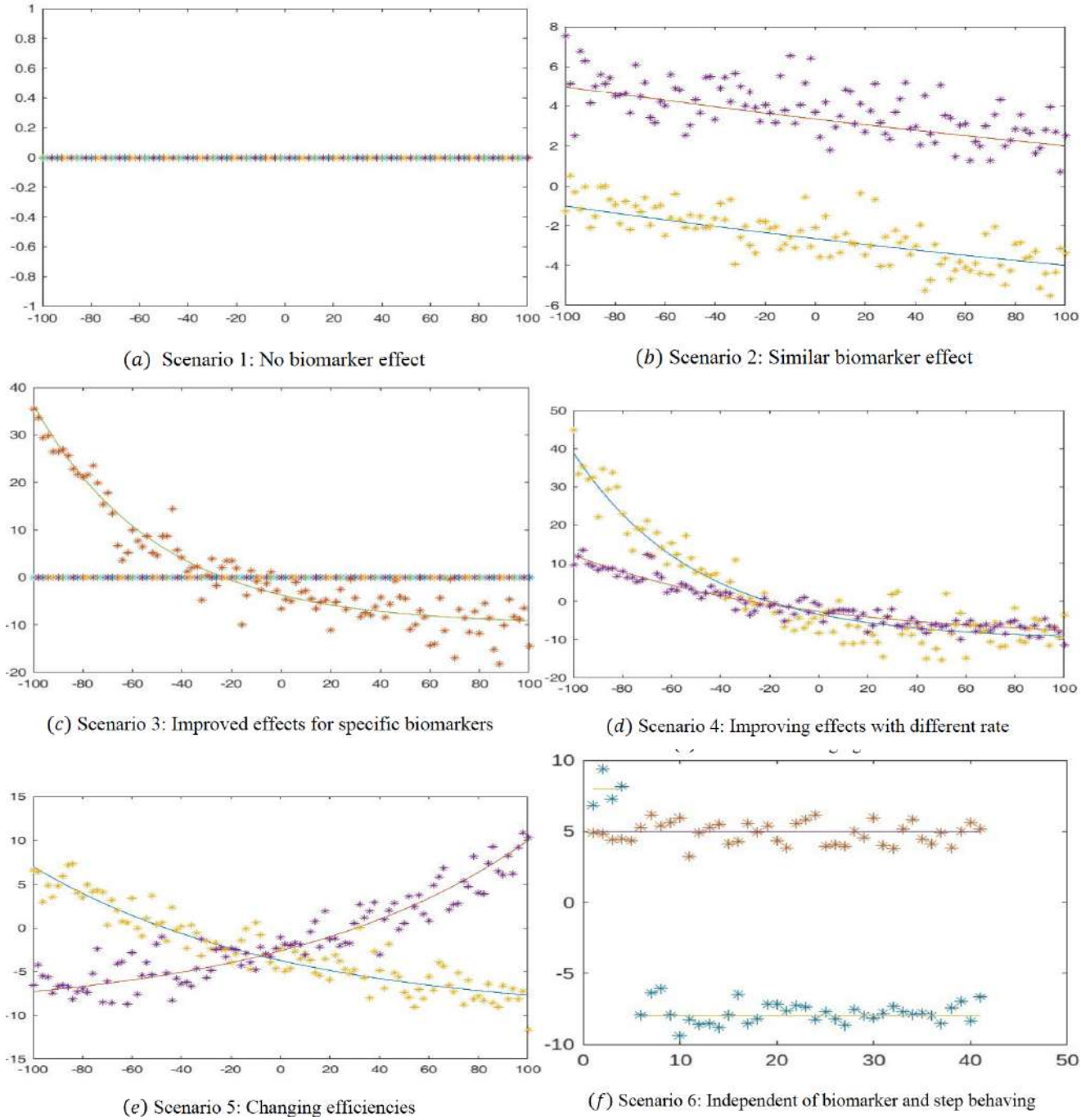

Figure 1: Relationships between patient's biomarker and the outcome of the treatments.

the regression methods considered, the GP model is found to be performing better in Scenarios 3, 4, 5 and 6. At the same time, the performance of GP model is very close in case of Scenarios 1 and 2 when compared to other models. The performance of the Spline regression is the least irrespective of different scenarios of treatments excluding Scenario 1. However, in specific, the Polynomial Regression is performing poorer than Spline for Scenario 6. Considering various scenarios as a whole, it is observed that regression methods show impressive proportion of allocation irrespective of the method except for Scenario 1. This is perhaps due to the fact that the difference between the observed outcomes of considered treatments is very small. However, the summary considering various scenarios is that the proportion of allocation is high when there is a significant difference between the treatments. It is needless to mention that the allocation efficiency will reduce when the difference between the outcomes of the treatments become smaller. It is interesting to note that performance of the proposed design using

Table 3: The Proportion of Patients Received the Best Treatment - Exponential Sequence

	NN	PR	Spline	RF	GP	FR
Scenario 1	0.5019	0.5016	**0.5027**	0.4858	0.4963	0.5008
Scenario 2	**0.7240**	0.7223	0.7006	0.7198	0.7237	0.4992
Scenario 3	0.6861	0.6912	0.6657	0.6795	**0.6914**	0.5009
Scenario 4	0.6263	0.6368	0.5900	0.6235	**0.6425**	0.5006
Scenario 5	0.6908	0.6928	0.6651	0.6885	**0.6945**	0.5018
Scenario 6	0.6956	0.6596	0.6633	0.6967	**0.7028**	0.4983

Table 4: The Proportion of Patients Received the Best Treatment - Linear Sequence

	NN	PR	Spline	RF	GP	FR
Scenario 1	0.5019	0.5026	0.5027	0.4999	0.4923	0.5008
Scenario 2	0.6794	0.6783	0.6661	0.6793	0.6796	0.4992
Scenario 3	0.6500	0.6540	0.6340	0.6457	0.6544	0.5009
Scenario 4	0.6035	0.6108	0.5729	0.6000	0.6142	0.5006
Scenario 5	0.6542	0.6559	0.6361	0.6552	0.6569	0.5018
Scenario 6	0.6570	0.6269	0.6270	0.6603	0.6643	0.4983

biomarkers with regression methods becomes still higher if more patients are involved to provide more information to the trial under consideration.

The average one-sided power indicates the probability of correctly identifying the experimental treatment having significant effect and is depending on the scenario considered in the analysis. This is illustrated in Table **5** for exponentially decreasing probability sequence and in Table **6** for linearly decreasing sequence. It can be seen from Tables **5** and **6** that Scenarios 1 and 3 using exponentially

Table 5: One Sided Power- Exponential Sequence

	NN	PR	Spline	RF	GP	FR
Scenario 1	0.0300	0.0260	0.0380	0.0400	0.0360	0.0260
Scenario 2	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Scenario 3	0.3920	0.3970	0.2680	0.2800	0.4020	0.0020
Scenario 4	0.0070	0.0020	0.0060	0.0200	0.0000	0.0660
Scenario 5	0.1540	0.1550	0.1720	0.1400	0.1760	0.1920
Scenario 6	1.0000	1.0000	0.9980	1.0000	1.0000	0.9280

Table 6: One Sided Power- Linear Sequence

	NN	PR	Spline	RF	GP	FR
Scenario 1	0.0300	0.0150	0.0250	0.0200	0.0260	0.0260
Scenario 2	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Scenario 3	0.2130	0.2130	0.1460	0.1200	0.2320	0.0020
Scenario 4	0.0050	0.0010	0.0090	0.0200	0.0000	0.0660
Scenario 5	0.1690	0.1670	0.1910	0.1400	0.1920	0.1920
Scenario 6	1.0000	1.0000	0.9960	1.0000	1.0000	0.9280

Table 7: Two Sided Power - Exponential Sequence

	NN	PR	Spline	RF	GP	FR
Scenario 1	0.0640	0.0610	0.0660	0.0800	0.0560	0.0550
Scenario 2	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Scenario 3	0.3920	0.3970	0.2680	0.2800	0.4020	0.1190
Scenario 4	0.7080	0.7340	0.4260	0.7200	0.7820	0.0710
Scenario 5	0.1550	0.1570	0.1730	0.1400	0.1760	0.1950
Scenario 6	1.0000	1.0000	0.9980	1.0000	1.0000	0.9280

Table 8: Two Sided Power - Linear Sequence

	NN	PR	Spline	RF	GP	FR
Scenario 1	0.0650	0.0480	0.0570	0.0400	0.0500	0.0550
Scenario 2	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Scenario 3	0.2150	0.2140	0.1470	0.1200	0.2340	0.1190
Scenario 4	0.5170	0.5440	0.2830	0.5000	0.6020	0.0710
Scenario 5	0.1720	0.1690	0.1920	0.1400	0.1920	0.1950
Scenario 6	1.0000	1.0000	0.9960	1.0000	1.0000	0.9280

decreasing probability sequence outperform both the RAR design using biomarker with linearly decreasing probability sequence and the FR design. In fact, either RF or GP performs well among various regression models in RAR design using biomarkers. It is interesting to note that the performance of all methods are equally good in Scenarios 2 and 6.

Similarly, here also the average two-sided power indicates the probability of correctly identifying the experimental treatment having significant effect and is depending on the scenario considered in the analysis. The results are given in Table **7** for exponentially increasing probability sequence and in Table **8** for linearly decreasing sequence. From Tables **7** and **8**, it is observed that the Scenarios 1, 3 and 4 with exponentially decreasing probability sequence outperform both the RAR design using biomarker with linearly decreasing probability and FR design. Here also all methods in the comparison are performing to the best in Scenarios 2 and 6 and at the same time, the FR stands as the best method for Scenario 5.

The justification for the observations about power of both the cases are apparent for Scenarios 2, 3 and 6. FR has the highest power in Scenario 5 for both the cases and is easy to understand that this is due to the difference between the treatments considered is large in this case. In Scenario 3 the allotment of patients by the FR design is a mixture of patients with high and low biomarker values and hence must be lesser than the RAR design using biomarker.

4. DISCUSSION AND CONCLUSIONS

RAR is to use the latest knowledge about the considered treatments to carry out allocation of the next patient. It is to ensure that the right for quality life of patients even when they are undergoing clinical trial against the FR in which patients are allotted to the considered treatments with equal probability. In the proposed method, the efficiency of RAR is enhanced by estimating the best treatment using regression models of treatment efficiency over the biomarker values of all the previous patients. Then, the consecutive patients are allocated to this estimated treatment, not purely based on the efficiency of the treatment, but in addition, a probability sequence is used with a higher probability to the efficient treatment. This is due to the fact that the best treatment can never be determined at this stage of clinical trial and confidence on the best treatment increases with the successive patients. The probability computed by exponentially decreasing sequence in the proposed method ensures the allocation to the estimated best treatment with higher probability compared to the previous approach and thus the proposed approach performs better than the previous one. Further, this is ethical in the sense that the allocation is done neither

to the efficient treatment by over confidence nor ignoring the available knowledge about the estimated efficient treatment.

The proposed work is significant in the sense that it investigates and answers the misconceptions about the use of RAR design saying that it has low power and cannot be practical to implement. The proposed method, in fact, aims to enhance the efficiency of the allotment of patients to treatments in clinical trials. The method considers the customization of treatments based on certain biomarkers related to the treatment outcomes. For this purpose, regression models are suggested to study the impact of the biomarker on the outcome of the treatment. At the same time, in the proposed method, unlike the existing approaches, the allocation of patients is not decided just based on the regression model results alone, it also incorporates exponentially decreasing probability sequence to allot efficient treatments to patients with a higher probability. The introduction of probability concept is to avoid misjudgment on the efficiency of the treatment due to early conclusion. The study investigates various regression models as well using simulation by considering various scenarios that mimics real time situation in clinical trial.

The simulation study considers the cases where there are only two treatments. The performance of the proposed allocation procedure is measured by the proportion of patients allotted, Type I error and statistical power. Comparative analyses are also performed to identify best treatments and the respective models. We have noticed that the proposed study is much beneficial to the patients of rare diseases as it attempts to ensure the best treatment to the needy patients even when they are in clinical trials. The simulation study alleviates the general apprehension over the low power of RAR design and its practicality. The proportion of allotment to the best treatment is well captured by the simulation study. The study reveals that the exponentially decreasing probability sequence improves the allotment efficiency compared to the existing study that adopted linearly decreasing probability sequence. Our study reveals that while all the regression models considered are showing improved results compared to the fixed randomization method, there are differences in the performance of these regression methods used under various scenarios.

Being a study on clinical trials for rare disease, the present study has some limitations related to sample size, mainly with regard to the study on the estimation of confidence intervals and standard errors. As a future study it is planned to extend the scope of the proposed study by considering more than one biomarker. Further we plan to apply the Advanced algorithms of ML and compare the same with the proposed method. It will be attempted to extend the proposed approach the interval valued biomarkers as well.

ACKNOWLEDGEMENT

Authors would like to thank the reviewer and the Editorial Director for their valuable comments and suggestions that have helped to bring considerable improvements to the earlier version of the paper.

REFERENCES

[1] Ravichandran J, Deepa S. Process capability indices for processes when the underlying data are interval-valued. International Journal of Statistics in Medical Research 2024; 13: 210-218. https://doi.org/10.6000/1929-6029.2024.13.19

[2] Vera-Ponce VJ, Zuzunaga-Montoya FE, Sanchez-Tamay NM, Vásquez-Romero LEM, Loayza-Castro JA, Huaman-Vega CH, Tapia-Limonchi R, De Carrillo CIG. Sample Size and Statistical Power Calculation in Multivariable Analyses: Development and Implementation of "Sample Size Multi" Packages in R. International Journal of Statistics in Medical Research 2024; 13: 259-274. https://doi.org/10.6000/1929-6029.2024.13.24

[3] Cao Y, Liang J. Multiple mean comparison for clusters of gene expression data through the t-SNE Plot and PCA dimension reduction, International Journal of Statistics in Medical Research 2025; 14: 1-14. https://doi.org/10.6000/1929-6029.2025.14.01

[4] Jackson H, Bowen S, Jaki T. Using biomarkers to allocate patients in response-adaptive clinical trial. Communications in Statistics-Simulation and Computation 2023; 52(12): 5946-5965. https://doi.org/10.1080/03610918.2021.2004420

[5] Barnett HY, Villar SS, Geys H, Jaki T. A novel statistical test for treatment differences in clinical trials using a response-adaptive forward-looking Gittins Index Rule. Biometrics 2023; 79(1): 86-97. https://doi.org/10.1111/biom.13581

[6] Tang F, Gajeswki BJ. Comparative effectiveness research using Bayesian adaptive design for rare diseases: Response adaptive randomization reusing participants. Statistics in Biopharmaceutical Research 2023; 15(1): 154-163. https://doi.org/10.1080/19466315.2021.1961854

[7] Jennison C, Turnbull BW. Group Sequential Methods with Applications to Clinical Trials. Chapman and Hall/CRC. U.K. 2000. https://doi.org/10.1201/9781584888581

[8] Rosenberger WF, Stallard N, Ivanova A, Harper CN, Ricks ML. Optimal adaptive designs for binary Response trials. Biometrics 2004; 57(3): 909-913. https://doi.org/10.1111/j.0006-341X.2001.00909.x

[9] Hu F, Zhang LX, Cheung SH, Chan WS. Doubly adaptive biased coin designs with delayed responses. The Canadian Journal of Statistics 2008; 36(4): 541-549. https://doi.org/10.1002/cjs.5550360404

[10] Robertson DS, Lee KM, Lopez-Kolkovska BC, Villar SS. Response-adaptive randomization in clinical trials: From

myths to practical considerations. Statistical Science 2023; 38(2): 185-208. https://doi.org/10.1214/22-STS865

[11] Cheung YK, Inoue LY, Wathen JK, Thall PF. Continuous Bayesian adaptive randomization based on event times with covariates. Statistics in Medicine 2006; 25(1): 55-70. https://doi.org/10.1002/sim.2247

[12] Proschan M, Evans S. Resist the temptation of response adaptive randomization. Clinical Infectious Diseases 2020; 1(11): 3002-3004. https://doi.org/10.1093/cid/ciaa334

[13] Rosenberger WF, Vidyashankar AN, Agarwal DK. Covariate adjusted response-adaptive designs for binary response. Journal of Biopharmaceutical Statistics 2001; 11(4): 227-236. https://doi.org/10.1081/BIP-120008846

[14] Villar SS, Rosenberger WF. Covariate-adjusted response-adaptive randomization for multi-arm clinical trials using a modified forward-looking Gittins index rule. Biometrics 2018; 74(1): 49-57. https://doi.org/10.1111/biom.12738

[15] Thall PF, Wathen JK. Covariate-adjusted adaptive randomization in a sarcoma trial with multi-stage treatments. Statistics in Medicine 2005; 24(13): 1947-1964. https://doi.org/10.1002/sim.2077

[16] Freidlin B, McShane LM, Korn EI. Randomized clinical trials with biomarkers: Design issues. Journal of the National Cancer Institute 2010; 102(3): 152-160. https://doi.org/10.1093/jnci/djp477

[17] Hu C, Dignam JJ. Biomarker-driven oncology clinical trials: Key design elements, types, features and practical considerations. Statistical Methods in Precision Oncology: Special Series 2019; 1-12. https://doi.org/10.1200/PO.19.00086

[18] Wong CH, Siah KW, Lo AW. Estimation of clinical trial success rates and related parameters. Biostatistics 2019; 20(2): 273-286. https://doi.org/10.1093/biostatistics/kxx069

[19] Yang Y, Zhu D. Randomized allocation with nonparametric estimation for a multi-armed bandit problem with covariates. The Annals of Statistics 2002; 30(1): 100-121. https://doi.org/10.1214/aos/1015362186

[20] Park Y. Challenges and opportunities in biomarker-driven trials: adaptive randomization. Annals of Translational Medicine 2022; 10(18): 1035. https://doi.org/10.21037/atm-21-6027

[21] Sechidis K, Papangelou K, Metcalfe PD, Svensson D, Weatherall J, Brown G. Distinguishing prognostic and predictive biomarkers: An information theoretic approach. Bioinformatics 2018; 34(19): 3365-76. https://doi.org/10.1093/bioinformatics/bty357

Received on 19-03-2025 Accepted on 15-04-2025 Published on 22-05-2025

https://doi.org/10.6000/1929-6029.2025.14.28