Revista Colombiana de Estadstica
July 2019, Volume 42, Issue 2, pp. 209 to 224
DOI: http://dx.doi.org/10.15446/rce.v42n2.65344

A Randomized Two Stage Adaptively Censored Design With Application to Testing

Un diseño aleatorio censurado adaptativo en dos etapas con aplicación

Uttam Bandyopadhyay[1,a], Rahul Bhattacharya[1,b]

[1]Department of Statistics, University of Calcutta, Kolkata, India

Abstract

A two stage randomized design is developed for two treatment clinical trials in which response variables are exponential and the observations are censored by using failure censoring and time censoring in the first and second stages respectively. The censoring time for the second stage is determined from the outcomes of the first stage. An application to testing for the equality of treatment effects is given along with a comparative study with relevant properties.

***Key words*:** Two stage design; Censoring; Inference.

Resumen

Se desarrolla un diseño aleatorizado en dos etapas para dos ensayos clínicos de tratamiento en los que las variables de respuesta son exponenciales y las observaciones se censuran utilizando la censura por fallo y la censura temporal en la primera y segunda etapas, respectivamente. El tiempo de censura para la segunda etapa se determina a partir de los resultados de la primera etapa. Una aplicación para probar la igualdad de efectos del tratamiento se da junto con un estudio comparativo con propiedades relevantes.

***Palabras clave*:** Diseño de dos etapas; Censura; Inferencia.

1. Introduction

Clinical trials involve a number of treatments and hence a comparison among the treatments is natural to identify the best treatment. The responses in a clinical trial are often time-to-event (e.g. remission) data. Since such trials recruit a

[a]PhD. E-mail: ubandyopadhyay08@gmail.com
[b]PhD. E-mail: rahul_bhattya@yahoo.com

number of human volunteers, a complete experiment results in loss of life. Thus it is not ethically appealing to allow the trial to continue until all the patients respond. Moreover, some of the patients do not visit the clinic to register the response. Therefore, it is neither feasible nor ethical to continue the trial indefinitely, and a curtailment becomes necessary.

Again ethics is one of the important concerns in any clinical trial, and hence we need to provide subjects the best possible care during the trial. This amounts to assign the subjects to the treatment doing better. But before the trial is actually conducted, the better treatment can not be determined, and hence we suggest to conduct the trial in two stages, where the subjects of the latter stage are assigned to the treatment doing better in the first stage. Specifically, we adopt a specified type of two stage (see, for example, Coad 1992) to compare two treatments based on a fixed number of patients, in which the first stage consists of assigning equal number of subjects to each treatment. Depending on the responses obtained in the first stage, the better performing treatment is decided and the remaining subjects are given this treatment exclusively.

The early work of Bandyopadhyay & Bhattacharya (2001) incorporated failure censoring in both the stages with an aim to explore the estimation aspects for the mean difference on reliability context. In a later work, Bandyopadhyay, Biswas & Bhattacharya (2009) adopted the same design with random censoring for every subject but determined the randomization probability for the second stage patients within a Bayesian framework. Use of censoring for response adaptive trials is further observed in Zhang & Rosenberger (2007), Sverdlov, Tymofyeyev & Wong (2011, 2014) and Biswas, Bhattacharya & Park (2016), among others. However, in all these instances though the allocation adopted is not two stage, censoring mechanism is fixed for all the subjects.

In the current work, we develop a two stage procedure with failure censoring in the first stage and determine the randomization probability of each second stage subject based on the first stage outcome. Unlike the available procedures, we suggest an adaptive time censoring in the second stage, which uses the first stage data in a convenient way to set the censoring time. Assuming exponential treatment responses, the procedure together with the likelihood ratio(LR) test for the hypothesis of equality of mean treatment responses is described in Section 2. A reasonable two stage competitor is suggested and compared with the proposed one through an extensive simulation based study in Section 3. Finally, Section 4 concludes, followed by some technical details in Appendix.

2. Two Stage Adaptive Time Censoring and Related Test Procedure

In this section we provide the proposed two stage adaptive sampling scheme with failure censoring in the first stage and time censoring in the second stage and a follow up test procedure. These are discussed in the following subsections.

2.1. Censoring Procedure and the Associated Estimation

We consider the situation, where two treatments, A and B, are on a clinical trial and the responses to the treatments are exponentially distributed with mean life times λ_A and λ_B, respectively. Let N be a prefixed number of subjects to be examined sequentially. We assume that the treatment with the higher capability of increasing the survival time is better. The allocation is carried out in two stages by assigning m subjects to each treatment arm in the first stage, where $m = \lfloor N\theta \rfloor$ and $\theta \in (0, \frac{1}{2})$. But, due to ethical and administrative convenience, we record only a preassigned number, r, of responses from each treatment group, where the censoring number r is so chosen that $r = \lfloor m\rho \rfloor$ for some $\rho \in (0, 1)$. Although curtailment of the study after a fixed number of events are observed (i.e. Type II censoring) is common in animal studies due to budget and time constraints (Xiong, Yan & Ming 2003), it is not unlikely in clinical trials with human beings. For example, consider a clinical trial, where the tumor-free survival times of the patients with squamous carcinoma of the oropharynx are the responses. Then with the passage of time progression in tumors is natural and therefore, it is not ethical to continue the trial up to a prefixed time or until all the responses are observed. Hence the experimenter needs to terminate the study after a fixed proportion of responses are observed. One such instance can be found in Bhattacharya (2007), where a real clinical trial involving patients with squamous carcinoma of the oropharynx (Kalbfleisch & Prentice 1980) is analysed adopting a Type II progressive censoring.

Now, to identify the better performing treatment(or the "winner") after the first stage, we find that the likelihood of the first stage data under failure censoring is proportional to

$$L_1 = \prod_{k=A,B} \left\{ {\lambda_k}^{-r} e^{-T_{1k}/\lambda_k} \right\},$$

where T_{1k} is the total observed lifetime in the first stage for the kth treatment, $k = A, B$. Thus T_{1k} is sufficient for λ_k and hence treatment k is decided as the winner after the first stage if T_{1k} is the largest. Naturally, an ethical strategy would be to assign the remaining $(N-2m)$ subjects exclusively to A or B as $T_{1A} > T_{1B}$ or $T_{1A} < T_{1B}$. However, such a strategy though appealing is not randomized. That is, it lacks the requirement of concealment of allocation (Rosenberger & Lachin 2015). Therefore, we incorporate randomization in the second stage through the allocation probabilities $\pi_{mi}, i \geq 2m + 1$ for the patients of the second stage, where π_{mi} depends on the first stage data and is set to favour the winner from the first stage for further allocation. Although a number of choices (Rosenberger & Lachin 2015) of π_{mi} is possible, we suggest to consider $\pi_{mi} = P^*(T_{1A} > T_{1B})$, where $P^*(T_{1A} > T_{1B})$ is a consistent estimator of $P(T_{1A} > T_{1B})$ based on the first stage data. Specifically, based on the first stage data, $\frac{2T_{1k}}{\lambda_k} \sim \chi^2_{2r}$ independently for each $k = A, B$ and hence $\frac{\lambda_B}{\lambda_A} \frac{T_{1A}}{T_{1B}} \sim F_{2r,2r}$. Then we can express $P(T_{1A} > T_{1B})$ as $P(F_{2r,2r} > \frac{\lambda_B}{\lambda_A})$, which depends on unknown (λ_A, λ_B) and we suggest to replace λ_k by its consistent estimator $\hat{\lambda}_{km} = \frac{T_{1k}}{r}$ based on the first stage to get $P^*(T_{1A} > T_{1B}) = P(F_{2r,2r} > \frac{\hat{\lambda}_{Bm}}{\hat{\lambda}_{Am}})$.

Furthermore, in the second stage, we adopt adaptive time censoring for each treatment arm with an aim to make the duration of the two stages identical. If $X_{(r)}$ and $Y_{(r)}$ denote, respectively, the r th order statistics obtained from the responses of treatment A and treatment B in the first stage, the duration of the first stage is $\max(X_{(r)}, Y_{(r)})$, and hence to make the observations of the second stage comparable with those of the first stage, we continue assignment of subjects of the second stage up to time $X_{(r)}$ for treatment A and $Y_{(r)}$ for treatment B.

Let $\delta_{mi}(= 1 - \bar{\delta}_{mi})$ be the treatment indicator in the second stage, where $\delta_{mi} = 1$ or 0 as treatment A or B is assigned. Then, given the first stage data, δ_{mi} are iid Bernoulli random variables with success probability π_{mi}. Thus the likelihood of the data from the entire trial is

$$L(\lambda_A, \lambda_B) = L_1 \prod_{i=2m+1}^{N} \pi_{mi}^{\delta_{mi}} (1 - \pi_{mi})^{\bar{\delta}_{mi}} \left\{ {\lambda_A}^{-n_A} e^{-T^*_{2A}/\lambda_A} \right\} \left\{ {\lambda_B}^{-n_B} e^{-T^*_{2B}/\lambda_B} \right\}, \quad (1)$$

where

$$T^*_{2A} = \sum_{i=2m+1}^{N} \delta_{mi} \min(X_i, X_{(r)}),\ n_A = \sum_{i=2m+1}^{N} \delta_{mi} I_{[X_i < X_{(r)}]},$$

$$T^*_{2B} = \sum_{i=2m+1}^{N} \bar{\delta}_{mi} \min(Y_i, Y_{(r)}),\ n_B = \sum_{i=2m+1}^{N} \bar{\delta}_{mi} I_{[Y_i < Y_{(r)}]}$$

with $I_{[.]}$ as the indicator function, $\{X_{2m+1}, X_{2m+2}, \ldots\}$ and $\{Y_{2m+1}, Y_{2m+2}, \ldots\}$ as the observations in the second stage corresponding to A and B, respectively. Hence, maximizing (1) with respect to (λ_A, λ_B), we get the maximum likelihood estimators of mean lifetimes as

$$\widehat{\lambda}_{kN} = \frac{T_{1k} + T^*_{2k}}{r + n_k}, k = A, B.$$

2.2. Test for the Mean Difference

In order to carry out inference following the proposed allocation design, we assume treatment A as experimental and treatment B as existing. Then a natural objective is to justify whether the experimental treatment is more promising than the existing one, that is, whether the experimental treatment enhances lifetime, on an average. Consequently, we consider testing $H_0 : \lambda_A = \lambda_B$ against $H : \lambda_A > \lambda_B$ and develop a test based on the likelihood ratio(LR) criterion

$$\Lambda_N = \frac{\sup_{H_0} L(\lambda_A, \lambda_B)}{\sup_{H_0 \cup H_1} L(\lambda_A, \lambda_B)}.$$

A little manipulation yields

$$\begin{aligned} \Lambda_N &= \left(\frac{\widehat{\lambda}_{AN}}{\widehat{\lambda}_{0N}}\right)^{r+n_A} \left(\frac{\widehat{\lambda}_{BN}}{\widehat{\lambda}_{0N}}\right)^{r+n_B} \quad \text{if } \widehat{\lambda}_{AN} > \widehat{\lambda}_{BN} \\ &= 1 \quad \text{if } \widehat{\lambda}_{AN} \leq \widehat{\lambda}_{BN}, \end{aligned}$$

where $\widehat{\lambda}_{0N}$, the maximum likelihood estimator obtained by maximizing (1) under the restriction $\lambda_A = \lambda_B = \lambda_0$, is of the form

$$\widehat{\lambda}_{0N} = \frac{T_{1A} + T^*_{2A} + T_{1B} + T^*_{2B}}{2r + n_A + n_B}.$$

As no further simplification of the criterion is possible, we reject the null hypothesis at a preassigned level $\alpha(0 < \alpha < 1)$ if

$$-2 \ln \Lambda_N > c$$

for some c satisfying

$$P_{H_0}(-2 \ln \Lambda_N > c) = \alpha.$$

It is shown (see Appendix) that, as $N \to \infty$,

$$-2 \ln \Lambda_N \to W$$

in distribution under the null hypothesis, where for any real w,

$$P(W \leq w) = \frac{1}{2} + \frac{1}{2} \int_0^w f_{\chi^2_1}(x) dx$$

with $f_{\chi^2_1}(x) = \frac{exp(-x/2)}{\sqrt{2\pi x}},\ x > 0.$

3. A Comparative Study

For a comparative assessment of the proposed allocation design, we consider a reasonable competitor and relevant performance measures.

3.1. Competitor

Here, we consider an allocation design (referred to as equal allocation), which is the same as the proposed one except that the subjects in the second stage are assigned to either treatment with equal probability. That is, for the competitor, we ignore the outcome of the first stage and suggest $\pi_{mi} = \frac{1}{2}$ for every incoming subject of the second stage. Thus, such a randomization procedure keeps the form of the likelihood function unchanged apart from a known multiplier. Consequently the corresponding LR criterion(say, $\tilde{\Lambda}_N$) for testing $H_0 : \lambda_A = \lambda_B$ against $H_1 : \lambda_A > \lambda_B$ takes the same form as that derived in Section 2 except that δ_{mi}, in this case, are iid Bernoulli($\frac{1}{2}$), independently of the first stage. Then it follows from Result A.2 in the appendix that, as $N \to \infty$,

$$-2 \ln \tilde{\Lambda}_N \to W$$

in distribution under the null hypothesis with W defined earlier.

3.2. Assessing Performance

For a meaningful assessment of the allocation design, we consider measures of both ethical and inferential aspects. We compute power of the LR test to investigate the discriminating ability of the proposed allocation design. Since ethics is one of our motivation in the development stage, expected allocation proportion(EAP) to the better treatment is used to measure the ethical aspect. Usually, out of N assignments in a two stage design, the intended number of allocation to treatment A is $m + \sum_{i=2m+1}^{N} \delta_{mi}$, but due to censoring in both the stages for the proposed design, the observed number of assignments reduces to $r + \sum_{i=2m+1}^{N} \delta_{mi} I_{[X_i < X_{(r)}]}$. Further, the actual number of total assignments, that is, the number of events under the proposed design is $2r + \sum_{i=2m+1}^{N} \delta_{mi} I_{[X_i < X_{(r)}]} + \sum_{i=2m+1}^{N} \bar{\delta}_{mi} I_{[Y_i < Y_{(r)}]}$ and hence the observed proportion of allocation to treatment A is simply

$$\frac{r + \sum_{i=2m+1}^{N} \delta_{mi} I_{[X_i < X_{(r)}]}}{2r + \sum_{i=2m+1}^{N} \delta_{mi} I_{[X_i < X_{(r)}]} + \sum_{i=2m+1}^{N} \bar{\delta}_{mi} I_{[Y_i < Y_{(r)}]}}.$$

The corresponding expectation, denoted by ϕ_{AN}, is the EAP to treatment A. Further, $E\left\{r + \sum_{i=2m+1}^{N} \delta_{mi} I_{[X_i < X_{(r)}]}\right\}$ gives the expected number of actual events for treatment A (denoted by ENA) and we also include it in addition to EAP.

Next to explore the small sample behaviour of the performance measures, we conduct a simulation study with 25,000 repetitions. For the simulation, we take different combinations of trial size N, the design parameters (r, m) and response parameters (λ_A, λ_B). However, for better understanding of the simulation output, we always consider treatment A to be better and fixing λ_B at unity, vary λ_A over 1.0 to 2.4 at an interval of 0.2. Although we get a lot from the simulation study, report only the EAP and ENA to treatment A (i.e. the better treatment) for $N = 60, 90$ and 120 in Tables 1,2 and 3. Moreover, we provide few comparative plots (see, Figures 1,2,3 and 4) of the empirical power for our test with adaptive time censoring and that under the equal allocation in the second stage for different combinations of (N, m, r).

TABLE 1: EAP and ENA to treatment A for the proposed allocation with $N = 60$.

	$m = 10$		$m = 20$	
$\Delta = \lambda_A - \lambda_B$	$r = 4$	$r = 6$	$r = 8$	$r = 16$
0.0	**.50 (.23)**	**.50 (.22)**	**.50 (.20)**	**.50 (.18)**
	16	**24**	**21**	**29**
0.2	**.55 (.22)**	**.56 (.21)**	**.56 (.20)**	**.58 (.17)**
	18	**27**	**23**	**31**
0.4	**.58 (.20)**	**.61 (.20)**	**.60 (.19)**	**.64 (.16)**
	19	**29**	**24**	**32**
0.6	**.62(.18)**	**.64 (.19)**	**.63 (.18)**	**.69 (.14)**
	20	**31**	**26**	**34**
0.8	**.64 (.18)**	**.67 (.18)**	**.66 (.17)**	**.72 (.12)**
	21	**32**	**27**	**35**
1.0	**.66 (.18)**	**.70 (.18)**	**.68 (.15)**	**.74 (.09)**
	22	**33**	**28**	**36**
1.2	**.68 (.18)**	**.72 (.17)**	**.70 (.15)**	**.76 (.08)**
	23	**34**	**28**	**37**
1.4	**.70 (.17)**	**.74(.17)**	**.71 (.13)**	**.77 (.06)**
	23	**35**	**29**	**39**

Figures within braces indicate the corresponding standard deviations. Second figures in each cell gives ENA (in nearest integers).

TABLE 2: EAP and ENA to treatment A for the proposed allocation with $N = 90$.

	$m = 15$		$m = 30$	
$\Delta = \lambda_A - \lambda_B$	$r = 6$	$r = 9$	$r = 12$	$r = 18$
0.0	**.50 (.24)**	**.50 (.23)**	**.50(.19)**	**.50 (.19)**
	25	**37**	**31**	**38**
0.2	**.56 (.23)**	**.57(.23)**	**.57 (.19)**	**.58 (.18)**
	28	**42**	**35**	**44**
0.4	**.60 (.24)**	**.63 (.22)**	**.62 (.18)**	**.64 (.16)**
	30	**46**	**38**	**48**
0.6	**.64 (.23)**	**.68 (.20)**	**.66 (.16)**	**.68 (.15)**
	32	**49**	**41**	**52**
0.8	**.67 (.22)**	**.71 (.19)**	**.69 (.14)**	**.71 (.13)**
	33	**51**	**42**	**54**
1.0	**.70 (.21)**	**.74 (.17)**	**.72 (.12)**	**.73(.11)**
	34	**53**	**43**	**55**
1.2	**.72 (.19)**	**.76 (.16)**	**.74 (.11)**	**.75 (.09)**
	35	**55**	**44**	**56**
1.4	**.73 (.18)**	**.78 (.14)**	**.76 (.10)**	**.76 (.08)**
	36	**56**	**45**	**57**

Figures within braces indicate the corresponding standard deviations. Second figures in each cell gives ENA (in nearest integers).

TABLE 3: EAP and ENA to treatment A for the proposed allocation with $N = 120$.

	$m = 20$		$m = 40$	
$\Delta = \lambda_A - \lambda_B$	$r = 8$	$r = 12$	$r = 16$	$r = 24$
0.0	**.50 (.23)**	**.50 (.23)**	**.50 (.19)**	**.50 (.18)**
	33	**49**	**41**	**61**
0.2	**.56 (.24)**	**.59 (.22)**	**.58 (.18)**	**.59 (.17)**
	37	**57**	**48**	**73**
0.4	**.62 (.23)**	**.65 (.21)**	**.64 (.16)**	**.67 (.15)**
	41	**63**	**52**	**81**
0.6	**.66 (.22)**	**.70(.19)**	**.69 (.15)**	**.72 (.12)**
	44	**68**	**56**	**86**
0.8	**.70 (.20)**	**.74(.17)**	**.72 (.12)**	**.75 (.09)**
	46	**71**	**58**	**89**
1.0	**.73 (.19)**	**.77 (.15)**	**.74(.10)**	**.77 (.07)**
	47	**74**	**60**	**91**
1.2	**.75 (.17)**	**.79 (.13)**	**.75 (.08)**	**.79 (.06)**
	49	**76**	**61**	**92**
1.4	**.77 (.16)**	**.81 (.11)**	**.76 (.08)**	**.80 (.05)**
	50	**77**	**62**	**93**

Figures within braces indicate the corresponding standard deviations. Second figures in each cell gives ENA (in nearest integers).

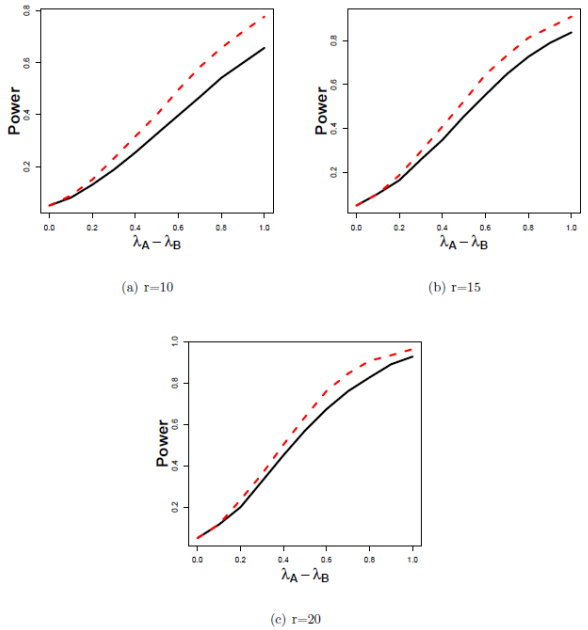
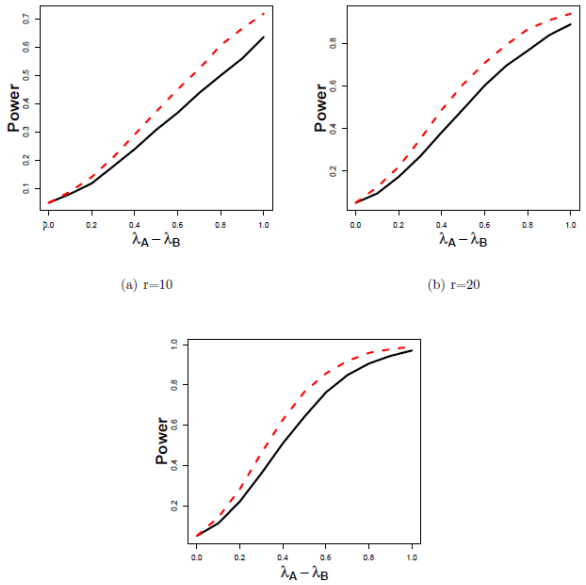


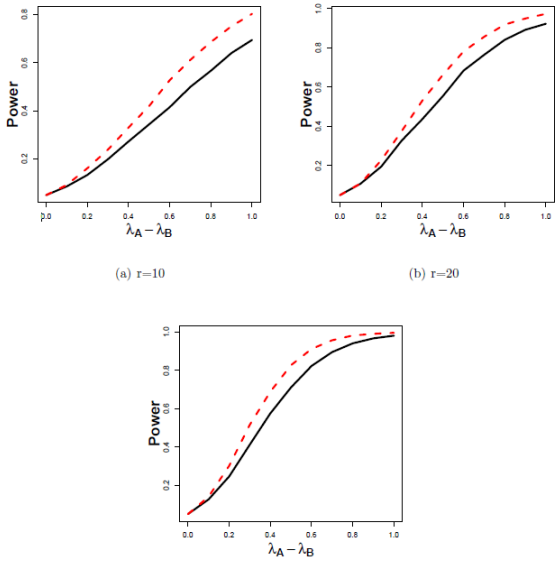
FIGURE 1: Empirical Power comparison between the proposed (solid line) and equal(dotted line) allocation designs for $N = 100$ and $m = 30$



(a) r=10 (b) r=20

(c) r=30

FIGURE 2: Empirical Power comparison between the proposed (solid line) and equal(dotted line) allocation designs for $N = 100$ and $m = 40$.



(a) r=10 (b) r=20

(c) r=30

FIGURE 3: Empirical Power comparison between the proposed (solid line) and equal(dotted line) allocation designs for $N = 150$ and $m = 40$.

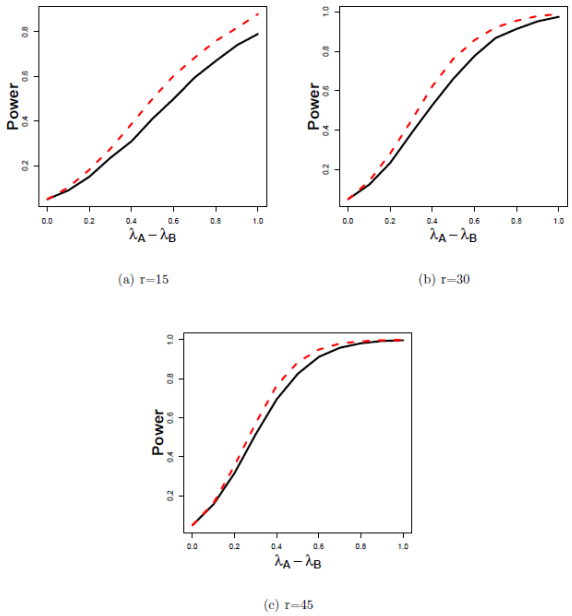


FIGURE 4: Empirical Power comparison between the proposed (solid line) and equal(dotted line) allocation designs for $N = 150$ and $m = 60$.

Discussion: The figures of Tables 1,2,3 and 4 reveal that irrespective of the choices of (r, m) and (λ_A, λ_B), the adaptive allocation, on an average, assigns more subjects to the "winner" of the first stage and makes the allocation unbalanced. This results a loss in statistical power, which is observed in Figures 1-4. However, power is seen to increase with an increase in r for fixed m. This can be best viewed if we consider any of Figures 1-4 and observe the variation in power for increasing r (i.e. figures indicated by (a),(b) and (c))with fixed m. The behaviour is quite natural because the second stage allocation probability depends only on r and hence increase in r amounts to increased first stage allocation, that is, decrease in the adaptive allocation of the second stage. The decrease in the second stage allocation causes most of the subjects to assign equally to either treatment and hence results in an increased power.

Although loss in power weakens the discriminating ability between the treatments but the usefulness of the proposed design lies in assigning the better performing treatment more frequently. Now, if we examine the ϕ_{AN} values, we find that $\phi_{AN} = \frac{1}{2}$ exactly for the equal allocation whereas for the adaptive allocation, ϕ_{AN} approaches $1 - \theta$ or θ as $\Delta = \lambda_A - \lambda_B > 0$ or < 0 for large N, where $m = \lfloor N\theta \rfloor$. Thus the proposed adaptive procedure assigns according to the treatment effectiveness Δ in the limit. The small sample figures are also in agreement with the limiting behaviour, where the higher proportion of allocation always correspond to treatment A, that is, the better treatment. Moreover, such a proportion consistently increase from 50% with the increase in Δ. Thus the proposed adaptive two stage censored procedure has the ability to assign patients

to the better treatment frequently but with a compromise in statistical precision. Censoring number r plays the crucial role in the proposed allocation. The effect of r in the allocation is two fold, namely, limiting the number of events in the first stage as well as controlling the duration of the second stage. Thus, a sensible choice of r could help the experimenter to achieve the goals of the concerned study. We have already observed that increase in r increases both EAP (and ENA) and power. However, a larger choice of r implies higher number of events in the trial, on an average and hence seems reasonable. But such a choice of r increases the duration of the trial and hence conflicts with the objective of censoring. Therefore, a larger choice is not reasonable despite the ability to produce promising performance measures. Consequently, motivated by the figures of Tables 1-3, as a compromise, we suggest to choose r or equivalently ρ in the range $(.60, .70)$.

4. Concluding Remarks

The current work uses a randomization probability based only on ethics. Naturally, other aspects like optimality(in some sense) is not taken into account for the development. Moreover, exponential responses are extensively used for our purpose though some other distributions such as Weibull or Gamma can be used. However, the development with such responses is not just a straightforward extension of the present formulation and requires a fresh look. The subsequent development incorporating optimality in addition to ethics considering a wide class of responses is, therefore, a topic for further investigation.

Acknowledgements

The authors are grateful to the Editor, Associate Editor and the anonymous reviewers for their helpful suggestions, which improved the quality of the submitted work.

[Received: April 2018 — Accepted: May 2019]

References

Bandyopadhyay, U. & Bhattacharya, R. (2001), 'Estimation in a two stage procedure for failure censored model', *Calcutta Statistical Association Bulletin* **51**, 179–190.

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.

Bhattacharya, B. (2007), 'Testing for ordered failure rates under general progressive censoring', *Journal of Statistical Planning and Inference* **137**, 1775–1786.

Biswas, A., Bhattacharya, R. & Park, E. (2016), 'On a class of optimal covariate-adjusted response adaptive designs for survival outcomes', *Statistical Methods in Medical Research* **25**, 2444–2456.

Coad, D. S. (1992), 'Some results on estimation for two stage clinical trials', *Sequential Analysis* **11**, 299–311.

Kalbfleisch, J. & Prentice, R. L. (1980), *The Statistical Analysis of Failure Time Data*, John Wiley & Sons Inc, New York.

Rosenberger, W. F. & Lachin, J. (2015), *Randomization in Clinical Trials: Theory and Practice, Second Edition*, John Wiley & Sons Inc, New York.

Sverdlov, O., Ryeznik, Y. & Wong, W. K. (2014), 'Efficient and ethical response adaptive randomization designs for multi-arm clinical trials with Weibull time-to-event outcomes', *Journal of Biopharmaceutical Statistics* **24**, 732–754.

Sverdlov, O., Tymofyeyev, Y. & Wong, W. K. (2011), 'Optimal response adaptive randomized designs for multi-armed survival trials', *Statistics in Medicine* **30**, 2890–2910.

Xiong, C., Yan, Y. & Ming, J. (2003), 'Sample sizes for comparing means of two lifetime intervention study', *Controlled Clinical Trials* **24**, 283?293.

Zhang, L. & Rosenberger, W. F. (2007), 'Response adaptive randomization for survival trials: The parametric approach', *Applied Statistics* **56**, 153–165.

Appendix

Result A1

As $N \to \infty$,

$$\pi_{mi} \to u(\lambda_A - \lambda_B),$$

almost surely, where $u(x) = 1, 0$ or $\frac{1}{2}$ as $x > 0, < 0$ or $= 0$.

Proof. Note that, $\frac{2T_{1k}}{\lambda_k} \sim \chi^2_{2r}, k = A, B$ independently, and hence writing $\lambda = \frac{\lambda_A}{\lambda_B}$ we get

$$\begin{aligned} P(T_{1A} > T_{1B}) &= P(\frac{2T_{1B}}{\lambda_B} \leq \lambda \frac{2T_{1A}}{\lambda_A}) \\ &= \int_0^\infty P(\chi^2_{2r} < \lambda y) f_{\chi^2_{2r}}(y) dy \end{aligned}$$

with $f_{\chi^2_{2r}}(y) = \frac{1}{\Gamma r 2^r} e^{(-y/2)} y^{r-1},\ y > 0$. Since we can express $P(\chi^2_{2r} < \lambda y) = \sum_{i=r}^{\infty} exp(-\lambda y/2)\frac{(\lambda y)^i}{2^i(i)!}$, we have the right hand number of the above as

$$\sum_{i=r}^{\infty} \int_0^{\infty} \frac{e^{(-\lambda y/2)}(\lambda y)^i}{2^i(i)!} \frac{e^{(-y/2)} y^{r-1}}{2^r(r-1)!} dy = \sum_{i=r}^{\infty} \frac{(r+i-1)!}{(r-1)!i!} \left(1 - \frac{1}{1+\lambda}\right)^i \left(\frac{1}{1+\lambda}\right)^r$$
$$= P(Z \geq r),$$

where Z has a negative binomial, $NB(r,p)$, distribution with $p = \frac{1}{1+\lambda} = \frac{\lambda_B}{\lambda_A+\lambda_B}$. Then, for fixed $p \in (0,1)$, as $r \to \infty$,

$$\frac{Z}{r} \to \frac{1-p}{p}$$

almost surely, where $\frac{1-p}{p} > 1$ or < 1 as $p < \frac{1}{2}$ or $> \frac{1}{2}$ or equivalently as $\lambda_A >$ or $< \lambda_B$. Moreover, as $r \to \infty$,

$$\frac{Z-r}{\sqrt{2r}} \to N(0,1)$$

in distribution for $p = \frac{1}{2}$. Thus, combining all these, we get $\lim_{r\to\infty} P(Z \geq r) =$ 1 or $\frac{1}{2}$ or 0 as $\lambda_A >$ or $=$ or $< \lambda_B$.

Now, by definition, $\pi_{mi} = P(Z \geq r | Z \sim NB(r, \widehat{p}))$, where $\widehat{p}$ is the same as p with λ_k replaced by its maximum likelihood estimator $\widehat{\lambda}_k = \frac{T_{1k}}{r}$ based on the first stage. Since $\widehat{p}$ is a strongly consistent estimator of p, the required result follows. □

Next we prove the following lemmas, which will be useful in deriving Result A2.

Lemma A1. *If X has an exponential distribution with mean μ, then for any $k \geq 1$ and positive constant a,*

$$E(X^k I_{[X \leq a]}) = \Gamma(k+1)\mu^k \left\{ p - (1-p) \sum_{j=1}^{k} \frac{(\frac{a}{\mu})^j}{j!} \right\},$$

where

$$p = P(X \leq a).$$

Proof. By definition,

$$E(X^k I_{[X \leq a]}) = \int_0^a e^{-\frac{x}{\mu}} \frac{x^k}{\mu}\, dx$$
$$= \Gamma(k+1)\mu^k \int_0^{\frac{a}{\mu}} e^{-y} \frac{y^k}{\Gamma(k+1)}\, dy$$
$$= \Gamma(k+1)\mu^k \left\{ 1 - \sum_{j=0}^{k} e^{-\frac{a}{\mu}} \frac{(\frac{a}{\mu})^j}{j!} \right\},$$

which gives the required result. □

Lemma A2. *Let $X_1, X_2, \ldots, X_n$ be iid as exponential with mean μ. Define*

$$\delta_i = I_{[X_i \le a]}, T^* = \sum_{i=1}^{n} \min(X_i, a), \text{and } n^* = \sum_{i=1}^{n} \delta_i.$$

Then

$$Var(T^* - \mu n^*) = np\mu^2.$$

Proof. Since, for each i,

$$\min(X_i, a) = \delta_i X_i + (1 - \delta_i)a,$$

writing

$$T^* - n^*\mu = \sum_{i=1}^{n} \delta_i (X_i - \mu) + a \sum_{i=1}^{n} (1 - \delta_i),$$

we obtain that

$$Var(T^* - n^*\mu) =$$
$$nVar\left\{\delta_1(X_1 - \mu)\right\} - 2anE(1 - \delta_1)E(\delta_1 X_1 - \mu\delta_1) + na^2 Var(1 - \delta_1). \qquad (*)$$

Now the representation

$$Var\left\{\delta_1(X_1 - \mu)\right\} = E(\delta_1 X_1^2) - 2\mu E(\delta_1) + \mu^2 E(\delta_1^2)$$

together with Lemma 1 gives

$$Var\left\{\delta_1(X_1 - \mu)\right\} = p\mu^2 - (1 - p)(2 - p)a^2.$$

Combining all these together with $Var(1-\delta_1) = p(1-p)$ in (*), we get the required result. □

Result A2

For any real x,

$$\lim_{N\to\infty} P_{H_0}(-2\ln\Lambda_N \le x) = \frac{1}{2} + \frac{1}{2}\int_0^x f_{\chi_1^2}(u)du.$$

Proof. Denoting $\widehat{\Delta}_N = \widehat{\lambda}_{AN} - \widehat{\lambda}_{BN}$, consider the representation

$$P_{H_0}\left(-2\ln\Lambda_N \le x\right) =$$
$$P_{H_0}\left(-2\ln\Lambda_N \le x, \widehat{\Delta}_N \le 0\right) + P_{H_0}\left(-2\ln\Lambda_N \le x, \widehat{\Delta}_N > 0\right).$$

Then it follows that

$$\lim_{N\to\infty} P_{H_0}\left(-2\ln\Lambda_N \le x\right) = \lim_{N\to\infty} P_{H_0}\left\{\sqrt{N}\widehat{\Delta}_N \le 0\right\} \text{ if } x \le 0$$

and

$$\lim_{N\to\infty} P_{H_0}\left(-2\ln\Lambda_N \le x\right) = \lim_{N\to\infty} P_{H_0}\left(\sqrt{N}\widehat{\Delta}_N \le 0\right) + \lim_{N\to\infty} P_{H_0}\left(-2\ln\Lambda_N \le x, \sqrt{N}\widehat{\Delta}_N > 0\right) \text{ if } x > 0.$$

Define the random variables $U_{kN} = \sqrt{N}\left(\frac{T_{1k}-r\lambda_0}{r}\right)$ and $V_{kN} = \sqrt{N}\left(\frac{T^*_{2k}-n_k\lambda_0}{N-2m}\right), k = A, B$, where λ_0 is the common unspecified value under the null hypothesis. Then a Taylor series expansion up to second order terms expresses $-2\ln\Lambda_N$ in terms of these variables as

$$\begin{aligned}-2\ln\Lambda_N = {} & \frac{r+n_A}{N\lambda_0^2}\left(\frac{r}{r+n_A}U_{AN} + \frac{N-2m}{r+n_A}V_{AN}\right)^2 \\ & + \frac{r+n_B}{N\lambda_0^2}\left(\frac{r}{r+n_B}U_{BN} + \frac{N-2m}{r+n_B}V_{BN}\right)^2 \\ & - \frac{2r+n_A+n_B}{N\lambda_0^2}\left\{\frac{r}{2r+n_A+n_B}(U_{AN}+U_{BN}) + \frac{N-2m}{2r+n_A+n_B}(V_{AN}+V_{BN})\right\}^2 \\ & + R_N,\end{aligned}$$

where $R_N \to 0$ in probability. Now, it is easy to observe that, for given the first stage data, n_A is conditionally $Binomial(\sum_{i=2m+1}^{N}\delta_{mi}, P[X_{2m+1} < X_{(r)}])$. Moreover, under H_0, δ_{mi} are iid $Bernoulli(\frac{1}{2})$, so that

$$\frac{\sum_{i=2m+1}^{N}\delta_{mi}}{N-2m} \to \frac{1}{2}$$

almost surely, and hence, as a consequence of the almost sure convergence of $X_{(r)} = X_{(\lfloor m\rho \rfloor)}$ to $-\lambda_A log(1-\rho)$, we get

$$\frac{n_A}{N-2m} \to \frac{\rho}{2}$$

almost surely. Then it follows that for any $k = A, B$,

$$\frac{n_k}{m} \to \rho\frac{1-2\theta}{2\theta} \tag{A1}$$

almost surely, $k = A, B$. Hence, under the null hypothesis, $-2\ln\Lambda_N$ and W_N^2, with

$$W_N = \frac{\sqrt{\rho}}{\lambda_0}\left\{\theta(U_{AN}-U_{BN}) + \frac{1-2\theta}{\rho}(V_{AN}-V_{BN})\right\},$$

have the same asymptotic distribution. But, given the first stage observations, W_N is the sum of independent random variables with conditional expectation

$$E(W_N) = \frac{\theta\sqrt{\rho}}{\lambda_0}(U_{AN}-U_{BN})$$

and conditional variance (see, Lemma 2, for details)

$$Var(W_N) = \frac{(1-2\theta)^2}{\rho}\left\{\frac{N\sum_{i=2m+1}^{N}\delta_{mi}}{(N-2m)^2}(1-e^{-\frac{X_{(r)}}{\lambda_0}}) + \frac{N\sum_{i=2m+1}^{N}\bar{\delta}_{mi}}{(N-2m)^2}(1-e^{-\frac{Y_{(r)}}{\lambda_0}})\right\}.$$

Since δ_{mi} are iid $Bernoulli(\frac{1}{2})$ under the null hypothesis, we have

$$\frac{\sum_{i=2m+1}^{N}\delta_{mi}}{N-2m} \to \frac{1}{2}$$

almost surely, and hence as $N \to \infty$

$$Var(W_N) \to (1-2\theta)$$

in probability. Again, U_{AN} and U_{BN} are independent and asymptotically $N\left(0, \frac{\lambda_0^2}{\rho\theta}\right)$ and consequently, as $N \to \infty$

$$\frac{\theta\sqrt{\rho}}{\lambda_0}(U_{AN} - U_{BN}) \to N(0, 2\theta)$$

in distribution. Hence, following Hajek et al.(1999, pp. 241-242), we get, unconditionally,

$$W_N \to N(0,1)$$

in distribution as $N \to \infty$. Further, we can express $\sqrt{N}\widehat{\Delta}_N$ as

$$\sqrt{N}\widehat{\Delta}_N = \frac{r}{r+n_A}U_{AN} - \frac{r}{r+n_B}U_{BN} + \frac{\sqrt{N-2m}}{r+n_A}V_{AN} - \frac{\sqrt{N-2m}}{r+n_B}V_{BN}. \quad \text{(A2)}$$

Then, under (A1), the right hand side of (A2) and $2\theta(U_{AN}-U_{BN})+\frac{2(1-2\theta)}{\rho}(V_{AN}-V_{BN})$ have the same asymptotic distribution. Since the latter quantity is nothing but $\frac{2\lambda_0 W_N}{\sqrt{\rho}}$ and W_N is asymptotically standard normal under the null hypothesis, we get

$$\lim_{N\to\infty} P_{H_0}\left(\sqrt{N}\widehat{\Delta}_N \le 0\right) = \frac{1}{2}$$

and

$$\lim_{N\to\infty} P_{H_0}\left(-2\ln\Lambda_N \le x, \sqrt{N}\widehat{\Delta}_N > 0\right) = \frac{1}{2}P(\chi_1^2 \le x),$$

and hence the result follows. □