

Bayesian response adaptive randomization in clinical trials with a time-trend: application to a real trial

Yunyun Jiang (✉ jyunyun2011@gmail.com)

George Washington University Milken Institute of Public Health <https://orcid.org/0000-0002-4438-1613>

Wenle Zhao

Medical University of South Carolina Department of Public Health Sciences

Valerie Durkalski-Mauldin

Medical University of South Carolina Department of Public Health Sciences

Research article

Keywords: Bayesian response adaptive randomization, time-trend, bias, adaptive allocation

Posted Date: November 12th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-35671/v2>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: In clinical trials with a large sample size, time-trend in response rates can affect the performance of Bayesian response adaptive randomization (BRAR).

Methods: To evaluate this impact, we utilize data from a previously completed randomized controlled trial that used a fixed 1:1 allocation. Subject response data from this study demonstrate a clear time-trend in the control group, but not in the treatment group. In this simulation study, we re-assign patients to treatment groups based on a BRAR algorithm, to examine the performance of BRAR as measured by the treatment effect estimation, the probability of early stopping, and the shift in adaptive allocation.

Results: Results from specific simulated study scenario show that in the presence of a time-trend, the timing of the first interim analysis is critical for the efficacy/futility decision making. Compared with fixed equal allocation, BRAR results in a higher probability of premature early stopping when time-trend effect exists. The magnitude of such impact varies among different BRAR algorithms.

Conclusions: Influential factors such as time trend, should be considered when planning the implementation of BRAR in large trials.

Trial Registration: URL: <https://www.clinicaltrials.gov>. Unique identifier: NCT00235495. Registered on October 10, 2005.

Background

Several large trials report using Bayesian response adaptive randomization (BRAR) to improve trial efficiency and patient ethics [1,2]. Changes in the study population profile and the response rate to the treatment over time can affect the operating characteristics of BRAR trials [3,4]. The cause of time-trend may involve standard of care improvements, subject protocol amendments, and recruitment site expansion. Altman et al. studied the time-trend and suggested investigators to monitor the time-trend graphically [4]. However, the response time-trend has rarely been examined in trial practice. The presence of a time-trend may bring bias in treatment effect estimates, and inflate type I error [5-12]. The impact of time trend on the performance of BRAR has been investigated under different time-trend patterns including no time trend, equal time trend and unequal time trend between two arms via simulation studies. [13, 14] The simulation results suggested that under BRAR with early stopping for efficacy, the type I error rate is inflated for all time-trend patterns, including unequal time-trends. Furthermore, treatment effect estimation can be biased under all time-trend patterns. In order to better understand the impact of time-trend in a real setting, we incorporated a BRAR design into a previously completed randomized controlled phase 3 trial with a time-trend. The goal of this simulation study is to investigate drawbacks of using BRAR in the presence of an unequal time-trend, and to explore the impact of the problem.

Method

The original ALIAS trial

ALIAS is a randomized, parallel-group, double-blind, phase 3 trial testing the treatment superiority of 25% albumin over saline in improving the outcome of acute ischemic stroke subjects [15]. The primary endpoint is favorable outcome, defined as either a modified Rankin scale score of 0 or 1, or an NIHSS score of 0 or 1, or both, at 90 days. With a fixed equal allocation, 422 and 419 subjects were randomized to the albumin (treatment) group and the saline (control) group respectively. Scheduled interim analyses were conducted at $n=275$ and $n=550$. The Data and Safety Monitoring Board requested an additional analysis at $n=732$, and stopped the trial for futility after 841 participants were randomized. The primary outcome did not differ by treatment assignment (albumin, 44.1%; saline, 44.2%. 95% confidence interval: 0.84 - 1.10 adjusting for baseline covariates). O'Brien and Fleming-type stopping guidelines were adopted for both efficacy and futility assessments. Unexpectedly, the response rate in the control arm increased over time while the response rate in the treatment arm remained stable across the trial. To visualize the time-trend of response rates, subjects are divided into 28 stages according to the enrollment sequence (i.e. every 30 subjects at each stage). The weighted least square (WLS) regression technique is applied to these data while taking into account the variation of arm size within each stage (**Figure 1**). The graph demonstrates a clear upward trend in the control arm's response rate; with the slope parameter $\beta=0.012$, $p\text{-value}<0.0005$. Meanwhile, the treatment arm's response rate is quite stable; with the slope parameter $\beta=0.00043$, $p\text{-value}=0.90$).

Simulation study design

The goal of this study design is to apply three different BRAR methods in redesigning of the ALIAS trial. The BRAR methods examined are the *probability-weighted allocation approach* (BRAR (1/2)) [16], the *natural lead-in allocation* (BRAR($n/2N$)) [17], and the *information-weighted allocation* (BRAR($1/2, \sigma^2$)) [18,19]. Information on the enrollment sequence, treatment assignment, and response outcome of the 841 subjects in the original ALIAS trial are used for this simulation study with BRAR. Interim analyses are conducted at $n=275$, $n=550$, and $n=732$, and the final analysis at $n=841$. Equal allocation is applied to the first 275 subjects. Response adaptive randomization probabilities are updated concurrently with each interim analysis. For each of the 4 enrollment segments (1-275, 276-550, 551-732, and 733-841), numbers of subjects in each arm are calculated based on the corresponding allocation probabilities. The trial will be stopped for efficacy if the posterior success probability for either arm being superior to the other exceeds the efficacy boundary of 0.995. The trial will be terminated for futility when the chance that posterior success probability for either arm being superior to the other is different from each other is less than 5%. Table 1 lists implementation parameters of the simulation study.

Table 1. Implementation parameters for simulating BRAR in the ALIAS trial

Parameters	Settings
Number of arms	2
Maximal sample size	841
RAR algorithms and parameters	1) Fixed equal randomization 2) BRAR($n/2N$) 3) BRAR($1/2, \sigma^2$) 4) BRAR($1/2$)
Balanced burn-in period	200, 350
Interim analyses and allocation ratio update time points	275,550,732
Efficacy stopping rule	$P(p_1^* > p_2^*) > 0.995$ or $P(p_2^* > p_1^*) > 0.995$
Futility stopping rule	$ P(p_1^* \geq p_2^*) - 0.5 < 0.05$
Estimation of early stopping	Proportion of simulation runs with: $P(p_1^* > p_2^*) > 0.995$ or $P(p_2^* > p_1^*) > 0.995$
Simulation iteration	10000 per scenario.

Simulation study dataset generation

Simulation study datasets, including enrollment sequence, treatment arm, and outcome, are generated by sampling from the ALIAS trial. For the treatment arm (without time-trend), simulation subjects are randomly selected from the entire ALIAS treatment arm with replacement. For the control arm (with time-trend), the i^{th} simulation subject is randomly selected from a group of 7 ALIAS control arm subjects with enrollment sequence in the range of $(i-3, i+3)$. For example, the 10th simulation subject in the control arm will be randomly selected from the 7 ALIAS subjects with enrollment sequence from 7 to 13 in the control arm.

BRAR algorithm

In this simulation study, we evaluated the operation characteristics of the trial under two burn-in period lengths (275 and 350) and three BRAR algorithms [13,14]. Burn-in period uses a fixed equal allocation for subject randomization. During the adaptive randomization phases, for each subject, the treatment assignment is generated based on the current treatment allocation ratio, which is obtained from the selected BRAR algorithm and responses of subjects enrolled prior to the current allocation ratio update. The fixed equal randomization is incorporated as a comparator.

The simulation study is re-executed 1000 times. Detailed information can be found in Figure 2 below. Since the goal of this study is to evaluate the impact of time trend on the performance of BRAR design, there is no safety monitoring procedures involved. The R, version 3.2.5 software (R Foundation for Statistical Computing) [20] was used for the analysis.

Results

Table 2 presents the results of the simulation studies. Regardless of what randomization algorithm is used, early trial termination occurs most frequently during the first interim analysis. Trials with first interim analysis at the sample size of 275, are about two times more likely to stop early than trials with first interim decision at sample size of 350 (39% vs 22%). The response rate difference between two arms is close to 0 at the third interim stage when fixed equal allocation is used, while BRAR has a greater treatment effect in favor of treatment arm at each interim analysis. Among three BRAR algorithms, BRAR($n/2M$) results in the least probability of early termination in the following interim analyses, followed by the BRAR ($1/2, \sigma^2$), and BRAR ($1/2$). It is worth noting that, with a shorter burn-in length, both BRAR ($1/2, \sigma^2$) (11%) and BRAR ($1/2$) (13%) result in a remarkably high probability of stopping early as compared with either fixed (4.3%) or BRAR ($n/2M$) (3.2%). Overall, there is a significant imbalance in treatment assignment between two groups when BRAR is used (greater than 5:3 for treatment versus control group), such imbalance increases with strength of adaptation of BRAR algorithms.

Table 2. Simulation results for BRAR implementation

Burn-in length	BRAR algorithm	Interimanalysis	Total EnrolledSubjects	Prob of early stopping for		n		% responses		Treatmenteffect (p1-p2)	Probability that one arm is better than the other arm	
				Efficacy	Futility	Arm1	Arm2	Arm1	Arm2	Mean±SD	P(trt>ctrl)	P(ctrl>trt)
275	Fixedequal (1:1)	1	1-275	0.346	0.007	137	138	0.44	0.313	0.127(0.057)	0.346	0
		2	276-550	0.043	0.043	137	138	0.432	0.385	0.048(0.036)	0.043	0
		3	551-732	0.003	0.146	93	89	0.434	0.43	0.004(0.029)	0	0.003
			732-841	----	----	55	54	0.435	0.448	-0.012(0.029)	---	---
		All	1-841	0.392	0.196	422	419				0.389	0.003
	BRAR($n/2N$)	1	1-275	0.346	0.007	137	138	0.44	0.313	0.127(0.057)	0.346	0
		2	276-550	0.032	0.039	171	104	0.432	0.385	0.049(0.035)	0.032	0
		3	551-732	0.052	0.073	119	63	0.434	0.403	0.031(0.048)	0.048	0.003
			732-841	----	----	70	37	0.435	0.418	0.017(0.029)	---	---
		All	1-841	0.43	0.119	497	342				0.426	0.003
	BRAR($1/2, \sigma^2$)	1	1-275	0.346	0.007	137	138	0.44	0.313	0.127(0.057)	0.346	0
		2	276-550	0.107	0.029	222	53	0.432	0.365	0.067(0.042)	0.107	0
		3	551-732	0.043	0.038	116	66	0.432	0.388	0.043(0.040)	0.043	0
			732-841	----	----	65	42	0.432	0.404	0.028(0.041)	---	---
		All	1-841	0.5	0.074	540	299				0.496	0
	BRAR(1/2)	1	1-275	0.346	0.007	137	138	0.44	0.313	0.127(0.057)	0.346	0
		2	276-550	0.125	0.031	220	55	0.433	0.366	0.066(0.045)	0.125	0
		3	551-732	0.059	0.044	133	49	0.433	0.386	0.047(0.043)	0.053	0.005
			732-841	----	----	80	28	0.433	0.394	0.039(0.043)	---	---
		All	1-841	0.53	0.082	570	270				0.524	0.005
350	Fixedequal (1:1)	1	1-350	0.17	0.026	175	175	0.441	0.352	0.089(0.051)	0.17	0
		2	351-550	0.052	0.027	100	100	0.437	0.381	0.056(0.034)	0.052	0
		3	551-732	0.001	0.115	92	90	0.437	0.428	0.009(0.031)	0	0.001
		4	732-841	----	----	56	53	0.438	0.445	0.007(0.030)	---	---
		All	1-841	0.223	0.168	423	418				0.222	0.001
	BRAR($n/2N$)	1	1-350	0.17	0.026	175	175	0.441	0.352	0.089(0.051)	0.17	0
		2	351-550	0.065	0.021	126	74	0.436	0.380	0.057(0.036)	0.062	0.002
		3	551-732	0.063	0.05	122	60	0.437	0.403	0.034(0.040)	0.061	0.001
		4	732-841	----	----	69	37	0.437	0.418	0.019(0.038)	0	0.002
		All	1-841	0.298	0.097	492	346				0.293	0.005
	BRAR($1/2, \sigma^2$)	1	1-350	0.17	0.026	175	175	0.441	0.352	0.089(0.051)	0.17	0
		2	351-550	0.116	0.027	153	47	0.436	0.372	0.064(0.043)	0.113	0.003
		3	551-732	0.044	0.045	124	58	0.435	0.394	0.041(0.042)	0.039	0.005
		4	732-841	----	----	70	36	0.436	0.405	0.031(0.042)	---	---
		All	1-841	0.33	0.098	522	316				0.322	0.008
	BRAR(1/2)	1	1-350	0.17	0.026	175	175	0.441	0.352	0.089(0.051)	0.17	0
		2	351-550	0.133	0.024	153	47	0.436	0.372	0.064(0.042)	0.133	0
		3	551-732	0.047	0.04	131	50	0.435	0.392	0.043(0.042)	0.044	0.003
		4	732-841	----	----	82	26	0.437	0.399	0.037(0.042)	---	---
		All	1-841	0.35	0.09	541	298				0.347	0.003

Figure 3 displays the path of allocation probability (for treatment arm) change across interim stages. The patterns differ by the algorithms. Among BRAR algorithms, BRAR ($n/2N$) has least shift from equal allocation, the overall randomization probability is maintained at 60%-70%. With BRAR ($1/2, \sigma^2$), the adaptive allocation starts from a high probability based on treatment benefit, and slides down quickly near the equal point (0.5). BRAR (1/2) also has a dramatic shift from equal allocation, after a gradual decline, it stabilizes at 70%-80%. BRAR ($1/2, \sigma^2$) shows the most change due to the time-trend and reduced treatment effect size over time; while the allocation ratio from BRAR ($n/2N$) is more stable.

Discussion

The time-trend phenomenon found in the ALIAS trial illustrates an average response rate difference between two arms that decreases over time due to patients' improved response outcome in the control arm. Given this existing time-trend pattern, the BRAR can induce higher early trial termination for efficacy as compared with fixed equal allocation. The BRAR tends to force the imbalance in treatment assignment between treatment and control arms based on the treatment effect in the early phase of the trial. This leads to the trial stopping before the response rate in control arm catches up with the treatment arm. Timing of the first interim analysis after the burn-in period is critical when implementing BRAR, particularly in the presence of a time-trend. The issue of time-trend with RAR design and the associated negative impacts have attracted great attention in recent years. Challenges still exist that limit our understanding of this problem. Current studies primarily examine time-trend effect based on a "linearity" assumption. [5,11] Nevertheless, in real application, the actual pattern of "time-trend" is often hard to predict. It can vary over time or differ by treatment arms, and is often concomitant with a certain degree of randomness. The ALIAS trial described in this paper provides a real example of a treatment-dependent trend effect, where the benefit in the control arm increases over time and the subjects' responses remain consistent in the treatment arm. In this research, we attempt to incorporate the BRAR into the original trial dataset and evaluate the performance under the influence of time-trend.

Simulation studies suggest that the adaptive randomization probability can be affected by treatment effect over time. As compared to fixed equal randomization, the BRAR is more likely to terminate trials early and has been shown to lead to inflated type I error and biased treatment estimates.[5-11] Having more subjects assigned to the better-performing arm may improve the efficiency to identify the superior treatment arm, depending on the model used for the efficacy analysis and success rate of the two treatments. However, in the presence of a time-trend, the trials that end prematurely are at risk of losing important information from the future subjects who are involved in the time-trend.

The re-design strategy proposed in this article is relatively novel. Luce BR, Connor JT, Broglio KR, et al. reported a re-analysis of the ALLHAT study using Bayesian adaptive design with pre-specified burn-in period length, timing and frequency of interim analyses, and the BRAR algorithm [21]. In order to present the time-trend in the control arm, subjects are randomly sampled from a series of moving blocks, which are arranged in a chronological enrollment order. For subjects in the arm without specific trend pattern, a random sampling is conducted among all treatment arm subjects. The purpose of doing this is to avoid any chance of trial early termination due to the reason that the number of subjects in the treatment runs out faster than those in the control arm. One challenge of the study redesign is that subjects need to be resampled, which increases the risk of having less heterogeneous subject population. On the other hand, the randomness of the study can be improved by increasing the number of times of trial executions (simulations).

Some argue that the impact of time-trend is quite minimal to the adaptive designs if only mild trend effect exists. [11] However the definition or threshold for what constitutes a 'mild' trend is difficult to determine. Our study shows that BRAR design can lead to erroneous conclusions about trial efficacy if the subjects' response changes over time. The trial which we studied here is a special case where the

trend is only present in the control arm. The allocation probability is continuously biased toward the treatment arm even though the observed treatment effect reduces over time.

An important goal of this research is to provide knowledge about the time-trend impact based on a real trial application, and further to propose useful recommendations on how to handle time-trend when using a BRAR design. Based on our findings, the strength of BRAR adaptation is associated with both magnitude of sample size imbalance and probability of incorrect premature early stopping. We observe a significantly elevated probability of early stopping in trials with BRAR ($1/2, \sigma^2$) and BRAR ($1/2$). This can be tempered by a larger burn-in period. Overall, among all BRAR algorithms, BRAR ($n/2N$) appears to maintain a good balance in treatment assignment and avoids premature early stopping in the presence of a time-trend. Thus, appropriate timing of interim analyses and a randomization approach that can lead to smaller variability in treatment allocation are recommended in trials that are more prone to a time-trend.

Conclusions

Influential factors such as time trend, should be considered when planning the implementation of BRAR in large trials.

Abbreviations

BRAR: Bayesian response adaptive randomization; ALIAS: Albumin in Acute Ischemic Stroke; ALLHAT: The Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial.

Declarations

Ethics approval and consent to participate

The ALIAS Principle Investigators provided access to the de-identified data of the ALIAS trial that was used for our study.

Consent for publication

Not applicable.

Availability of Data and Materials

The ALIAS dataset is available in the NINDS data repository of archived clinical research datasets (<https://www.ninds.nih.gov/Current-Research/Research-Funded-NINDS/Clinical-Research/Archived-Clinical-Research-Datasets>) and can be accessed upon request to the NINDS Clinical Research Liaison.

Competing interests

The authors declare that they have no competing interests.

Funding

This research is partly supported by the NIH/NINDS grants U01NS0059041 (NETT), U24NS100655 (SIREN) and U01NS087748 (StrokeNet). The trial was funded by cooperative agreements from the U.S. National Institutes of Health/National Institute of Neurological Disorders and Stroke. A representative of NINDS (CSM) participated in the study leadership committee and contributed to design decisions as well as the interpretation and writing of the manuscript. The listed NIH grants supported the efforts of the authors however only the listed authors participated in the analysis and interpretation of the simulation study and in the writing of the manuscript.

Authors' contributions

YJ, WZ and VDM were responsible for study conception and study design. YJ drafted the initial version of the manuscript, with input from WZ and VDM. WZ and VDM critically reviewed and provided feedback on the study design and manuscript. All authors have read and approved the manuscript.

Acknowledgement

This research is partly supported by the NIH/NINDS grants U01NS0059041 (NETT), U01NS087748 (StrokeNet) and U24 NS100655 (SIREN) as partial support. The authors thank the anonymous reviewers for their careful review and great comments for this manuscript. We would like to thank the ALIAS investigators for providing access to the trial data.

References

1. Yin G, Chen N, Lee JJ. Phase II trial design with Bayesian adaptive randomization and predictive probability. *J R Stat Soc Ser C Appl Stat.* 2012; 61: 219–235. doi:10.1111/j.1467-9876.2011.01006.x.
2. Lee JJ, Gu X, Liu S. Bayesian adaptive randomization designs for targeted agent development. *Clin Trials.* 2010; 7(5): 584-596.
3. Lipsky AM, Greenland S. Confounding due to changing background risk in adaptively randomized trials. *Clin Trials.* 2011; 8(4): 390-7.
4. Altman DG, Royston JP. The hidden effect of time. *Stat Med,* 1988; 7(6): 629-37.
5. Karrison TG, Huo D, Chappell R. A group sequential, response adaptive design for randomized clinical trials. *Control Clin Trials.* 2003; 24(5): 506-22.
6. Thall PF, Fox PS, Wathan JK. Some Caveats for Outcome Adaptive Randomization in Clinical Trials. In: Sverdlov O (ed). *Modern adaptive randomized clinical trials: statistical, operational, and regulatory aspects.* Randomization in clinical trials. Oxford: Taylor & Francis, 2015.
7. Kairalla JA. et al. Adaptive trial designs: a review of barriers and opportunities. 2012; 13: 145.

8. Simon R, Simon NR. Using Randomization Tests to Preserve Type I Error With Response adaptive and Covariate adaptive Randomization. *Stat Probab Lett.* 2011; **81**(7): 767-72.
9. Bowden J, Trippa L. Unbiased estimation for response adaptive clinical trials. *Stat Methods Med Res.* 2017;**26**(5):2376-2388. DOI:10.1177/0962280215597716.
10. Thall PF, Wathen Practical Bayesian Adaptive Randomization in Clinical Trials. *Eur J Cancer.* 2007; **43**(5): 859-66.
11. Cook JD. The effect of population drift on adaptively randomized trial. UT MD Anderson Cancer Center Department of Biostatistics Working Paper Series 2017.
12. Thall PF, Fox PS, Wathen JK. Statistical controversies in clinical research: scientific and ethical problems with adaptive randomization in comparative clinical trials. *Ann Oncol.* 2015; 26(8): 1621-8.
13. Jiang Y, Zhao W, Mauldin V. Time-trend impact on treatment estimation in two-arm clinical trials with a binary outcome and Bayesian response adaptive randomization, *Journal of Biopharmaceutical Statistics*, 2020; 30(1): 69-88, DOI: [1080/10543406.2019.1607368](https://doi.org/10.1080/10543406.2019.1607368) .
14. Jiang Y, Zhao W, Mauldin V. Impact of adaptation algorithm, timing, and stopping boundaries on the performance of Bayesian response adaptive randomization in confirmative trials with a binary endpoint. *Contemporary Clin Trials.* 2017; **62**:114-120.
15. Ginsberg MD, Palesch YY and Hill MD et al. High-dose albumin treatment for acute ischemic stroke (ALIAS) Part 2: a randomized, double-blind, phase 3, placebo-controlled trial. *Lancet Neurol.* 2013; **12**(11):1049-58.
16. Thompson WR. On the Likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika.* 1933; 25(3-4): 285-294.
17. Bello GA, Sabo Outcome-adaptive allocation with natural lead-in for three-group trials with binary outcomes. *Journal of Statistical Computation and Simulation.* 2016; 86(12): 2441-2449.
18. Connor JT, Luce BR, Broglio KR, et al. Do Bayesian adaptive trials offer advantages for comparative effectiveness research? Protocol for the RE-ADAPT study. *Clin Trials.* 2013; 10(5): 807-27.
19. Connor JT, Elm JJ, Broglio KR. Bayesian adaptive trials offers advantages in comparative effectiveness trials: an example in status epilepticus. *J Clin Epidemiol.* 2013; 66(8): 130-137.
20. R Development Core Team (2014), *R: A Language and Environment for Statistical Computing.* Vienna, Austria: the R Foundation for Statistical Computing. Available online at <http://www.R-project.org/>.
21. Luce BR, Connor JT, Broglio KR, et al. Using Bayesian Adaptive Trial Designs for Comparative Effectiveness Research: A Virtual Trial Execution. *Ann Intern Med.* 2016; **165**(6):431-438. doi: 10.7326/M15-0823.

Figures

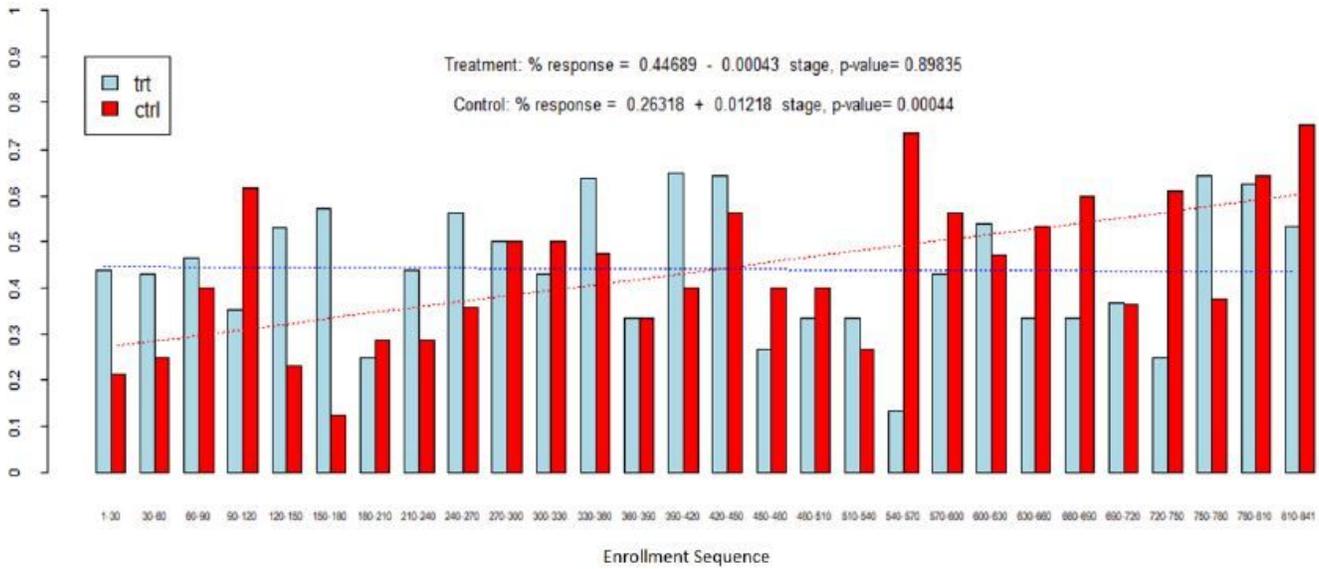


Figure 1

Response rates by enrollment stage (per 30 subjects) and arm

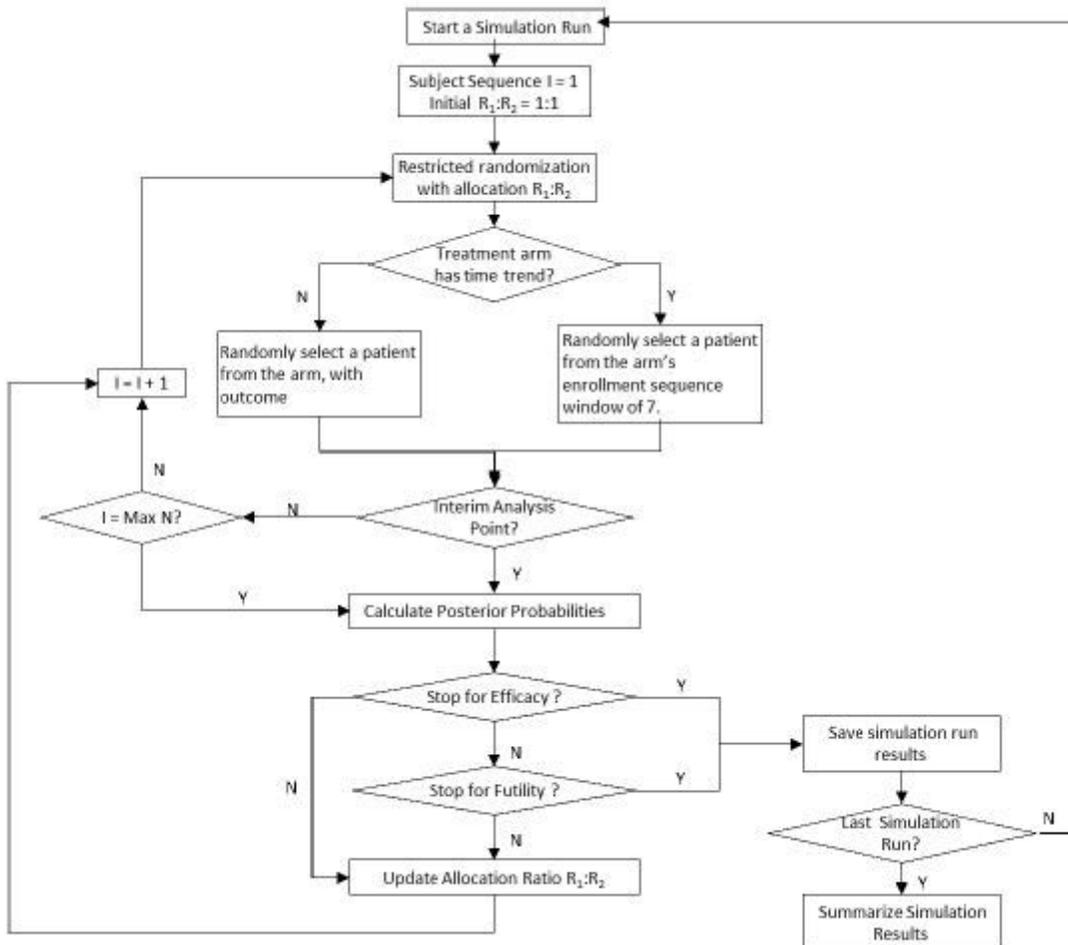


Figure 2

Procedures of simulation BRAR studies with ALIAS trial data

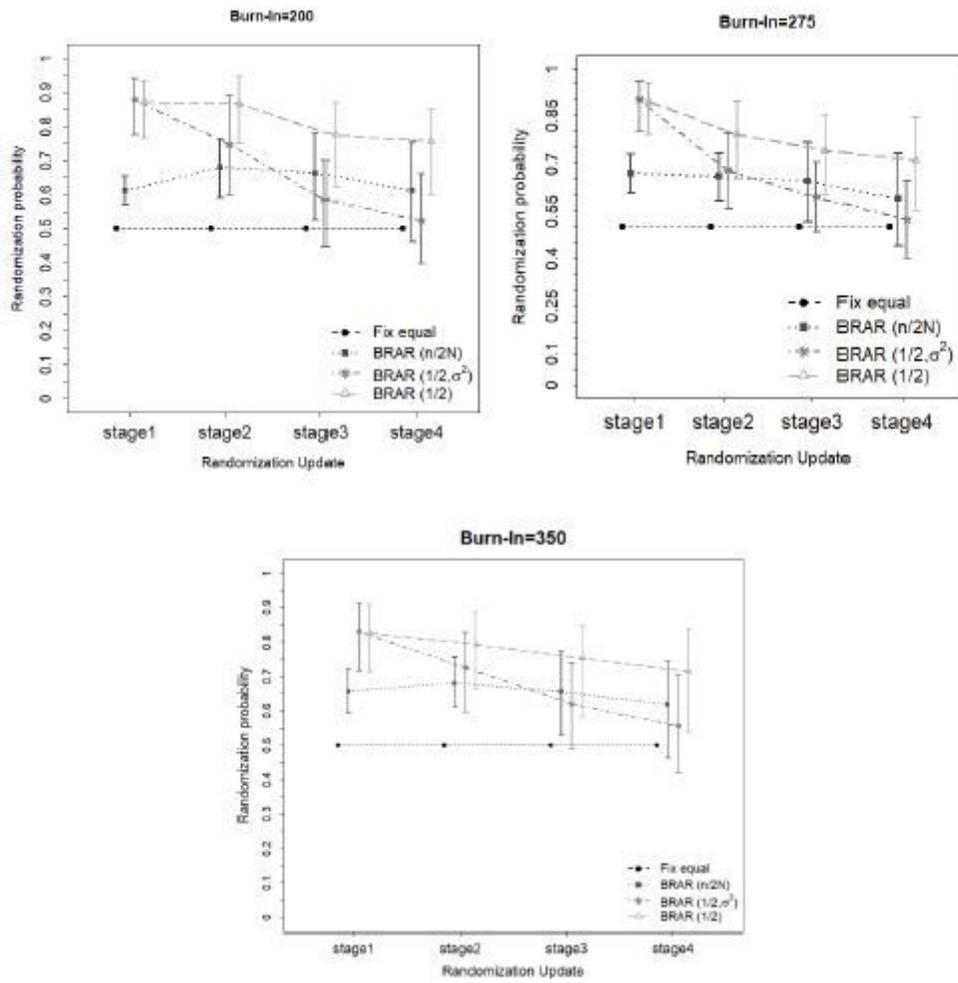


Figure 3

Randomization probability over time at each randomization update.