International Journal of Statistical Sciences *ISSN 1683–5603*
Vol. 24(2) Special, December, 2024, pp 67-79
DOI: https://doi.org/10.3329/ijss.v24i20.78215

Effect of Misspecification on a Multi Treatment Two Stage Adaptive Design for Survival Outcomes

Soumyadeep Das[1*] and Rahul Bhattacharya[2]

[1]Department of Statistics, Bidhannagar College,
EB-2, Sector 1, Salt Lake City, Kolkata – 700 064, India
[2]Department of Statistics, University of Calcutta,
35 Ballygunge Circular Road, Kolkata – 700 019, India
Email: rahul_bhattya@yahoo.com
[*]Correspondence should be addressed to Soumyadeep Das
(Email: das.soumyadeep1992@gmail.com)

[Received June 30, 2024; Accepted December 1, 2024]

Abstract

Multi-treatment two stage adaptive design for survival responses is generally developed under different assumptions. In this work, we explore the performance of such a design when the assumptions are violated. As a choice of design, we consider a specific design of Bhattacharya and Shome (2019), which uses random censoring, exponential response, Koziol-Green model (Koziol and Green, 1976), etc. Several ethical and inferential criteria of the design are studied under model misspecification for different parameter configurations as well as for a data arising from a real clinical trial.

Keywords: Clinical Trials, Two Stage Allocation, Random Censoring, Model misspecification.

AMS Classification: 62N02.

1. Introduction

Clinical trial generally deals with new drugs/therapies tested on human patients by studying its relative performance in terms of several parameters over the other competing treatments. In the era of adaptive randomization, allocation designs which can skew the allocation to the better performing treatment satisfying ethical norms

of a trial, have become popular and subject to study over several decades. Two stage design is one of such data dependant adaptive procedure for the purpose of skewing the allocation to the treatment doing better, where the first stage uses complete randomization (CR), which allocates the available treatments with equal probabilities to the patients. After an interim analysis, second stage allocation probabilities are determined with an objective to assign more patients to the better performing treatment of the first stage and the second stage patients are allocated with those probabilities. Several two stage design are discussed in the works of Colton (1963) and Coad (1992, 1994). As their designs used deterministic procedures for the second stage patients, selection bias is natural (Matthews, 2006). Consequently, Bandyopadhyay and Bhattacharya (2007) developed an ethical allocation incorporating randomization at both stages.

Sometimes, the responses of the patients are time to event data (e.g. the lifetimes of cancer patients in terms of progression free survival (PFS) or overall survival (OS)). However, in a clinical trial with survival outcome, it is not always possible to continue the study until all patient respond as survival times may vary from time to fatal event (i.e. death, relapse or remission). Naturally, censoring is common in this situation. But survival data analysis becomes difficult by the presence of censoring and hence, to update the allocation strategy after each response is not feasible in practice. Therefore, a purely sequential adaptive procedure like response adaptive design (Rosenberger and Lachin, 2015) can not be applied in a simple way and it is better to update the allocation strategy after a group of responses are observed. Thus, a two stage design is a better alternative to carry out the analysis with censored observations. Several two stage adaptive designs have been studied in the literature in the recent past. One of them is a multi-treatment two stage design by Bhattacharya and Shome (2019). They proposed a two stage procedure where the second stage allocation probabilities are determined by *asymptotic p values* (Silvapulle and Sen, 2005) of a suitable score test for testing the presence of a unique superior treatment based on the accrued data from the first stage patients. Although the procedure is well studied, several rigid assumptions were deployed to develop the design. First of all, they assumed that the lifetime and censoring variables both follow exponential distributions. Secondly, they used Koziol-Green model (Koziol and Green, 1976), where not only the lifetime and censoring variables are independent but also the log of their survival functions are proportional. Moreover, they assumed the presence of a unique superior treatment. But in reality, neither of these assumptions may be valid and hence it is of interest to study the behaviour of this design under the violation of the assumptions with respect to relevant operating characteristics. Specifically, in the present paper, we consider some lifetime distributions like Gamma, Weibull and Lognormal and then use their method as if the distributions are misspecified as Exponential. Moreover, we introduce dependent exponential responses and apply the two stage design as if they are independent. Lastly, we allow some configurations where there is no unique superior treatment and observe how such a two stage design performs. Additionally,

we redesign the same real clinical trial (i.e. recurrent glioblastoma trial of Batchelor et al. (2013) as considered in Bhattacharya and Shome (2019) and study the deviations under model misspecification.

After a brief discussion of the two stage design of Bhattacharya and Shome (2019) in section 2, we evaluate the performance of the design empirically under model violations in section 3. In section 4, we discuss how model misspecification affects the performance of a real clinical trial and lastly, we conclude the work with a relevant discussion in section 5.

2. The two stage allocation design of Bhattacharya and Shome (2019)

Consider a clinical trial involving $t(\geq 3)$ treatments and N prefixed subjects where $tm(\leq N)$ subjects are assigned to t treatments, m to each treatment, in the first stage. Depending on the results of the first stage, a set of allocation probabilities $(\rho_{1m}, \rho_{2m}, \ldots, \rho_{tm})$ satisfying $\sum_{k=1}^{t} \rho_{km} = 1$ is derived for the allocation of second stage patients.

First, denoting the survival (censoring) time corresponding to the i th patient when given treatment k by X_{ki} (C_{ki}) and defining the censoring indicator $I_{ki} = I(X_{ki} < C_{ki})$, $k = 1, 2, .., t$, $i = 1, 2, \ldots$, the observations are obtained as (Y_{ki}, I_{ki}), where $Y_{ki} = \min(X_{ki}, C_{ki})$. Moreover, denoting $f_k(g_k)$ as the density function of the survival (censoring) distribution and $\bar{F}_k$ ($\bar{G}_k$) as the corresponding survival function corresponding to treatment k, they assumed the exponential distributions under Koziol-Green model (Koziol and Green, 1976) as $\bar{F}_k(t) = \exp\left(-\frac{t}{\mu_k}\right)$, $\bar{G}_k(t) = \left(\bar{F}_k(t)\right)^{\nu_k}$, $k = 1, 2, \ldots, t$ under the independence of X_{ki} and C_{ki} for each i and k.

Secondly, for the derivation of second stage allocation probabilities, the framework of multiple comparisons with the best (Hsu, 1996) is implemented. In particular, taking treatment k as the best, the following statistical hypotheses is considered: $H_{0k} : \mu_k \leq \max_{r \neq k} \mu_r$ against $H_{1k} : \mu_k > \max_{r \neq k} \mu_r$, where μ_r is the treatment effect measure for treatment r, $r = 1, 2, \ldots, t$. As it is easy to observe that H_{0k} is equivalent to the union of a number of sub hypotheses $H_{0kr} : \mu_k \leq \mu_r$ and similarly, H_{1k} can be expressed as the intersection of the sub hypotheses $H_{1kr} : \mu_k > \mu_r$, the global null hypothesis H_{0k} is rejected if all the tests for H_{0kr} against $H_{1kr}, k = 1, 2, .., t$ are rejected. Then the corresponding *asymptotic p values* (Silvapulle and Sen, 2005) of relevant score tests for testing directional hypotheses are obtained as p_{kr}. As, lower value of p_{kr} indicates higher evidence of superiority of treatment k over treatment r based on the first stage data, $q_{kr} = 1 - p_{kr}$ was defined as evidence of superiority of treatment k over treatment r. Following the notions of intersection union tests of hypothesis (Berger and Hsu, 1996), an evidence measure of the superiority of treatment k over all the treatments is simply $\pi_{km} = \min_{r \neq k} q_{kr}$, where the higher the value of π_{km}, the higher is the superiority of treatment k among others. Combining all these,

they suggested to assign any incoming subject of second stage to treatment k with probability $\rho_{km} = \frac{\pi_{km}}{\pi_{1m}+\pi_{2m}+\ldots+\pi_{tm}}$.
Moreover, denoting N_k as the number of patients assigned treatment k, Bhattacharya and Shome (2019) also proved that, as $m, N \to \infty$, $m/N \to \theta$ with $\theta \in (0, 1/t)$, the observed allocation proportion to treatment k (i.e. $\frac{N_k}{N}$) approaches $(1-\theta)$ or θ in probability as $\mu_k > \max_{r\neq k} \mu_r$ or $\mu_k < \max_{r\neq k} \mu_r$. Hence, the observed allocation proportion to treatment k behaves ethically in the limit, whenever $\mu_k > \max_{r\neq k} \mu_r$, that is, when treatment k is the unique superior.

3. Performance Evaluation under misspecification

Beside stating the asymptotic properties of their design, Bhattacharya and Shome (2019) also performed some small sample studies with $t = 3$ treatments under the assumptions, they used for developments. As operating characteristics, they considered

1. Type I error rates
2. The distribution of expected allocation proportion (EAP) to different treatments along with the standard deviations and
3. The power of a test of the superiority of treatment 1 considering the hypothesis $H_0 : \mu_1 \leq \max_{k=2,3} \mu_k$ against $H_1 : \mu_1 > \max_{k=2,3} \mu_k$ based on the data from two stages.

To study the design after relaxing the assumptions, we consider the following aspects:

1. The density f_k of data generating process is different from exponential. Specifically, we use Gamma, Weibull and Lognormal as the alternative lifetime distributions.
2. X_{ki} and C_{ki} are dependent with a structure described through Clayton copula (Clayton, 1978) in data generating process; Kendall's τ gives the strength of association.
3. Existence of more than one superior treatment.

Under these model violating situations, we consider several configurations to study how the operating characteristics vary from the original cases. The parameters of the distributions are so taken that the expected values for lifetime as well as censoring random variables remain same for head to head comparison. After a simulation with 20,000 repetitions of the relevant procedure for $m = 30$ and $N = 150$, we report the results in Tables 1-3. On the other hand, the comparison is provided through boxplots and power curves (Figures 1 and 2, respectively) in case of $m = 40$ and $N = 200$ for visual representations. Moreover, in practical situations, there may present more than one superior treatment. But Bhattacharya and Shome (2019) developed their design assuming only one superior treatment. So, here we allow more than one superior treatment and study the effect of it under different values of m, n, ν_i's and μ_i's in Table 4.

Table 1: Operating characteristics for $\nu_1 = \nu_2 = \nu_3 = 1$ and $\mu_2 = \mu_3 = 2$ with varying μ_1 for $m = 30$ and $N = 150$.

μ_1	Model	Trt-1	Trt-2	Trt-3	Power
	Exponential (mean=μ_k)	0.340 (0.170)	0.330 (0.170)	0.320 (0.170)	0.050
	Gamma (shape=3, scale=$\mu_k/3$)	0.340 (0.166)	0.341 (0.167)	0.320 (0.166)	0.050
	Weibull (shape=3, scale=$\mu_k/0.893$)	0.340 (0.166)	0.340 (0.166)	0.320 (0.166)	0.050
2.0	Lognormal (meanlog=log(μ_k), sdlog=3)	0.340 (0.169)	0.341 (0.170)	0.319 (0.169)	0.050
	Associated Exponential ($\tau = 0.091$)	0.341 (0.167)	0.340 (0.167)	0.320 (0.167)	0.050
	Associated Exponential ($\tau = 0.500$)	0.339 (0.167)	0.342 (0.168)	0.319 (0.167)	0.050
	Associated Exponential ($\tau = 0.909$)	0.339 (0.167)	0.341 (0.167)	0.321 (0.167)	0.050
	Exponential (mean=μ_k)	0.450 (0.170)	0.290 (0.140)	0.270 (0.140)	0.250
	Gamma (shape=3, scale=$\mu_k/3$)	0.472 (0.162)	0.273 (0.125)	0.254 (0.126)	0.345
	Weibull (shape=3, scale=$\mu_k/0.893$)	0.480 (0.158)	0.271 (0.122)	0.249 (0.121)	0.357
2.6	Lognormal (meanlog=log(μ_k), sdlog=3)	0.371 (0.178)	0.324 (0.163)	0.305 (0.163)	0.089
	Associated Exponential ($\tau = 0.091$)	0.441 (0.173)	0.289 (0.139)	0.270 (0.141)	0.243
	Associated Exponential ($\tau = 0.500$)	0.439 (0.174)	0.292 (0.142)	0.269 (0.140)	0.221
	Associated Exponential ($\tau = 0.909$)	0.445 (0.172)	0.288 (0.139)	0.266 (0.137)	0.247
	Exponential (mean=μ_k)	0.520 (0.140)	0.250 (0.100)	0.230 (0.100)	0.530
	Gamma (shape=3, scale=$\mu_k/3$)	0.541 (0.110)	0.239 (0.079)	0.220 (0.082)	0.688
	Weibull (shape=3, scale=$\mu_k/0.893$)	0.549 (0.099)	0.235 (0.073)	0.216 (0.074)	0.720
3.2	Lognormal (meanlog=log(μ_k), sdlog=3)	0.398 (0.180)	0.312 (0.157)	0.289 (0.155)	0.127
	Associated Exponential ($\tau = 0.091$)	0.513 (0.140)	0.253 (0.104)	0.234 (0.105)	0.510
	Associated Exponential ($\tau = 0.500$)	0.507 (0.144)	0.256 (0.108)	0.237 (0.109)	0.473
	Associated Exponential ($\tau = 0.909$)	0.516 (0.137)	0.252 (0.102)	0.233 (0.102)	0.517
	Exponential (mean=μ_k)	0.550 (0.090)	0.230 (0.070)	0.220 (0.070)	0.740
	Gamma (shape=3, scale=$\mu_k/3$)	0.568 (0.067)	0.226 (0.051)	0.206 (0.052)	0.874
	Weibull (shape=3, scale=$\mu_k/0.893$)	0.571 (0.060)	0.224 (0.047)	0.204 (0.047)	0.894
3.8	Lognormal (meanlog=log(μ_k), sdlog=3)	0.422 (0.179)	0.299 (0.150)	0.278 (0.149)	0.175
	Associated Exponential ($\tau = 0.091$)	0.548 (0.104)	0.236 (0.076)	0.216 (0.077)	0.710
	Associated Exponential ($\tau = 0.500$)	0.543 (0.111)	0.238 (0.082)	0.219 (0.082)	0.675
	Associated Exponential ($\tau = 0.909$)	0.551 (0.101)	0.235 (0.074)	0.215 (0.074)	0.724
	Exponential (mean=μ_k)	0.570 (0.070)	0.230 (0.060)	0.210 (0.050)	0.860
	Gamma (shape=3, scale=$\mu_k/3$)	0.577 (0.046)	0.222 (0.038)	0.202 (0.038)	0.948
	Weibull (shape=3, scale=$\mu_k/0.893$)	0.578 (0.043)	0.222 (0.037)	0.201 (0.036)	0.956
4.4	Lognormal (meanlog=log(μ_k), sdlog=3)	0.440 (0.177)	0.291 (0.144)	0.270 (0.143)	0.219
	Associated Exponential ($\tau = 0.091$)	0.566 (0.076)	0.227 (0.056)	0.207 (0.058)	0.837
	Associated Exponential ($\tau = 0.500$)	0.562 (0.083)	0.229 (0.062)	0.209 (0.062)	0.807
	Associated Exponential ($\tau = 0.909$)	0.566 (0.074)	0.227 (0.055)	0.207 (0.057)	0.851

SD of each EAP is inside the braces. Power under equal μ_k's denotes the type I error rate.

Table 2: Operating characteristics for $\nu_1 = 2, \nu_2 = \nu_3 = 1$ and $\mu_2 = \mu_3 = 2$ with varying μ_1 for $m = 30$ and $N = 150$.

μ_1	Model	Trt-1	Trt-2	Trt-3	Power
	Exponential (mean=μ_k)	0.340 (0.170)	0.330 (0.170)	0.320 (0.170)	0.050
	Gamma (shape=3, scale=$\mu_k/3$)	0.428 (0.175)	0.297 (0.145)	0.275 (0.144)	0.050
	Weibull (shape=3, scale=$\mu_k/0.893$)	0.461 (0.167)	0.279 (0.131)	0.260 (0.131)	0.050
2.0	Lognormal (meanlog=log(μ_k), sdlog=3)	0.279 (0.130)	0.370 (0.176)	0.351 (0.176)	0.050
	Associated Exponential ($\tau = 0.091$)	0.353 (0.171)	0.332 (0.165)	0.315 (0.165)	0.050
	Associated Exponential ($\tau = 0.500$)	0.448 (0.173)	0.287 (0.139)	0.265 (0.138)	0.050
	Associated Exponential ($\tau = 0.909$)	0.577 (0.051)	0.222 (0.041)	0.201 (0.040)	0.050
	Exponential (mean=μ_k)	0.440 (0.170)	0.290 (0.140)	0.270 (0.140)	0.220
	Gamma (shape=3, scale=$\mu_k/3$)	0.526 (0.128)	0.246 (0.093)	0.227 (0.095)	0.245
	Weibull (shape=3, scale=$\mu_k/0.893$)	0.546 (0.105)	0.238 (0.080)	0.216 (0.076)	0.249
2.6	Lognormal (meanlog=log(μ_k), sdlog=3)	0.297 (0.146)	0.363 (0.175)	0.340 (0.175)	0.087
	Associated Exponential ($\tau = 0.091$)	0.443 (0.173)	0.288 (0.139)	0.269 (0.140)	0.203
	Associated Exponential ($\tau = 0.500$)	0.516 (0.139)	0.253 (0.105)	0.231 (0.102)	0.158
	Associated Exponential ($\tau = 0.909$)	0.580 (0.039)	0.220 (0.035)	0.200 (0.034)	0.074
	Exponential (mean=μ_k)	0.510 (0.150)	0.260 (0.110)	0.240 (0.110)	0.480
	Gamma (shape=3, scale=$\mu_k/3$)	0.565 (0.077)	0.228 (0.057)	0.208 (0.058)	0.502
	Weibull (shape=3, scale=$\mu_k/0.893$)	0.572 (0.058)	0.224 (0.046)	0.204 (0.045)	0.494
3.2	Lognormal (meanlog=log(μ_k), sdlog=3)	0.318 (0.158)	0.352 (0.172)	0.330 (0.171)	0.132
	Associated Exponential ($\tau = 0.091$)	0.507 (0.145)	0.256 (0.108)	0.237 (0.110)	0.425
	Associated Exponential ($\tau = 0.500$)	0.549 (0.103)	0.236 (0.077)	0.215 (0.075)	0.288
	Associated Exponential ($\tau = 0.909$)	0.581 (0.034)	0.219 (0.032)	0.200 (0.032)	0.099
	Exponential (mean=μ_k)	0.540 (0.110)	0.240 (0.080)	0.220 (0.080)	0.690
	Gamma (shape=3, scale=$\mu_k/3$)	0.576 (0.049)	0.222 (0.040)	0.202 (0.040)	0.700
	Weibull (shape=3, scale=$\mu_k/0.893$)	0.579 (0.039)	0.221 (0.034)	0.201 (0.035)	0.687
3.8	Lognormal (meanlog=log(μ_k), sdlog=3)	0.333 (0.165)	0.344 (0.170)	0.323 (0.170)	0.180
	Associated Exponential ($\tau = 0.091$)	0.544 (0.110)	0.237 (0.079)	0.219 (0.083)	0.632
	Associated Exponential ($\tau = 0.500$)	0.565 (0.079)	0.227 (0.058)	0.207 (0.059)	0.421
	Associated Exponential ($\tau = 0.909$)	0.581 (0.033)	0.220 (0.031)	0.199 (0.031)	0.119
	Exponential (mean=μ_k)	0.560 (0.080)	0.230 (0.060)	0.210 (0.050)	0.820
	Gamma (shape=3, scale=$\mu_k/3$)	0.579 (0.038)	0.220 (0.034)	0.201 (0.033)	0.814
	Weibull (shape=3, scale=$\mu_k/0.893$)	0.581 (0.034)	0.220 (0.032)	0.200 (0.033)	0.799
4.4	Lognormal (meanlog=log(μ_k), sdlog=3)	0.352 (0.173)	0.335 (0.167)	0.313 (0.166)	0.234
	Associated Exponential ($\tau = 0.091$)	0.562 (0.084)	0.229 (0.062)	0.210 (0.062)	0.777
	Associated Exponential ($\tau = 0.500$)	0.573 (0.060)	0.223 (0.046)	0.203 (0.047)	0.533
	Associated Exponential ($\tau = 0.909$)	0.582 (0.033)	0.219 (0.031)	0.199 (0.031)	0.140

SD of each EAP is inside the braces. Power under equal μ_k's denotes the type I error rate.

Table 3: Operating characteristics for $\nu_1 = 1, \nu_2 = \nu_3 = 2$ and $\mu_2 = \mu_3 = 2$ with varying μ_1 for $m = 30$ and $N = 150$.

μ_1	Model	Trt-1	Trt-2	Trt-3	Power
	Exponential (mean=μ_k)	0.340 (0.170)	0.340 (0.170)	0.320 (0.170)	0.050
	Gamma (shape=3, scale=$\mu_k/3$)	0.273 (0.124)	0.374 (0.175)	0.353 (0.176)	0.050
	Weibull (shape=3, scale=$\mu_k/0.893$)	0.254 (0.102)	0.382 (0.177)	0.364 (0.177)	0.050
2.0	Lognormal (meanlog=log(μ_k), sdlog=3)	0.415 (0.177)	0.302 (0.147)	0.238 (0.148)	0.050
	Associated Exponential ($\tau = 0.091$)	0.329 (0.163)	0.346 (0.169)	0.326 (0.169)	0.050
	Associated Exponential ($\tau = 0.500$)	0.257 (0.107)	0.383 (0.178)	0.361 (0.178)	0.050
	Associated Exponential ($\tau = 0.909$)	0.224 (0.032)	0.399 (0.181)	0.377 (0.181)	1.000
	Exponential (mean=μ_k)	0.440 (0.170)	0.290 (0.140)	0.270 (0.140)	0.210
	Gamma (shape=3, scale=$\mu_k/3$)	0.370 (0.174)	0.325 (0.160)	0.305 (0.161)	0.298
	Weibull (shape=3, scale=$\mu_k/0.893$)	0.339 (0.166)	0.343 (0.168)	0.318 (0.166)	0.320
2.6	Lognormal (meanlog=log(μ_k), sdlog=3)	0.451 (0.171)	0.286 (0.137)	0.263 (0.134)	0.083
	Associated Exponential ($\tau = 0.091$)	0.420 (0.176)	0.300 (0.147)	0.280 (0.147)	0.200
	Associated Exponential ($\tau = 0.500$)	0.312 (0.155)	0.355 (0.173)	0.333 (0.172	0.187
	Associated Exponential ($\tau = 0.909$)	0.224 (0.035)	0.400 (0.181)	0.376 (0.181)	1.000
	Exponential (mean=μ_k)	0.500 (0.150)	0.260 (0.110)	0.240 (0.110)	0.430
	Gamma (shape=3, scale=$\mu_k/3$)	0.458 (0.167)	0.281 (0.132)	0.261 (0.132)	0.618
	Weibull (shape=3, scale=$\mu_k/0.893$)	0.433 (0.174)	0.294 (0.142)	0.273 (0.141)	0.664
3.2	Lognormal (meanlog=log(μ_k), sdlog=3)	0.479 (0.161)	0.270 (0.122)	0.252 (0.124)	0.121
	Associated Exponential ($\tau = 0.091$)	0.487 (0.155)	0.266 (0.119)	0.247 (0.119)	0.399
	Associated Exponential ($\tau = 0.500$)	0.368 (0.175)	0.327 (0.163)	0.306 (0.162)	0.371
	Associated Exponential ($\tau = 0.909$)	0.226 (0.043)	0.396 (0.181)	0.378 (0.181)	1.000
	Exponential (mean=μ_k)	0.540 (0.110)	0.240 (0.080)	0.220 (0.080)	0.610
	Gamma (shape=3, scale=$\mu_k/3$)	0.518 (0.133)	0.251 (0.099)	0.231 (0.099)	0.830
	Weibull (shape=3, scale=$\mu_k/0.893$)	0.501 (0.146)	0.260 (0.112)	0.238 (0.109)	0.875
3.8	Lognormal (meanlog=log(μ_k), sdlog=3)	0.499 (0.150)	0.260 (0.113)	0.240 (0.112)	0.150
	Associated Exponential ($\tau = 0.091$)	0.527 (0.126)	0.245 (0.092)	0.227 (0.096)	0.571
	Associated Exponential ($\tau = 0.500$)	0.419 (0.177)	0.302 (0.149)	0.279 (0.147)	0.558
	Associated Exponential ($\tau = 0.909$)	0.230 (0.057)	0.395 (0.181)	0.375 (0.181)	1.000
	Exponential (mean=μ_k)	0.560 (0.090)	0.230 (0.070)	0.210 (0.060)	0.730
	Gamma (shape=3, scale=$\mu_k/3$)	0.548 (0.101)	0.236 (0.075)	0.216 (0.074)	0.927
	Weibull (shape=3, scale=$\mu_k/0.893$)	0.537 (0.114)	0.241 (0.082)	0.222 (0.085)	0.957
4.4	Lognormal (meanlog=log(μ_k), sdlog=3)	0.516 (0.138)	0.253 (0.103)	0.232 (0.102)	0.187
	Associated Exponential ($\tau = 0.091$)	0.551 (0.099)	0.234 (0.072)	0.215 (0.074)	0.702
	Associated Exponential ($\tau = 0.500$)	0.459 (0.167)	0.279 (0.131)	0.261 (0.133)	0.696
	Associated Exponential ($\tau = 0.909$)	0.234 (0.070)	0.393 (0.181)	0.373 (0.180)	1.000

SD of each EAP is inside the braces. Power under equal μ_k's denotes the type I error rate.

Table 4: Operating characteristics for different parameter configurations under the presence of more than one superior treatment.

(ν_1, ν_2, ν_3)	(μ_1, μ_2, μ_3)	Trt-1	Trt-2	Trt-3	Power
		$m = 30, N = 150$			
	(2.6, 2.6, 2.0)	0.378 (0.176)	0.379 (0.177)	0.243 (0.113)	0.074
(1, 1, 1)	(3.2, 3.2, 2.0)	0.392 (0.178)	0.392 (0.178)	0.216 (0.070)	0.091
	(3.8, 3.8, 2.0)	0.396 (0.178)	0.397 (0.179)	0.207 (0.048)	0.108
	(4.4, 4.4, 2.0)	0.399 (0.179)	0.396 (0.179)	0.205 (0.037)	0.103
	(2.6, 2.6, 2.0)	0.375 (0.176)	0.380 (0.176)	0.244 (0.115)	0.075
(2, 1, 1)	(3.2, 3.2, 2.0)	0.388 (0.177)	0.395 (0.178)	0.217 (0.072)	0.081
	(3.8, 3.8, 2.0)	0.390 (0.179)	0.402 (0.178)	0.208 (0.049)	0.088
	(4.4, 4.4, 2.0)	0.392 (0.179)	0.404 (0.178)	0.205 (0.038)	0.089
	(2.6, 2.6, 2.0)	0.378 (0.176)	0.370 (0.176)	0.252 (0.123)	0.069
(1, 2, 2)	(3.2, 3.2, 2.0)	0.392 (0.178)	0.385 (0.177)	0.223 (0.085)	0.083
	(3.8, 3.8, 2.0)	0.400 (0.179)	0.389 (0.178)	0.212 (0.061)	0.095
	(4.4, 4.4, 2.0)	0.401 (0.179)	0.392 (0.179)	0.207 (0.048)	0.098
		$m = 50, N = 200$			
	(2.6, 2.6, 2.0)	0.367 (0.119)	0.368 (0.119)	0.265 (0.069)	0.069
(1, 1, 1)	(3.2, 3.2, 2.0)	0.375 (0.119)	0.376 (0.119)	0.249 (0.039)	0.079
	(3.8, 3.8, 2.0)	0.377 (0.119)	0.376 (0.119)	0.247 (0.031)	0.081
	(4.4, 4.4, 2.0)	0.377 (0.120)	0.376 (0.119)	0.247 (0.030)	0.085
	(2.6, 2.6, 2.0)	0.367 (0.119)	0.368 (0.119)	0.265 (0.069)	0.066
(2, 1, 1)	(3.2, 3.2, 2.0)	0.372 (0.119)	0.378 (0.120)	0.250 (0.040)	0.073
	(3.8, 3.8, 2.0)	0.373 (0.120)	0.380 (0.119)	0.247 (0.032)	0.076
	(4.4, 4.4, 2.0)	0.372 (0.120)	0.381 (0.120)	0.247 (0.030)	0.078
	(2.6, 2.6, 2.0)	0.366 (0.119)	0.363 (0.118)	0.271 (0.076)	0.072
(1, 2, 2)	(3.2, 3.2, 2.0)	0.376 (0.119)	0.371 (0.119)	0.253 (0.047)	0.088
	(3.8, 3.8, 2.0)	0.378 (0.119)	0.374 (0.119)	0.248 (0.035)	0.096
	(4.4, 4.4, 2.0)	0.380 (0.119)	0.373 (0.119)	0.247 (0.031)	0.099

SD of each EAP is inside the braces.

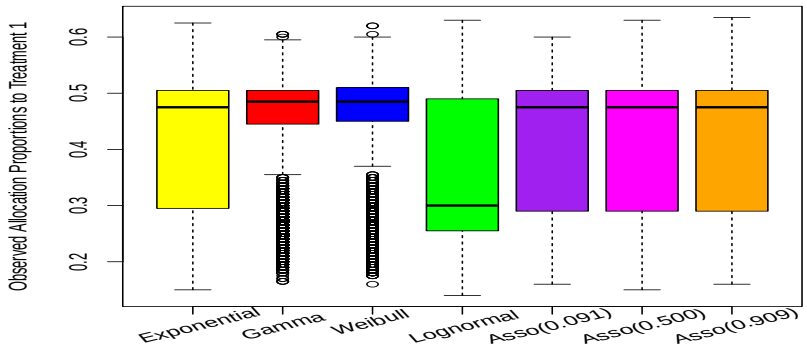

Figure 1: Boxplots of observed allocation proportions to treatment 1 for $\mu_1 = 2.6, \mu_2 = \mu_3 = 2$, $\nu_1 = \nu_2 = \nu_3 = 1$ with $m = 40$ and $N = 200$.

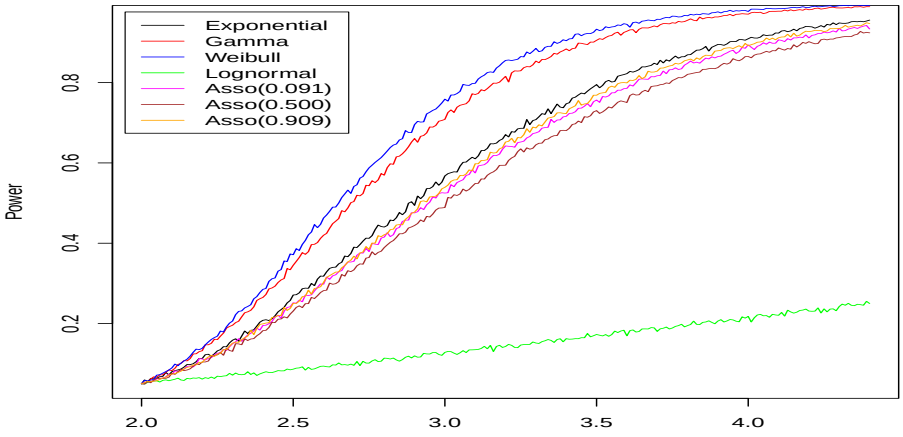

Figure 2: Comparison of powers for $\mu_2 = \mu_3 = 2$, $\nu_1 = \nu_2 = \nu_3 = 1$ with varying μ_2 from 2.0 to 4.4 under $m = 40$ and $N = 200$.

Remarks: First of all, the type I error rates are more or less maintained at the nominal level for all the cases except one. From Tables 1-2 and Figures 1-2, it is evident that higher proportion of patients are assigned to the best treatment with higher power when the response is originally Gamma or Weibull rather than the Exponential but it is lower when the response variable changes to Lognormal. Moreover, when the response variable is Exponential but associated, then the power gradually drops as we increase the intensity of association (i.e. τ). The EAP's for best treatment are fluctuating without specifying any pattern.

Moreover, Tables 2-3 are showing that, even if μ_k's are same, all the three treatments have different EAP values if ν_k's are different when the model assumption is violated. One very unusual situation for Table 3 is when the Exponential distributions have high association. In those cases, all the operating characteristics (i.e. type I error rate, EAP and power) are giving such values which are completely unexpected and non-interpretable.

In addition, Table 4 shows that the EAP figures do not vary for the two superior treatments and have higher values than inferior treatment, which is expected. But the powers are really very low for all the cases. Hence, the performance of the design is not satisfactory when the actual situation is different from the assumed conditions.

4. Redesigning Recurrent Glioblastoma trial of Batchelor et al. (2013)

For the evaluation of the allocation design of Bhattacharya and Shome (2019) from the perspective of a real clinical practitioner, we consider the same real clinical trial

considered in Bhattacharya and Shome (2019), that is, the recurrent glioblastoma trial of Batchelor et al. (2013), where the efficacies of combination of Cediranib 20 mg and Lomustine (Treatment 1), Cediranib 30 mg (Treatment 2) and Lomustine (Treatment 3) alone are evaluated. 325 patients with recurrent glioblastoma were randomly assigned using a 2:2:1 randomization to receive treatments 1, 2 and 3, respectively. However, the study did not show significant prolongation in progression free survival (PFS) for any treatment.

As only the summary statistics of the trial data are given publicly by Batchelor et al. (2013), it is not possible to check the validity of the Koziol-Green model. So, we have assumed that PFS for these treatments follow the Koziol-Green model and estimated the parameters. It is worthwhile to mention that Bhattacharya and Shome (2019) assumed the same to carry out the relevant analysis based on this data. According to the Koziol-Greeen model, the relevant estimates (in months) are $\hat{\mu}_1 = 6.011, \hat{\mu}_2 = 4.423, \hat{\mu}_3 = 3.943$ and $\hat{\nu}_1 = 0.554, \hat{\nu}_2 = 0.456, \hat{\nu}_3 = 0.757$. Treating these estimates as the true ones, we redesign the trial in two stages with the trial size $N = 325$ for two choices of θ as $1/5$ and $1/4$ under different models. Relevant findings with respect to expected allocation proportions (EAP) and expected allocation counts (EAC) with their respective SD's are reported in Table 5.

The figures from the Table 5 show that the two stage adaptive procedure of Bhattacharya and Shome (2019) performs better than the actual situation and may save a number of patients assigning to the inferior treatment whatever be the distribution of responses.

Table 5: Redesigning the recurrent glioblastoma trial (Batchelor et al., 2013) and its evaluation with respect to the operating characteristics under different models.

	EAP (SD)				EAC (SD)		
Model	Trt-1	Trt-2	Trt-3	Trt-1	Trt-2	Trt-3	
$\theta = 1/5$							
Exponential (mean=μ_k)	0.552 (0.114)	0.237 (0.099)	0.210 (0.062)	179.529 (36.963)	77.072 (32.127)	68.4 (20.076)	
Gamma (shape=3, scale=$\mu_k/3$)	0.577 (0.070)	0.216 (0.051)	0.207 (0.051)	187.526 (22.759)	70.265 (16.503)	67.209 (16.698)	
Weibull (shape=3, scale=$\mu_k/0.893$)	0.581 (0.058)	0.213 (0.036)	0.206 (0.049)	188.876 (18.889)	69.138 (11.796)	66.986 (15.878)	
Lognormal (meanlog=log(μ_k), sdlog=3)	0.394 (0.192)	0.360 (0.187)	0.247 (0.126)	128.040 (62.335)	116.839 (60.930)	80.121 (40.827)	
Associated Exponential ($\tau = 0.091$)	0.549 (0.118)	0.237 (0.099)	0.213 (0.070)	178.587 (38.229)	77.127 (32.057)	69.286 (22.610)	
Associated Exponential ($\tau = 0.500$)	0.557 (0.107)	0.221 (0.065)	0.222 (0.088)	181.150 (34.739)	71.748 (21.124)	72.102 (28.702)	
Associated Exponential ($\tau = 0.909$)	0.573 (0.080)	0.221 (0.067)	0.206 (0.048)	186.222 (26.133)	71.839 (21.749)	66.939 (15.730)	
$\theta = 1/4$							
Exponential (mean=μ_k)	0.476 (0.063)	0.271 (0.057)	0.253 (0.036)	154.644 (0.491)	88.115 (18.391)	82.241 (11.598)	
Gamma (shape=3, scale=$\mu_k/3$)	0.489 (0.037)	0.260 (0.030)	0.251 (0.031)	158.773 (12.055)	84.651 (9.78)	81.576 (10.159)	
Weibull (shape=3, scale=$\mu_k/0.893$)	0.489 (0.033)	0.259 (0.026)	0.251 (0.031)	159.048 (10.826)	84.289 (8.347)	81.664 (9.971)	
Lognormal (meanlog=log(μ_k), sdlog=3)	0.373 (0.119)	0.352 (0.117)	0.275 (0.077)	121.22 (38.802)	114.355 (38.032)	89.425 (24.867)	
Associated Exponential ($\tau = 0.091$)	0.475 (0.065)	0.271 (0.056)	0.255 (0.040)	154.213 (21.206)	87.979 (18.312)	82.808 (13.019)	
Associated Exponential ($\tau = 0.500$)	0.478 (0.060)	0.263 (0.039)	0.259 (0.051)	155.229 (19.548)	85.572 (12.585)	84.198 (16.673)	
Associated Exponential ($\tau = 0.909$)	0.486 (0.044)	0.263 (0.039)	0.251 (0.030)	157.882 (14.457)	85.529 (12.809)	81.588 (9.791)	
Actual	0.397	0.403	0.200	129	131	65	

5. Recommendations

The present paper discusses the effect of model misspecification on the two stage design of Bhattacharya and Shome (2019). Although the procedure is meritorious when all the assumptions made by them are valid, it has been revealed that, the procedure may perform poorly in certain situations. So it is necessary to check the model validity before performing any particular design which is very much dependent on some assumptions. Real clinical trials should also be performed very cautiously under these scenarios. Some distribution free and robust methods may be applied, in general, to get rid of some unsatisfactory results.

Acknowledgements

The authors thank the anonymous reviewers, the Editors and the Associate Editors for their insightful comments which led to an improvement over the earlier version of this work.

References

[1] Bandyopadhyay, U. and Bhattacharya, R. (2007). On an ethical cum optimal adaptive allocation design. *Statistics* 41, 471-483.

[2] Batchelor, T. T., Mulholland, P., Neyns, B., Nabors, L. B., Campone, M., Wick, A., Mason, W., Mikkelsen, T., Phuphanich, S., Ashby, L. S., DeGroot, J., Gattamaneni, R., Cher, L., Rosenthal, M., Payer, F., Jürgensmeier, J.M., Jain, R.K., Sorensen, A.G., Xu, J., Liu, Q. and Bent, M. (2013). Phase III Randomized Trial Comparing the Efficacy of Cediranib As Monotherapy, and in Combination With Lomustine, Versus Lomustine Alone in Patients With Recurrent Glioblastoma. *Journal Of Clinical Oncology* 31, 3212-3218.

[3] Berger, R. L. and Hsu, J. C. (1996). Bioequivalence Trials,Intersection-Union Tests and Equivalence Confidence Sets. *Statistical Science* 11, 283-302.

[4] Bhattacharya, R. and Shome, M. (2019). A multi-treatment two stage adaptive allocation for survival outcomes, *Communications in Statistics- Theory and Methods*, 48, 2001-2013.

[5] Clayton, D. G. (1978). A Model for Association in Bivariate Life Tables and Its Application in Epidemiological Studies of Familial Tendency in Chronic Disease Incidence. *Biometrika*, 65(1), 141-151.

[6] Coad, D. S. (1992). Some results on estimation for two stage clinical trials. *Sequential Analysis* 11, 299-311.

[7] Coad, D. S. (1994). Sequential estimation for two-stage and three-stage clinical trials. *Journal of Statistical Planning and Inference* 43, 343-351.

[8] Colton, T. (1963). A model for selecting one of two medical treatments. *Journal of American Statistical Association* 58, 388-400.

[9] Hsu, J. C. (1996). *Multiple Comparisons Theory And Methods*, Chapman & Hall, London, UK.

[10] Koziol, J. A. and Green, S. B. (1976). A Cramer-von Mises statistic for randomly censored data. *Biometrika* 63, 465-474.

[11] Matthews, J. N. S. (2006). *Introduction to Randomized Controlled Clinical Trials, 2nd Edition*, Routledch, Taylor & Francis Group.

[12] Rosenberger, W. F. and Lachin, J. L. (2015). *Randomisation in Clinical Trials: Theory and Practice, Second Edition* Wiley, New York.

[13] Silvapulle, M. J. and Sen, P. K. (2005). *Constrained Statistical Inference : Inequality, Order, and Shape Restrictions* John Wiley & Sons, Inc., Hoboken, New Jersey.