



ON A CLASS OF OPTIMAL TYPE COVARIATE ADJUSTED RESPONSE ADAPTIVE ALLOCATIONS FOR NORMAL TREATMENT RESPONSES

Firstname Lastname

Affiliation

Abstract

In the context of clinical trials for comparing two treatments, a new class of covariate adjusted response adaptive procedures is developed to achieve a balance between clinical optimality and inferential precision. Assuming normally distributed responses, several exact and asymptotic properties are studied and compared with a reasonable alternative under the presence of treatment covariate interaction.

Keywords: Response adaptive design, Covariate adjusted allocation, Optimality.

1. Introduction

Response adaptive allocation is a useful technique in clinical trials to skew the allocation of incoming subjects towards the better performing treatment by using the available information. Assuming homogeneity among the subjects, a number of response adaptive allocation designs were developed in the last few decades. But heterogeneity with respect to some covariate (e.g. age, sex etc.) is quite obvious in trials involving human beings. The presence of covariate information often influences the responses of the subjects and this suggests to incorporate such information into the process of randomization. Therefore, a sensible allocation design should randomize any entering subject on the basis of the data accrued thus far together with his/her covariate information. Such allocations are referred to as *covariate adjusted allocation designs* (Hu and Rosenberger, 2006).

A number of attempts (see, for example, Begg and Iglewicz (1980), Atkinson (1982) and Wong and Zhu (2008)) is made to incorporate covariates into the randomisation process using the Efron's biased coin methodology in the framework of optimum design theory. But these schemes do not allow to incorporate the information of the previous treatment outcomes, and hence are not response adaptive and consequently lacks the ethical requirement of reducing the assignment fraction to the inferior treatment. However, in the context of continuous response trials, only a few response adaptive procedures incorporate covariate information, in addition. Although Bandyopadhyay and Biswas (2001) and Biswas et al. (2006) incorporated available responses, allocations and covariate information in randomisation, ignored the cur-

rent covariate information. Therefore, the allocation designs developed with covariates either minimize the variability of the estimated treatment effect incorporating covariates(see, for example, Begg and Iglewicz(1980),Atkinson(1982) and Wong and Zhu(2008)) without caring allocation of patients or favour the better performing treatment for allocation (see, for example, Bandyopadhyay and Biswas(2001) and Biswas et al.(2006)) accepting a significant loss in statistical precision. Although Jennison and Turnbull(2000) provided an unified approach to find an allocation ratio as a trade off between the ethical need of optimizing a clinically relevant criteria (e.g., number of allocations to the inferior treatment, total expected response or number of failures) and preservation of power for testing the equality of treatment effects, ignored the possibility of any covariate information. Along the line of Jennison and Turnbull(2000), Biswas and Mandal(2004) made an attempt to incorporate covariate information with an aim to minimise treatment failure but ignored the assessment of the so claimed optimality. It is worth mentioning at this point that, recently Azriel et al.(2012) derived an optimal proportion considering only the maximization of power of a test of equality of treatment effects but in the context of covariate independent binary response trials.

The existence of interaction between treatments and covariates is a common phenomena in any clinical trial. But the effect of such interactions under a response adaptive set up is not explored completely. In fact, most of the works related to covariate either ignored the possibility of such interactions(e.g.,Bandyopadhyay and Biswas(2001), Biswas and Mandal(2004))or, if incorporated, avoid to assess the effect in details. However, the ethical goal consists always in assigning a greater fraction of subjects to the better performing treatment independently of the presence or absence of treatment-covariate interaction. But the only difference is that in the presence of such interactions, the benefit of the treatments depends on the kind of patients(Friedman et al.(2010)). The present work, in the presence of treatment covariate interaction, is an attempt to develop a class of target allocation proportions by combining the views of both statistician and clinician when the response variable is continuous. The relevant measures of optimality, optimal target proportions and their implementation, along with some limiting results, are all given in Section 2. In Section 3 we explore and compare the small sample properties of the proposed allocation for a hypothetical clinical trial. Section 4 concludes with a discussion of some related issues.

2. The Proposed Allocation

2.1. Clinical Optimality

Suppose two treatments, denoted by A and B , are under a clinical trial and the patients appear in the clinic sequentially. Each patient is treated once, the responses are immediate and a lower response indicates a favourable situation. Let Y_{ki} denote the potential outcome of the i th subject on treatment $k, k=A, B$ with \mathbf{Z}_i as the associated p component vector of covariates. In practical situations either Y_{Ai} or Y_{Bi} is observed for the i th subject. We assume that the conditional distribution of Y_{ki} given $\mathbf{Z}_i = \mathbf{z}$ is normal with mean $\mu_k(\mathbf{z}) = \alpha_k + \beta_k^T \mathbf{z}$ and variance σ_k^2 . We further assume that the variables \mathbf{Z}_i 's are independently and identically distributed (*iid*) according to a p -variate distribution which is completely/partially known. However, in real clinical trials, the covariates are the measurements on some clinical characteristics of human beings(e.g. age, blood pressure,glucose level), and hence they vary within certain limits. Therefore, it is clinically meaningful to assume that the covariates $\{\mathbf{Z}_j, j \geq 1\}$ are uniformly bounded.

As indicated earlier, the first step in deriving an optimal allocation is to define a measure of ethical sensitivity. Treatment failures, allocations to the inferior treatment and the expected treatment responses are often used to measure the ethical sensitivity. But, before quantifying such measures, we need to define a treatment success and an effective treatment. Unlike a binary response trial, a treatment success, in the context of a continuous response trial, is not uniquely defined. However, if we look at real clinical trials to reduce hypertension,

blood pressure or blood sugar, we find that the response is continuous and a lower response, irrespective of the covariate value, is favourable. Therefore, if a lower response indicates a favourable situation, a treatment producing a response lower than a clinically relevant threshold can be regarded as successful. Thus treatment k is successful/beneficial for the i th subject if $Y_{ki} \leq c$ for some clinically relevant threshold c . Moreover, if a lower response is favourable, treatment k is more effective for the i th patient than treatment k' if $Y_{ki} < Y_{k'i}$, $k \neq k'$. Therefore, we can set the following choices of the ethical metric $\gamma_k(\mathbf{Z})$ for treatment k : $\gamma_{k1}(\mathbf{Z}) = P(Y_{k1} > Y_{k'1} | \mathbf{Z})$, $\gamma_{k2}(\mathbf{Z}) = P(Y_{k1} \geq c | \mathbf{Z})$ and $\gamma_{k3}(\mathbf{Z}) = P(Y_{k1} \geq c, Y_{k1} > Y_{k'1} | \mathbf{Z})$, where lower values are desired for any treatment k . Suppose n_k subjects are assigned to treatment $k (= A, B)$ in a non-randomized manner and it remains to define a combined metric. However, different covariate profiles are of different importance and hence deriving a combined metric requires further consideration. For a better apprehension, assume that the covariate Z is categorical taking only the values z_1, z_2 and z_3 with the consideration that $z_1(z_3)$ is the most(least) favorable covariate profile. That is, $z_1 < z_2 < z_3$ and hence, the response of a subject with covariate profile z_1 to a particular treatment is expected to be the smallest. Then an ethical metric corresponding to treatment k can be expressed as $\sum_{j=1}^3 w_j \gamma_k(z_j)$, where w_j is the non-negative weight assigned to covariate profile z_j , $j = 1, 2, 3$. Since, a lower value of $\gamma_k(z)$ indicates higher treatment effectiveness for covariate profile z , $\gamma_k(z)$ must be increasing in z . Now, under the presence of treatment covariate interaction, treatment effectiveness depends on the covariate profile and in the current context we assume that treatment A is most effective for $Z = z_1$ and least effective for $Z = z_3$ but for $Z = z_2$ both of the treatments are equally effective. Thus the relations $\gamma_A(z_1) \leq \gamma_B(z_1)$, $\gamma_A(z_2) = \gamma_B(z_2)$ and $\gamma_A(z_3) \geq \gamma_B(z_3)$, are natural. Although different choices of weights provide different metrics but we use $w_j = P(Z = z_j)$, $j = 1, 2, 3$ in the current work. Such a choice reduces the treatment effectiveness measure for treatment k to $E_{\mathbf{Z}} \{\gamma_k(\mathbf{Z})\}$ and in the current work, we continue our development with this. Now, if we consider the first criterion, it can be observed that $E_{\mathbf{Z}} \{\gamma_{A1}(\mathbf{Z})\} = P(Y_{A1} \geq Y_{B1})$ represents the probability that treatment A is inferior than treatment B . Naturally, $n_A P(Y_{A1} \geq Y_{B1}) + n_B P(Y_{B1} \geq Y_{A1})$ can be looked upon as the expected number of assignments to the inferior treatment. Again, $E_{\mathbf{Z}} \{\gamma_{A2}(\mathbf{Z})\} = P(Y_{A1} \geq c)$ gives the probability that treatment A is not beneficial and hence the expected number of subjects not benefited under the trial can be expressed as $n_A P(Y_{A1} \geq c) + n_B P(Y_{B1} \geq c)$. In a similar fashion, $n_A P(Y_{A1} \geq Y_{B1}, Y_{A1} \geq c) + n_B P(Y_{B1} \geq Y_{A1}, Y_{B1} \geq c)$ represents the expected number of subjects assigned to the inferior and non-beneficial treatment. Therefore, a combined measure of the ethical sensitivity corresponding to any ethical metric $\gamma_k(\mathbf{Z})$ can be expressed as

$$G(n_A, n_B) = n_A E_{\mathbf{Z}} \{\gamma_A(\mathbf{Z})\} + n_B E_{\mathbf{Z}} \{\gamma_B(\mathbf{Z})\}.$$

Clearly, with these choices of $\gamma_k(\mathbf{Z})$, minimization of $G(n_A, n_B)$ is required to maintain the ethical requirements. Under normality of responses, the ethical metrics can be expressed as $\gamma_{k1}(\mathbf{Z}) = \Phi\left(\frac{\mu_k(\mathbf{Z}) - \mu_{k'}(\mathbf{Z})}{\sqrt{\sigma_A^2 + \sigma_B^2}}\right)$ and $\gamma_{k2}(\mathbf{Z}) = \Phi\left(\frac{\mu_k(\mathbf{Z}) - c}{\sigma_k}\right)$. For the derivation of the third criterion, we note that the conditional (i.e. for given \mathbf{Z}) joint distribution of Y_{k1} and $Y_{k1} - Y_{k'1}$ is bivariate normal with means $\mu_k(\mathbf{Z})$ and $\mu_k(\mathbf{Z}) - \mu_{k'}(\mathbf{Z})$, variances σ_k^2 and $\sigma_A^2 + \sigma_B^2$ and correlation coefficient $\zeta_k = \frac{\sigma_k}{\sqrt{\sigma_A^2 + \sigma_B^2}}$. Hence $\gamma_{k3}(\mathbf{Z})$ can be expressed as $\Phi_2\left(\frac{\mu_k(\mathbf{Z}) - c}{\sigma_k}, \frac{\mu_k(\mathbf{Z}) - \mu_{k'}(\mathbf{Z})}{\sqrt{\sigma_A^2 + \sigma_B^2}}, \zeta_k = \frac{\sigma_k}{\sqrt{\sigma_A^2 + \sigma_B^2}}\right)$, where $\Phi_2(\cdot, \cdot, \zeta)$ is the distribution function of a standard bivariate normal distribution with correlation coefficient ζ . Now it is observed that $\gamma_{k1}(\mathbf{Z})$ is sensitive only the difference in the treatment effectiveness and $\gamma_{k2}(\mathbf{Z})$ is sensitive only to the departure from the threshold whereas $\gamma_{k3}(\mathbf{Z})$ is sensitive to both difference in the treatment effectiveness and the departure from the threshold.

2.2. Statistical Optimality

In a statistical optimal procedure one can search for an allocation rule by controlling the

variability of a distance measure between the two parameters $\boldsymbol{\eta}_A$ and $\boldsymbol{\eta}_B$, where $\boldsymbol{\eta}_k = (\alpha_k, \boldsymbol{\beta}_k^T)^T$, $k = A, B$. However, the true parameters are never known in advance and hence, as an alternative, we suggest to control the variability of a relevant sample distance measure. In the current work we suggest to control the variability of the Wald type distance measure, used in testing the equality of $\boldsymbol{\eta}_A$ and $\boldsymbol{\eta}_B$. If $\hat{\boldsymbol{\eta}}_{kn}$ denotes the maximum likelihood (ML) estimate of $\boldsymbol{\eta}_k$ based on $n = n_A + n_B$ observations, then such distance is defined by the statistic

$$W_n = (\hat{\boldsymbol{\eta}}_{An} - \hat{\boldsymbol{\eta}}_{Bn})^T \hat{V}_n^{-1} (\hat{\boldsymbol{\eta}}_{An} - \hat{\boldsymbol{\eta}}_{Bn}),$$

where \hat{V}_n is the ML estimate of V_n , the conditional (given all the covariate values) dispersion matrix of $\hat{\boldsymbol{\eta}}_{An} - \hat{\boldsymbol{\eta}}_{Bn}$. Under the normal response model, the conditional distribution (given all the covariate values) of $\hat{\boldsymbol{\eta}}_{An} - \hat{\boldsymbol{\eta}}_{Bn}$ is $(p + 1)$ variate normal with mean vector $\boldsymbol{\eta}_A - \boldsymbol{\eta}_B$ and dispersion matrix

$$V_n = \frac{\sigma_A^2}{n_A} S_A^{-1} + \frac{\sigma_B^2}{n_B} S_B^{-1},$$

where $S_k = \frac{1}{n_k} \sum_{j=1}^{n_k} \boldsymbol{\xi}_j \boldsymbol{\xi}_j^T$ and $\boldsymbol{\xi}_j = (1, \mathbf{Z}_j^T)^T$ and is independent of \hat{V}_n . Then, since W_n is a quadratic form involving $\hat{\boldsymbol{\eta}}_{An} - \hat{\boldsymbol{\eta}}_{Bn}$ and \hat{V}_n , the conditional variance of W_n can be expressed as (see, for example, Seber and Lee(2003))

$$4(\boldsymbol{\eta}_A - \boldsymbol{\eta}_B)^T \hat{V}_n^{-1} V_n \hat{V}_n^{-1} (\boldsymbol{\eta}_A - \boldsymbol{\eta}_B) + 2 \text{Trace}(\hat{V}_n^{-1} V_n \hat{V}_n^{-1} V_n).$$

We now assume that $n \rightarrow \infty$ and $\frac{n_k}{n}$ converges to a limit belonging to $(0, 1)$. Then, if $E_{\mathbf{Z}_1}(\boldsymbol{\xi}_1 \boldsymbol{\xi}_1^T)$ is positive definite, \hat{V}_n is asymptotically equivalent in probability to V_n and hence the asymptotic variance of W_n comes out to be a decreasing function of $\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}$. Therefore, fixing the asymptotic variance of W_n or equivalently $\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}$, a statistically optimal procedure can be obtained.

2.3. The Optimal Target

Thus we have in our hand the measures of both statistical and clinical optimality, the prerequisites to derive an optimal allocation ratio. But minimisation of $G(n_A, n_B)$ alone provides degenerate rules and this necessitates the consideration of statistical optimality. Naturally, to find the optimal ratio $\frac{n_A}{n_A + n_B}$, we need to minimise $G(n_A, n_B)$ fixing $\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}$ below a suitable quantity. We, therefore, formulate the following optimization problem:

$$\min_{n_A, n_B} [n_A E_{\mathbf{Z}} \{\gamma_A(\mathbf{Z})\} + n_B E_{\mathbf{Z}} \{\gamma_B(\mathbf{Z})\}]$$

subject to

$$\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B} = \kappa,$$

where κ is a prefixed positive quantity. With $\Psi_k = E_{\mathbf{Z}} \{\gamma_k(\mathbf{Z})\} (> 0)$, the optimal solution to the optimisation problem above is

$$\pi_A = \left(\frac{n_A}{n_A + n_B} \right)_{opt} = \frac{\sigma_A \sqrt{\Psi_B}}{\sigma_A \sqrt{\Psi_B} + \sigma_B \sqrt{\Psi_A}}.$$

2.4. CARA Allocation in practice

It is worth mentioning at this point that the presence of treatment covariate interaction implies that a treatment is not uniformly effective for every covariate profile and hence the allocation probability to a treatment must depend on the patient's covariate profile. For

example, suppose treatment A is more(less) effective than treatment B for a given covariate profile. Naturally, the allocation probability of a subject with that covariate profile must be higher(lower) for treatment A than treatment B. However, the optimal ratio π_k does not depend on the covariate profile \mathbf{Z} and hence to assign patients according to their covariate profile, we need to use the current patient's covariate for his/her assignment. In most of the trials the covariate information is available at the time of randomisation, and hence it is not difficult to incorporate such information in the randomisation procedure. In addition, the optimal ratio π_k depends on the unknown parameter $\boldsymbol{\theta} = (\boldsymbol{\theta}_A^T, \boldsymbol{\theta}_B^T)$ with $\boldsymbol{\theta}_k = (\boldsymbol{\eta}_k^T, \sigma_k)^T$ and hence we need to replace them by suitable estimates. Therefore, in order to develop a reasonable randomization procedure based on the optimal ratio, we need to use the current patient's covariate and the available parameter estimates in a reasonable way. In the present work, for the assignment of the $(i+1)$ th subject, we suggest to replace the unknown quantity $E_{\mathbf{Z}}\{\gamma_k(\mathbf{Z})\}$ by its projection $\hat{\gamma}_{ki}(\mathbf{Z}_{i+1})$, where the unknown parameters are replaced by the estimates, calculated on the basis of the response, allocation and covariate data prior to the randomization of the $(i+1)$ th subject and \mathbf{Z} by the current covariate value \mathbf{Z}_{i+1} . If δ_{kj} is the treatment indicator of the j th subject, that is, $\delta_{kj} = 1$ if treatment k is applied and 0 otherwise and $\mathcal{F}_i, i \geq 1$ denotes the information on allocations, responses and covariates up to and including the i th subject then the suggested CARA randomization corresponding to any metric $\gamma_k(\mathbf{Z})$ can be described by the allocation probability

$$P(\delta_{k,i+1} = 1 | \mathcal{F}_i, \mathbf{Z}_{i+1}) = \pi_{ki}(\hat{\boldsymbol{\theta}}_i, \mathbf{Z}_{i+1})$$

where $\hat{\boldsymbol{\theta}}_i$ is the ML estimate of $\boldsymbol{\theta}$ based on history of the data up to the i -th entry of the incoming subject and

$$\pi_{Ai}(\hat{\boldsymbol{\theta}}_i, \mathbf{Z}_{i+1}) = 1 - \pi_{Bi}(\hat{\boldsymbol{\theta}}_i, \mathbf{Z}_{i+1}) = \frac{\hat{\sigma}_{Ai} \sqrt{\hat{\gamma}_{Bi}(\mathbf{Z}_{i+1})}}{\hat{\sigma}_{Bi} \sqrt{\hat{\gamma}_{Ai}(\mathbf{Z}_{i+1})} + \hat{\sigma}_{Ai} \sqrt{\hat{\gamma}_{Bi}(\mathbf{Z}_{i+1})}}.$$

Clearly $\pi_{ki}(\hat{\boldsymbol{\theta}}_i, \mathbf{Z}_{i+1})$ is an estimate based on the available data where $E_{\mathbf{Z}}\{\gamma_k(\mathbf{Z})\}$ is replaced by its projection $\hat{\gamma}_{ki}(\mathbf{Z}_{i+1})$ and σ_k is replaced by its available estimate $\hat{\sigma}_{ki}, k = A, B$. However, for implementation in practice, we need to assign an initial number n_0 of subjects to each treatment arm to get the starting estimates of the parameters. For the estimation of parameters, we suggest to use sequentially updated maximum likelihood estimates. Specifically, with the response model as $Y_{kj} \sim N(\boldsymbol{\eta}_k^T \boldsymbol{\xi}_j, \sigma_k^2), j \geq 1$ and observations up to stage i , the likelihood function of $\boldsymbol{\theta}$, ignoring the constant of proportionality, can be expressed as

$$\mathcal{L}_i(\boldsymbol{\theta}) = \mathcal{L}_i(\boldsymbol{\theta}_A) \mathcal{L}_i(\boldsymbol{\theta}_B),$$

where

$$\log \mathcal{L}_i(\boldsymbol{\theta}_k) = -\frac{\sum_{j=1}^i \delta_{kj}}{2} \log \sigma_k^2 - \frac{1}{2} \sum_{j=1}^i \delta_{kj} \left(\frac{Y_{kj} - \boldsymbol{\eta}_k^T \boldsymbol{\xi}_j}{\sigma_k} \right)^2.$$

Now, writing

$$\begin{aligned} N_{ki} &= \sum_{j=1}^i \delta_{kj}, \bar{y}_{ki} = \frac{\sum_{j=1}^i \delta_{kj} Y_{kj}}{N_{ki}}, \bar{\mathbf{z}}_{ki} = \frac{\sum_{j=1}^i \delta_{kj} \mathbf{Z}_j}{N_{ki}} \\ S_{zz,i}^{(k)} &= \sum_{j=1}^i \delta_{kj} (\mathbf{Z}_j - \bar{\mathbf{z}}_{ki})(\mathbf{Z}_j - \bar{\mathbf{z}}_{ki})^T \\ \mathbf{s}_{zY,i}^{(k)} &= \sum_{j=1}^i \delta_{kj} (\mathbf{Z}_j - \bar{\mathbf{z}}_{ki}) Y_{kj}, \end{aligned}$$

we can express the ML estimates based on the data available up to i th stage as

$$\hat{\boldsymbol{\beta}}_{ki} = S_{zz,i}^{(k)-1} \mathbf{s}_{zY,i}^{(k)}, \quad \hat{\alpha}_{ki} = \bar{y}_{ki} - \hat{\boldsymbol{\beta}}_{ki}^T \bar{\mathbf{z}}_{ki}, \quad \hat{\sigma}_{ki}^2 = \frac{1}{N_{ki}} \sum_{j=1}^i \delta_{kj} (Y_{kj} - \hat{\alpha}_{ki} - \hat{\boldsymbol{\beta}}_{ki}^T \mathbf{Z}_j)^2.$$

Let $\hat{\boldsymbol{\theta}}_{kn}$ denote the ML estimate of $\boldsymbol{\theta}_k$ based on the data available up to the n th stage. Then irrespective of any metric $\gamma_k(\mathbf{Z})$ we have the following results, proved in Appendix A.

Result 2.1: As $n \rightarrow \infty$,

$$\hat{\boldsymbol{\theta}}_{kn} \rightarrow \boldsymbol{\theta}_k$$

almost surely, $k = A, B$.

Result 2.2: As $n \rightarrow \infty$,

$$\frac{N_{kn}}{n} \rightarrow E_{\mathbf{Z}_1} \pi_k(\boldsymbol{\theta}, \mathbf{Z}_1)$$

almost surely, $k = A, B$.

Now to examine the extent of benefited proportion of subjects under the proposed allocation, we define the indicator variable $F_i = \delta_{Ai} I_{[Y_{Ai} \leq c]} + \delta_{Bi} I_{[Y_{Bi} \leq c]}$, $i \geq 1$ for the i th patient. We use $\bar{F}_n = n^{-1} \sum_{i=1}^n F_i$ as an indicative measure of the proportion benefited under the trial, which takes into account the covariate information. Then we have the following result, which is proved in Appendix B.

Result 2.3: As $n \rightarrow \infty$,

$$\bar{F}_n \rightarrow \psi = E_{\mathbf{Z}_1} \left\{ \sum_{k=A,B} \pi_k(\boldsymbol{\theta}, \mathbf{Z}_1) P(Y_{k1} \leq c \mid \mathbf{Z}_1) \right\},$$

almost surely.

However, the role of covariate information can be best explained if we can separate the covariate levels. For example, considering \mathbf{Z} to be categorical, we denote the number of subjects having covariate level \mathbf{z} assigned to treatment k by $N_{kn}(\mathbf{z})$ and the corresponding total number of subjects for a given covariate level \mathbf{z} by $N_n(\mathbf{z})$, after the completion of n assignments. Then we get $\frac{N_{kn}(\mathbf{z})}{N_n(\mathbf{z})}$ as the conditional proportion of subjects assigned to treatment k at $\mathbf{Z} = \mathbf{z}$. Now we have the following result, the proof of which follows from Zhang et al.(2007).

Result 2.4: As $n \rightarrow \infty$,

$$\frac{N_{kn}(\mathbf{z})}{N_n(\mathbf{z})} \rightarrow \pi_k(\boldsymbol{\theta}, \mathbf{z})$$

almost surely, for a given covariate, provided $P(\mathbf{Z}_1 = \mathbf{z}) > 0$.

In a similar fashion, the conditional proportion of subjects benefited in the trial can be expressed as $\frac{F_n(\mathbf{z})}{N_n(\mathbf{z})}$, where $F_n(\mathbf{z})$ denote the number of subjects benefited under the trial for a given covariate category \mathbf{z} . Proceeding as in Zhang et al.(2007), for a given covariate \mathbf{z} , we find the almost sure limit of the conditional proportion of benefited subjects as $\sum_{k=A,B} \pi_k(\boldsymbol{\theta}, \mathbf{z}) P(Y_{k1} \leq c \mid \mathbf{z})$, provided $P(\mathbf{Z}_1 = \mathbf{z}) > 0$.

However, assessing the optimality in the presence of covariate is not only critical but also a debatable issue. In a covariate ignored procedure, optimality means achieving the optimal target proportions in the limit. But there is no real guideline on deriving and defining an optimal allocation in the presence of covariate information. The current work is a reasonable extension of the framework developed in Jennison and Turnbull(2000) to derive the optimal target proportions. Although the derivation of the optimal proportion is based on average patients but in actual implementation, the current patient's covariate is used to calculate his/her assignment probability. In particular, we suggested to replace $E_{\mathbf{Z}} \{\gamma_k(\mathbf{Z})\}$ by its projection $\hat{\gamma}_{ki}(\mathbf{Z}_{i+1})$ to allow the assignment probability of the $(i+1)$ th patient to depend on his/her covariate. This makes the allocation *covariate adjusted* but leads to a limiting allocation proportion which is different from the target optimal proportion. On the other hand, if one

replaces $E_{\mathbf{Z}} \{\gamma_k(\mathbf{Z})\}$ by the conventional estimate $i^{-1} \sum_{j=1}^i \hat{\gamma}_{kj}(\mathbf{Z}_j)$ for the assignment of the $(i+1)$ th subject, the optimal target proportion π_k is realized in the limit and makes the allocation *conventionally optimal*. But in any clinical trial the ethical imperative lies in ensuring the best possible care for the subjects using all available information, and a *conventionally optimal* procedure does not incorporate the current patient's covariate for his/her allocation even if it is available prior to randomization and hence lacks sensibility. Naturally, the proposed allocation seemed to be more desirable than the conventionally optimal allocation in the context of a real clinical trial. In fact, we have modified the unconditional optimal proportion π_k in accordance with the clinician's requirement to get the conditional(conditional on the covariate) treatment allocation probabilities and hence, refer the allocation as *near or pseudo optimal*.

3. Performance Evaluation

3.1. A hypothetical clinical trial

To assess the performance of the proposed allocation, we consider a hypothetical clinical trial on reducing systolic blood pressure of the patients. Since, in general, the response to a treatment of blood pressure reduction varies with age, we consider age as the only covariate and assume a realistic situation, where the experimental treatment (i.e. treatment A) is more efficient than the Control(i.e. treatment B) for all age groups but the effectiveness for both the treatments is the highest for the young patients, moderate for the middle aged and the lowest for the old aged. This naturally convinces the presence of a quantitative treatment-covariate interaction. The patients enter the study sequentially, their ages are noted and depending on the age they are classified either Young(age <30 years)or Middle-aged(age falls in the interval 30-60 years) or Old-aged (age > 60 years). Naturally, the effect of the treatment is expected to be the most for the young patients and the least for the old-aged patients and hence the young patient category is the most favourable(MF), the middle-aged category is only favourable(F) whereas the old-aged category is least favourable(LF). Since we are categorizing a continuous variable(i.e. age), it becomes necessary to assign appropriate values to different categories. If the relative weights of different categories are known, then one can use the well known Likert's category scaling(Garret(1966)) to determine the representative numerical values. Specifically, depending on the proportion of patients in the *MF,F* and *LF* categories, the corresponding values of the categorizing variable Z can be determined.

3.2. Performance Measures

For the present hypothetical trial a treatment is naturally beneficial if it is capable of reducing the systolic blood pressure of the subject below the *Normal* limit. Recent reports (e.g. "Understanding blood pressure readings", American Heart Association, 2011; <http://www.heart.org/HEARTORG/Conditions/HighBloodPressure>) reveal that a systolic blood pressure of less than 120 mmHg is defined as *Normal* and this suggests to take c as 120 (in mmHg), irrespective of the other prognostic factors. Being consistent with the development so far, we assume that the response of a subject assigned to treatment k has a normal distribution with mean $\alpha_k + \beta_k z$ and constant variance σ^2 , for given covariate value z of Z following a discrete probability distribution. Suppose the proportion of patients in the *MF,F* and *LF* categories are respectively $\frac{1}{6}$, $\frac{1}{3}$ and $\frac{1}{2}$. The categories are ordered and hence to run the allocation, we need to assign appropriate numerical weights to each category. Then Likert's method under the assumption of normality of age gives, respectively, the weights -1.5, -.45 and .80 for the most favourable, favourable and the least favourable categories. Thus we can

take

$$Z = \begin{cases} -1.5 & \text{for a young patient} \\ -.45 & \text{for a middle-aged patient} \\ +.80 & \text{for an old-aged patient} \end{cases} .$$

with $p_{MF} = P(Z = -1.5) = \frac{1}{6}$, $p_F = P(Z = -.45) = \frac{1}{3}$, and $p_{LF} = P(Z = .80) = \frac{1}{2}$. For the evaluation of performance, we consider

- The overall and conditional allocation proportions to treatment A together with the standard errors.
- The overall and conditional proportions of benefited subjects together with the relevant standard errors.
- The power of a test of equality of treatment effects $H_0 : \alpha_A = \alpha_B, \beta_A = \beta_B$.

Now to assess the performance of the class of allocation procedures in small samples, we consider the allocation designs corresponding to each of the metrics γ_{k1}, γ_{k2} and γ_{k3} where the allocation design corresponding to the metric γ_{kj} is indicated by Dj . We, in addition, consider equiprobability randomization(denoted by $D4$), where either of the treatments is assigned with probability $\frac{1}{2}$ and conduct a simulation study and calculate the sample size necessary to attain 80% power using $D4$ for a test of equality of treatment effects under specific alternatives and evaluate the performance at this sample size. Specifically, for each fixed \mathbf{z} , we examine $\frac{N_{kn}(\mathbf{z})}{N_n(\mathbf{z})}$, the conditional proportion of allocations (CAP) to treatment k and $\frac{F_n(\mathbf{z})}{N_n(\mathbf{z})}$, the conditional proportion of subjects benefited (CBP) in the trial when the trial is conducted with n subjects. The conditional proportion of allocations to treatment A(denoted by CAP_A), expected allocation proportions to treatment A (denoted by EAP_A), the conditional and overall benefited proportions (indicated, respectively, by CBP and EBP) are also computed together with their standard errors. The relevant test statistic is W_n^* ,

which is the same as W_n except that the estimates used are obtained following the adaptive procedure with n_k replaced by the corresponding observed number N_{kn} . A larger value of W_n^* indicates possible rejection of the null hypothesis. We always take $\alpha_B = 121, \alpha_A = 119.5$ and vary σ, β_k and $\mathbf{p} = (p_{MF}, p_F, p_{LF})^T$. For the purpose of computation, we always choose β_k 's to ensure highest(lowest) expected response for the patients with most(least) favourable covariate categories. Now the covariate categories may be of equal as well as unequal importance. But equal proportions in \mathbf{p} is not convincing in practice and therefore, we always assume that p_{MF} is the highest. We provide the relevant performance measures assuming unequal weights for the covariate categories in Tables 1.

Remarks: As expected, the presence of treatment covariate interaction led to reasonable outcomes. Except for design D4, we observe the experience of the lowest and the highest allocation proportions by the subjects with the most favourable and the least favourable covariate category, respectively. That is, the allocation provides the safeguard to the subjects who actually needs the better therapy(i.e. patients with $Z=0.80$). The conditional proportions(both allocation and benefited) vary sensibly with both the covariate values and treatment effectiveness. However, the precision level(i.e. statistical power) decreases for all the CARA allocation designs(i.e. D1,D2 and D3) with an increase in β_B . In fact, the loss in power is minimum when $\beta_A = \beta_B$ and increases with a decrease in $\beta_A - \beta_B$. It is interesting to note that among the CARA allocation designs, the loss in precision(i.e. statistical power) compared to 80% mark is minimum for $D1$. This is not surprising if we look at the corresponding CAP(or EAP) values, where lower difference in $\beta_A - \beta_B$ indicates larger allocation towards the better treatment and with higher loss in power. However, such a loss can be compensated with the significant gain in the ethics measures. Thus we have derived a class of allocations which uses all available information, even the covariate information of the entering subject. Moreover it

Table 1: Performance Comparison at 80% power with $\alpha_A = 119.50, \alpha_B = 121, \mathbf{p} = (1/6, 1/3, 1/2)$.

| $(n, \beta_A, \beta_B, \sigma)$ | Design | $CAP_A(SD)$ with covariate profile | | | $EAP_A(SD)$ | | | $CBP(SD)$ with covariate profile | | | EBP(SD) | Power |
|---------------------------------|--------|------------------------------------|---------------|---------------|---------------|---------------|---------------|----------------------------------|---------------|-------|---------|-------|
| | | MF | F | LF | MF | F | LF | MF | F | LF | | |
| (81,-3,-3,1) | D1 | 0.599 (0.111) | 0.615 (0.091) | 0.622 (0.083) | 0.616 (0.056) | 0.012 (0.001) | 0.116 (0.063) | 0.968 (0.028) | 0.499 (0.052) | 0.787 | | |
| | D2 | 0.598 (0.112) | 0.617 (0.092) | 0.693 (0.086) | 0.651 (0.057) | 0.010 (0.001) | 0.118 (0.063) | 0.974 (0.025) | 0.502 (0.051) | 0.748 | | |
| | D3 | 0.501 (0.111) | 0.528 (0.091) | 0.682 (0.085) | 0.595 (0.056) | 0.014 (0.001) | 0.102 (0.059) | 0.973 (0.026) | 0.497 (0.052) | 0.773 | | |
| | D4 | 0.500 (0.110) | 0.500 (0.090) | 0.501 (0.076) | 0.501 (0.052) | 0.002 (0.001) | 0.096 (0.056) | 0.959 (0.032) | 0.489 (0.052) | 0.801 | | |
| (88,-3,-2.4,1) | D1 | 0.552 (0.112) | 0.609 (0.087) | 0.682 (0.079) | 0.633 (0.054) | 0.106 (0.001) | 0.122 (0.062) | 0.942 (0.035) | 0.491 (0.05) | 0.753 | | |
| | D2 | 0.547 (0.114) | 0.612 (0.089) | 0.767 (0.076) | 0.674 (0.054) | 0.012 (0.001) | 0.121 (0.060) | 0.957 (0.031) | 0.498 (0.052) | 0.661 | | |
| | D3 | 0.498 (0.107) | 0.528 (0.086) | 0.753 (0.077) | 0.629 (0.053) | 0.014 (0.011) | 0.106 (0.056) | 0.954 (0.032) | 0.492 (0.051) | 0.673 | | |
| | D4 | 0.500 (0.106) | 0.499 (0.086) | 0.500 (0.074) | 0.500 (0.049) | 0.003 (0.001) | 0.101 (0.055) | 0.910 (0.044) | 0.468 (0.051) | 0.798 | | |
| (112,-3,-1.8,1) | D1 | 0.482 (0.109) | 0.604 (0.082) | 0.753 (0.069) | 0.653 (0.051) | 0.014 (0.002) | 0.128 (0.056) | 0.918 (0.035) | 0.485 (0.045) | 0.623 | | |
| | D2 | 0.48 (0.111) | 0.606 (0.081) | 0.841 (0.06) | 0.696 (0.048) | 0.016 (0.002) | 0.127 (0.055) | 0.945 (0.030) | 0.498 (0.045) | 0.307 | | |
| | D3 | 0.500 (0.099) | 0.523 (0.078) | 0.819 (0.063) | 0.662 (0.046) | 0.015 (0.002) | 0.114 (0.053) | 0.939 (0.033) | 0.492 (0.045) | 0.364 | | |
| | D4 | 0.499 (0.099) | 0.501 (0.078) | 0.499 (0.066) | 0.499 (0.045) | 0.001 (0.002) | 0.105 (0.052) | 0.834 (0.051) | 0.439 (0.045) | 0.806 | | |
| (125,-3,-3,3/2) | D1 | 0.595 (0.102) | 0.605 (0.077) | 0.608 (0.071) | 0.605 (0.049) | 0.022 (0.01) | 0.188 (0.061) | 0.915 (0.035) | 0.506 (0.043) | 0.802 | | |
| | D2 | 0.597 (0.102) | 0.606 (0.078) | 0.663 (0.079) | 0.632 (0.052) | 0.024 (0.010) | 0.189 (0.062) | 0.923 (0.034) | 0.511 (0.043) | 0.784 | | |
| | D3 | 0.503 (0.096) | 0.531 (0.074) | 0.648 (0.075) | 0.583 (0.048) | 0.028 (0.010) | 0.172 (0.059) | 0.922 (0.035) | 0.504 (0.043) | 0.769 | | |
| | D4 | 0.499 (0.095) | 0.499 (0.074) | 0.501 (0.063) | 0.499 (0.043) | 0.002 (0.009) | 0.165 (0.058) | 0.899 (0.039) | 0.491 (0.043) | 0.804 | | |
| (134,-3,-2.4,3/2) | D1 | 0.542 (0.103) | 0.593 (0.075) | 0.652 (0.069) | 0.612 (0.048) | 0.033 (0.011) | 0.197 (0.060) | 0.888 (0.038) | 0.497 (0.042) | 0.769 | | |
| | D2 | 0.541 (0.104) | 0.594 (0.076) | 0.715 (0.075) | 0.643 (0.051) | 0.035 (0.011) | 0.197 (0.061) | 0.904 (0.037) | 0.505 (0.042) | 0.723 | | |
| | D3 | 0.501 (0.093) | 0.529 (0.073) | 0.693 (0.072) | 0.604 (0.047) | 0.033 (0.010) | 0.185 (0.059) | 0.898 (0.038) | 0.497 (0.042) | 0.736 | | |
| | D4 | 0.501 (0.094) | 0.501 (0.072) | 0.500 (0.061) | 0.501 (0.042) | 0.002 (0.012) | 0.177 (0.058) | 0.851 (0.044) | 0.472 (0.042) | 0.802 | | |
| (174,-3,-1.8,3/2) | D1 | 0.483 (0.096) | 0.58 (0.068) | 0.701 (0.061) | 0.621 (0.044) | 0.055 (0.014) | 0.209 (0.054) | 0.867 (0.036) | 0.493 (0.037) | 0.714 | | |
| | D2 | 0.481 (0.095) | 0.583 (0.069) | 0.771 (0.062) | 0.656 (0.045) | 0.057 (0.013) | 0.208 (0.054) | 0.891 (0.035) | 0.505 (0.037) | 0.564 | | |
| | D3 | 0.500 (0.085) | 0.526 (0.065) | 0.741 (0.061) | 0.626 (0.041) | 0.054 (0.013) | 0.199 (0.053) | 0.881 (0.037) | 0.498 (0.038) | 0.598 | | |
| | D4 | 0.501 (0.084) | 0.501 (0.064) | 0.500 (0.053) | 0.500 (0.037) | 0.005 (0.013) | 0.195 (0.052) | 0.794 (0.044) | 0.454 (0.037) | 0.799 | | |

Figures within the parentheses indicate the corresponding standard error

assigns subjects according to the importance of the covariate and therefore, has the ability to perform sensibly in real situations.

4. Concluding Remarks

The development of the current work is made for two treatments assuming continuous treatment outcomes. The corresponding development with several treatments for any type of responses incorporating some additional clinically relevant issues is left for future work.

References

- Atkinson, A.C. (1982). Optimum biased coin designs for sequential clinical trials with prognostic factors. *Biometrika*, **69**, 61-67.
- Azriel D., Mandel M. and Rinott Y. (2012). Optimal allocation to maximize power of two-sample tests for binary response. *Biometrika*, **99**, 101-113.
- Bandyopadhyay, U. and Biswas, A. (2001). Adaptive designs for normal responses with prognostic factors. *Biometrika*, **88**, 409-419.
- Begg, C.B. and Iglewicz, B. (1980). A treatment allocation procedure for sequential clinical trials. *Biometrics*, **36**, 81-90.
- Biswas, A., Huang, H.H. and Huang, W.T.(2006). Covariate-adjusted adaptive designs for continuous responses in a Phase III clinical trial: Recommendation for practice. *Journal of Biopharmaceutical Statistics* **16**, 227-239.
- Biswas, A. and Mandal, S. (2004). Optimal adaptive designs in phase III clinical trials for continuous responses with covariates. In *mODa 7 - Advances in Model-Oriented Design and Analysis*, (Eds. A. Di. Bucchianico, H.L.Álauter and H.P. Wynn), 51-58. Heidelberg: Physica-Verlag.
- Friedman, L.M., Furberg, C.D. and DeMets, D.L.(2010). *Fundamentals of Clinical Trials, Fourth Edition* Springer-Verlag, New York.
- Garret, H.E.(1966). *Statistics in Psychology and Education*. Longmans, Green.
- Hu, F. and Rosenberger, W.F.(2006). *The theory of response-adaptive randomization in clinical trials*. Wiley Interscience, New York.
- Jennison, C. and Turnbull, B.W.(2000) . *Group Sequential Methods with Applications to Clinical Trials*. Boca Raton, Florida: Chapman and Hall/CRC , Chapter 17.
- Seber, G.A.F and Lee. A.J.(2003) *Linear Regression Analysis, Second Edition* John Wiley & Sons, Inc., Hoboken, New Jersey.
- Shirayev, A.N.(1996) *Probability, Second Edition*(Translated by R.P. Boas) Springer, New York.
- Wong, W. K. and Zhu, W.(2008). Optimum treatment allocation rules under a variance heterogeneity model. *Statistics in Medicine* **27** 4581-4595.
- Zhang, L.X.,Hu , F., Cheung, S.H. and Chan, W.S.(2007). Asymptotic properties of covariate-adjusted response-adaptive designs. *Annals of Statistics* **35** 1166-1182.

Zhang, L.X. and Hu, F.(2009). A new family of covariate-adjusted response-adaptive designs and their properties. *Applied Mathematical Journal of Chinese Universities* **24** 1-13.

Appendix A

Proof of Result 2.1: We note that the ML estimator $\hat{\boldsymbol{\theta}}_{kn}$ is a solution to the equation

$$\frac{\partial \log \mathcal{L}_n(\boldsymbol{\theta}_k)}{\partial \boldsymbol{\theta}_k} = \mathbf{0}.$$

Since $\{\mathbf{Z}_j, j \geq 1\}$ are assumed uniformly bounded, under the normal response model it follows from the Corollary 3.1 of Zhang and Hu(2009) that for $\epsilon > 0$ small enough

$$\log \mathcal{L}_n(\boldsymbol{\theta}_k^*) < \log \mathcal{L}_n(\boldsymbol{\theta}_k)$$

with probability 1 whenever $\|\boldsymbol{\theta}_k^* - \boldsymbol{\theta}_k\| = \epsilon$ and n is large enough. This in turn implies

$$\hat{\boldsymbol{\theta}}_{kn} \rightarrow \boldsymbol{\theta}_k$$

almost surely as $n \rightarrow \infty$. Hence the strong consistency of the parameters under the response adaptive set up is established. \square

Proof of Result 2.2: It is easy to observe that $\hat{\boldsymbol{\eta}}_{kn}$, the ML estimate of $\boldsymbol{\eta}_k$, satisfies the equation

$$\sum_{j=1}^n \delta_{kj}(Y_{kj} - \hat{\boldsymbol{\eta}}_{kn}^T \boldsymbol{\xi}_j) \boldsymbol{\xi}_j = \mathbf{0},$$

which can be rewritten as

$$\sum_{j=1}^n \delta_{kj}(Y_{kj} - \boldsymbol{\eta}_k^T \boldsymbol{\xi}_j) \boldsymbol{\xi}_j + \sum_{j=1}^n \delta_{kj}(\boldsymbol{\xi}_j^T \boldsymbol{\eta}_k - \boldsymbol{\xi}_j^T \hat{\boldsymbol{\eta}}_{kn}) \boldsymbol{\xi}_j = \mathbf{0}.$$

Therefore, $\hat{\boldsymbol{\eta}}_{kn}$ satisfies the following equation

$$\frac{1}{n} \sum_{j=1}^n \delta_{kj} \boldsymbol{\xi}_j \boldsymbol{\xi}_j^T (\hat{\boldsymbol{\eta}}_{kn} - \boldsymbol{\eta}_k) = \frac{1}{n} \sum_{j=1}^n \delta_{kj}(Y_{kj} - \boldsymbol{\eta}_k^T \boldsymbol{\xi}_j) \boldsymbol{\xi}_j. \quad (A.1)$$

It can be observed that the matrices

$$A_j = \sum_{i=1}^j \frac{1}{i} \boldsymbol{\xi}_i \boldsymbol{\xi}_i^T \{ \delta_{ki} - E(\delta_{ki} | \mathcal{F}_{i-1}) \}, j \geq 1$$

and the vectors

$$\mathbf{b}_j = \sum_{i=1}^j \frac{1}{i} [\delta_{ki}(Y_{ki} - \boldsymbol{\eta}_k^T \boldsymbol{\xi}_i) \boldsymbol{\xi}_i - E \{ \delta_{ki}(Y_{ki} - \boldsymbol{\eta}_k^T \boldsymbol{\xi}_i) \boldsymbol{\xi}_i | \mathcal{F}_{i-1} \}], j \geq 1$$

are both martingales. Now noting the fact that $|\delta_{ki} - E(\delta_{ki} | \mathcal{F}_{i-1})| \leq 1$ and using the martingale property, we get

$$E[\text{Trace}(A_j A_j^T)] \leq E_{\mathbf{Z}_1}[\text{Trace}(\boldsymbol{\xi}_1 \boldsymbol{\xi}_1^T)]^2 \sum_{i=1}^{\infty} \frac{1}{i^2}$$

and

$$E[\text{Trace}(\mathbf{b}_j \mathbf{b}_j^T)] \leq \sigma_k^2 E_{\mathbf{Z}_1}[\text{Trace}(\boldsymbol{\xi}_1 \boldsymbol{\xi}_1^T)] \sum_{i=1}^{\infty} \frac{1}{i^2}, k = A, B$$

for every $j \geq 1$. Since $\{\mathbf{Z}_j, j \geq 1\}$ are uniformly bounded and the response variances are finite, the right hand members of the above two inequalities are both finite. Thus using L_2 -martingale convergence theorem together with Kronecker's lemma (see, for example, Theorem 4 of Shirayev(1996),pp.519), we have as $n \rightarrow \infty$

$$\frac{1}{n} \sum_{j=1}^n \left\{ \delta_{kj} \boldsymbol{\xi}_j \boldsymbol{\xi}_j^T - E(\delta_{kj} \boldsymbol{\xi}_j \boldsymbol{\xi}_j^T | \mathcal{F}_{j-1}) \right\} \rightarrow \mathbf{0}^{p+1 \times p+1}$$

and

$$\frac{1}{n} \sum_{j=1}^n \left[\delta_{kj} (Y_{kj} - \boldsymbol{\eta}_k^T \boldsymbol{\xi}_j) \boldsymbol{\xi}_j - E \left\{ \delta_{kj} (Y_{kj} - \boldsymbol{\eta}_k^T \boldsymbol{\xi}_j) \boldsymbol{\xi}_j | \mathcal{F}_{j-1} \right\} \right] \rightarrow \mathbf{0}$$

almost surely. Since

$$E(\delta_{kj} | \mathcal{F}_{j-1}, \mathbf{Z}_j) = \pi_{kj}(\hat{\boldsymbol{\theta}}_{j-1}, \mathbf{Z}_j),$$

we have

$$E(\delta_{kj} \boldsymbol{\xi}_j \boldsymbol{\xi}_j^T | \mathcal{F}_{j-1}) - E_{\mathbf{Z}_1} \left\{ \pi_{kj}(\hat{\boldsymbol{\theta}}_{j-1}, \mathbf{Z}_1) \boldsymbol{\xi}_1 \boldsymbol{\xi}_1^T \right\} = \mathbf{0}.$$

Further, as $\pi_k(\boldsymbol{\theta}, \mathbf{Z})$ is continuous in $\boldsymbol{\theta}$ for fixed \mathbf{Z} , it follows from the strong consistency of $\hat{\boldsymbol{\theta}}_n$ that as $n \rightarrow \infty$

$$\frac{1}{n} \sum_{j=1}^n E_{\mathbf{Z}_1} \left\{ \pi_{kj}(\hat{\boldsymbol{\theta}}_{j-1}, \mathbf{Z}_1) \boldsymbol{\xi}_1 \boldsymbol{\xi}_1^T \right\} \rightarrow I_{\pi k}(\boldsymbol{\theta})$$

almost surely, where

$$E_{\mathbf{Z}_1} \left\{ \pi_k(\boldsymbol{\theta}, \mathbf{Z}_1) \boldsymbol{\xi}_1 \boldsymbol{\xi}_1^T \right\} = I_{\pi k}(\boldsymbol{\theta}), k = A, B.$$

Hence, assuming $I_{\pi k}(\boldsymbol{\theta})$ to be non-singular, equation (A.1) implies

$$(\hat{\boldsymbol{\eta}}_{kn} - \boldsymbol{\eta}_k) - I_{\pi k}^{-1}(\boldsymbol{\theta}) \frac{1}{n} \sum_{j=1}^n \delta_{kj} (Y_{kj} - \boldsymbol{\eta}_k^T \boldsymbol{\xi}_j) \boldsymbol{\xi}_j \rightarrow \mathbf{0}$$

almost surely, $k=A, B$. Thus the required conditions of Theorem 2.1 of Zhang et al.(2007) are satisfied if we take

$$h_{kj}(Y_{kj}, \boldsymbol{\xi}_j) = I_{\pi k}^{-1}(\boldsymbol{\theta}) (Y_{kj} - \boldsymbol{\eta}_k^T \boldsymbol{\xi}_j) \boldsymbol{\xi}_j,$$

where $E \left\{ (Y_{kj} - \boldsymbol{\eta}_k^T \boldsymbol{\xi}_j) \boldsymbol{\xi}_j | \mathbf{Z}_j \right\} = \mathbf{0}$ for any $j, k=A, B$. In a similar fashion, it follows that as $n \rightarrow \infty$

$$(\hat{\sigma}_{kn} - \sigma_k) - I_{\pi k}^{*-1}(\boldsymbol{\theta}) \frac{1}{n} \sum_{j=1}^n \delta_{kj} \left\{ \frac{(Y_{kj} - \boldsymbol{\eta}_k^T \boldsymbol{\xi}_j)^2}{\sigma_k^3} - \frac{1}{\sigma_k} \right\} \rightarrow 0$$

almost surely, where

$$I_{\pi k}^*(\boldsymbol{\theta}) = \frac{2}{\sigma_k^2} E_{\mathbf{Z}_1} \left\{ \pi_k(\boldsymbol{\theta}, \mathbf{Z}_1) \right\}$$

is non-singular, $k=A, B$. Thus the required conditions of Theorem 2.1 of Zhang et al.(2007) are satisfied with

$$h_{kj}(Y_{kj}, \boldsymbol{\xi}_j) = I_{\pi k}^{*-1}(\boldsymbol{\theta}) \left\{ \frac{(Y_{kj} - \boldsymbol{\eta}_k^T \boldsymbol{\xi}_j)^2}{\sigma_k^3} - \frac{1}{\sigma_k} \right\}, j = 1, 2, \dots$$

Hence the result follows. □

Appendix B

Proof of Result 2.3: Writing

$$\frac{1}{n} \sum_{i=1}^n F_i = \frac{1}{n} \sum_{k=A,B} \sum_{i=1}^n \{\delta_{ki} F_{ki} - E(\delta_{ki} F_{ki} | \mathcal{F}_{i-1})\} + \frac{1}{n} \sum_{i=1}^n E(F_i | \mathcal{F}_{i-1}), \quad (B.1)$$

where $F_{ki} = I_{[Y_{ki} \leq c]}$ represents the failure indicator for the i -th patient with treatment $k, k=A, B$. Since $|\delta_{ki} F_{ki} - E(\delta_{ki} F_{ki} | \mathcal{F}_{i-1})| \leq 1$, we as in the proof of Result 2.2, apply L_2 -martingale convergence theorem on the martingale

$$M_j = \sum_{i=1}^j \frac{1}{i} \{\delta_{ki} F_{ki} - E(\delta_{ki} F_{ki} | \mathcal{F}_{i-1})\}, \quad j \geq 1$$

and get that

$$\frac{1}{n} \sum_{i=1}^n \{\delta_{ki} F_{ki} - E(\delta_{ki} F_{ki} | \mathcal{F}_{i-1})\} \rightarrow 0 \quad (B.2)$$

almost surely, $k=A, B$. Thus the convergence will be established if we can show that, as $n \rightarrow \infty$,

$$P(F_n = 1 | \mathcal{F}_{n-1}) \rightarrow \psi$$

almost surely. Moreover, since the covariates are iid, we have

$$E(F_n | \mathcal{F}_{n-1}) = E_{\mathbf{Z}_1} \left\{ \sum_{k=A,B} P(F_{kn} = 1 | \delta_{kn} = 1, \mathbf{Z}_1, \mathcal{F}_{n-1}) P(\delta_{kn} = 1 | \mathbf{Z}_1, \mathcal{F}_{n-1}) \right\}. \quad (B.3)$$

It can be easily seen that that

$$P(F_{kn} = 1 | \delta_{kn} = 1, \mathbf{Z}_1, \mathcal{F}_{n-1}) = P(Y_{k1} \leq c | \mathbf{Z}_1), \quad (B.4)$$

and, as in Zhang et al.(2007),

$$P(\delta_{kn} = 1 | \mathbf{Z}_1, \mathcal{F}_{n-1}) \rightarrow \pi_k(\boldsymbol{\theta}, \mathbf{Z}_1) \quad (B.5)$$

almost surely. Hence, by the dominated convergence theorem together with Kronecker's lemma, the required result follows from (B.1) through (B.2)-(B.5). □

Affiliation:

Firstname Lastname

Affiliation

Address, Country

E-mail: **name@address**

URL: **http://link/to/webpage/**

Austrian Journal of Statistics

published by the Austrian Society of Statistics

Volume VV

MMMMMM YYYY

http://www.ajs.or.at/

http://www.osg.or.at/

Submitted: yyyy-mm-dd

Accepted: yyyy-mm-dd