Optimal Clinical Trial Designs to Detect Treatment-Biomarker Interaction

Hongjian Zhu, Feifang Hu and Hongyu Zhao

University of Texas School of Public Health, University of Virginia, Yale University

hongjian.zhu@uth.tmc.edu

Abstract

Biomarkers play a crucial role in the design and analysis of clinical trials for personalized medicine. One major goal of these trials is to derive an optimal treatment scheme based on each patient's biomarker level. Although completely randomized trials may be employed, a more efficient design can be attained when patients are adaptively allocated to different treatments throughout the trial using biomarker information. Therefore, we propose a new adaptive allocation method based on using multiple regression models to study treatment-biomarker interactions. We show that this perspective simplifies the derivation of optimal allocations. Moreover, when implemented in real clinical trials, our method can integrate all the covariates which may not be related to the treatment-biomarker interaction into a single mode for a joint analysis. Our general idea can be applied to diverse models to derive optimal allocations. Simulation results show that both the optimal allocation and the proposed design can lead to a more efficient trial.

1. Problem

Consider the following **model**

$$Y_i = \beta_0 + \beta_1 Z_{i1} + \ldots + \beta_K Z_{iK} + \beta_T T_i + \beta Z_{i1} T_i + \epsilon_i,\ i = 1, \ldots, n, \tag{1}$$

- Y_i: Treatment response with independent errors $\epsilon_i \sim N(0, \sigma_i^2)$
- $Z_{i1}, \ldots, Z_{iK}$: Covariates.
- T_i: Treatment assignment taking value 1 or 0 for two treatments.
- $(\beta_0, \beta_1, \ldots, \beta_K, \beta_T, \beta)$: Unknown parameters.

Aim: To find the optimal allocation of patients to two treatments to maximize the power of testing the following hypothesis

$$H_0 : \beta = 0 \text{ versus } H_1 : \beta \neq 0. \tag{2}$$

and implement it through a randomized sequential design.

Remarks:

- We are interested in the interaction between treatments and a certain biomarker (Z_1) and its practical meaning in the context of personalized medicine that patients with different prognostic factors may respond to treatments differently.
- Z_1 is required to be a binary covariate, which may indicate the mutation status (widetype or mutated) of a gene or the expression level (high or low) of a gene or protein.
- $Z_2, \ldots, Z_K$ are expected to have influence on responses.
- The model could be either homoscedastic or heteroscedastic.

2. Optimal Allocations

(1) Homoscedastic models without interaction term.

We start with a simpler model:

$$Y_i = \beta_0 + \beta_1 Z_{i1} + \ldots + \beta_K Z_{iK} + \beta_T T_i + \epsilon_i,\ i = 1, \ldots, n, \tag{3}$$

$$H_0 : \beta_T = 0 \text{ versus } H_1 : \beta_T \neq 0. \tag{4}$$

The data is naturally divided into two groups based on treatment assignments,

$$E(Y_i^1) = \boldsymbol{Z}_i^1 \boldsymbol{\beta'} + \beta_T,\ i = 1, \ldots, n_1$$

and

$$E(Y_i^2) = \boldsymbol{Z}_i^2 \boldsymbol{\beta'},\ i = 1, \ldots, n_2$$

- $Y_i^j, j = 1, 2$: Responses for two independent groups 1 and 2.
- $\boldsymbol{Z}_i^j = (Z_{i1}^j, \ldots, Z_{iK}^j)^T$: Covariate vector for the ith patient in group j.
- n_j: The number of patients in group j.
- $\boldsymbol{\beta'} = (\beta_0, \beta_1, \ldots, \beta_K)^T$: Regression coefficients.

Idea: Maximize efficiency by minimizing the number of groups to be compared.

Intuition: If the mean value of the covariate part in group 1 is the same as that in group 2, i.e.,

$$\sum_{i=1}^{n_1} \boldsymbol{Z}_i^1 \boldsymbol{\beta'}/n_1 = \sum_{i=1}^{n_2} \boldsymbol{Z}_i^2 \boldsymbol{\beta'}/n_2 = \mu, \tag{5}$$

testing hypothesis (4) is equivalent to testing whether the mean ($\bar{y}^1$) of the samples in group 1 equals to that ($\bar{y}^2$) in group 2, i.e. $\mu + \beta_T = \mu$. In other words, we are performing the hypothesis test by comparing the means of two groups. On the contrary, if condition (5) is not satisfied, we have to test the hypothesis by comparing more groups of data, so that the power is likely lowered.

Theorem 1 *Consider model (3) and hypothesis (4), the optimal allocation for maximizing the power requires both of the following two conditions hold*

(A) $\sum_{i=1}^{n_1} \boldsymbol{Z}_i^1 \boldsymbol{\beta'}/n_1 = \sum_{i=1}^{n_2} \boldsymbol{Z}_i^2 \boldsymbol{\beta'}/n_2$,

(B) $n_1 = n_2$.

(2) Heteroscedastic models with interaction term

For model (1), we naturally divide the data into the following four groups,

$$\begin{aligned} E(Y_i^1) &= \beta_0 + \beta_2 Z_{i2}^1 + \ldots + \beta_K Z_{iK}^1 + \beta_1 + \beta_T + \beta \\ &= \boldsymbol{Z}_i^1 \boldsymbol{\beta''} + \beta_1 + \beta_T + \beta,\ i = 1, \ldots, n_1, \end{aligned} \tag{6}$$

$$\begin{aligned} E(Y_i^2) &= \beta_0 + \beta_2 Z_{i2}^2 + \ldots + \beta_K Z_{iK}^2 + \beta_T \\ &= \boldsymbol{Z}_i^2 \boldsymbol{\beta''} + \beta_T,\ i = 1, \ldots, n_2, \end{aligned} \tag{7}$$

$$\begin{aligned} E(Y_i^3) &= \beta_0 + \beta_2 Z_{i2}^3 + \ldots + \beta_K Z_{iK}^3 + \beta_1 \\ &= \boldsymbol{Z}_i^3 \boldsymbol{\beta''} + \beta_1,\ i = 1, \ldots, n_3, \end{aligned} \tag{8}$$

$$\begin{aligned} E(Y_i^4) &= \beta_0 + \beta_2 Z_{i2}^4 + \ldots + \beta_K Z_{iK}^4 \\ &= \boldsymbol{Z}_i^4 \boldsymbol{\beta''},\ i = 1, \ldots, n_4, \end{aligned} \tag{9}$$

- $Y_i^j, j = 1, 2, 3, 4$: Responses for these four independent groups.
- $\boldsymbol{Z}_i^j = (1, Z_{i2}^j, \ldots, Z_{iK}^j)^T$.
- n_j: The number of patients in group j.
- $\boldsymbol{\beta''} = (\beta_0, \beta_2, \ldots, \beta_K)^T$.
- $Var(\epsilon_i^j) = \sigma_j^2$,

Theorem 2 *Consider model (1) with specified variance of error as above and hypothesis (2), the optimal allocation for maximizing the power requires both of the following conditions hold*

(A) $\sum_{i=1}^{n_1} \boldsymbol{Z}_i^1 \boldsymbol{\beta''}/n_1 = \sum_{i=1}^{n_2} \boldsymbol{Z}_i^2 \boldsymbol{\beta''}/n_2 = \sum_{i=1}^{n_3} \boldsymbol{Z}_i^3 \boldsymbol{\beta''}/n_3 = \sum_{i=1}^{n_4} \boldsymbol{Z}_i^4 \boldsymbol{\beta''}/n_4$,

(B) $\frac{n_1}{n_1+n_3} = \frac{\sigma_1}{\sigma_1+\sigma_3}$ *and* $\frac{n_2}{n_2+n_4} = \frac{\sigma_2}{\sigma_2+\sigma_4}$

3. Design

Suppose patients come to clinical trial sequentially and n data points $(y_i, z_{i1}, \ldots, z_{iK}, T_i,\ i = 1, \ldots, n)$ have been observed. When the $(n+1)$th patient with covariates $(z_{(n+1)1}, \ldots, z_{(n+1)K})$ enters the trial, we take the following steps in assigning the current patient to one of the two treatments.

(A) Obtain the estimator $\hat{\boldsymbol{\beta}}_n''$ of parameter $\boldsymbol{\beta''} = (\beta_0, \beta_2, \ldots, \beta_K)$ and the estimator $\hat{\sigma}_i, i = 1, 2, 3, 4$ of standard deviations for the four groups by least squares.

(B) Count the number of patients in each of the four groups, i.e. (n_1, n_2, n_3, n_4).

(C) Assume the $(n+1)$th patient is assigned to treatment 1. Calculate the variance (VAR_1) of the four items $(\frac{\sum \boldsymbol{Z}_i^1 \hat{\boldsymbol{\beta}}_n''}{n_1}, \frac{\sum \boldsymbol{Z}_i^2 \hat{\boldsymbol{\beta}}_n''}{n_2}, \frac{\sum \boldsymbol{Z}_i^3 \hat{\boldsymbol{\beta}}_n''}{n_3}, \frac{\sum \boldsymbol{Z}_i^4 \hat{\boldsymbol{\beta}}_n''}{n_4})$.

(D) Suppose the $(n+1)$th patient is assigned to treatment 2. Calculate the corresponding variance (VAR_2) in the same way as in step (C). If $z_{(n+1)1} = 1$, go to step (E1), otherwise go to step (E2).

(E1) Calculate $D = w_1(VAR_1 - VAR_2) + w_2(\frac{n_1}{n_1+n_3} - \frac{\hat{\sigma}_1}{\hat{\sigma}_1+\hat{\sigma}_3})$, where $w_1, w_2 > 0$.

(E2) Calculate $D = w_1(VAR_1 - VAR_2) + w_2(\frac{n_2}{n_2+n_4} - \frac{\hat{\sigma}_2}{\hat{\sigma}_2+\hat{\sigma}_4})$.

(F) Assign the next patient to treatment 1 with the following probability

$$\psi = \begin{matrix} p & D < 0 \\ 0.5 & D = 0 \\ 1-p & D > 0 \end{matrix}, \tag{10}$$

where $0.5 < p < 1$.

4. Simulations

- Sample size: $n = 1000$.
- $(\beta_0, \beta_1, \beta_2, \beta_3, \beta_T) = (1, 10, 5, 3, 8)$.
- $Z_i, i = 1, 2, 3 \sim Bernoulli(0.5)$.
- $(\sigma_1, \sigma_2, \sigma_3, \sigma_4) = (1, 1, 2, 2)$.
- weight $w_1 = 1, w_2 = 1$.
- $p = 0.8$.
- Randomization: Complete Randomization (CR), Proposed Method (PM), Atkinson's (1982)
- All the results are based on 1000 replications.

Table 1: *Comparisons of various designs in the heteroscedastic case with all binary covariates*

Randomization	β	power	$\hat{\beta}$	VAR	$\rho_1 = \frac{n_1}{n_1+n_3}$	$\rho_2 \frac{n_2}{n_2+n_4}$
CR	0.5	0.705	0.504	0.035	0.500	0.500
			(0.201)	(0.028)	(0.022)	(0.023)
PM	0.5	0.748	0.498	0.012	0.334	0.335
			(0.190)	(0.014)	(0.014)	(0.014)
Atkinson	0.5	0.702	0.495	0.034	0.500	0.499
			(0.196)	(0.028)	(0.022)	(0.021)
CR	0.6	0.848	0.600	0.035	0.500	0.501
			(0.204)	(0.028)	(0.023)	(0.023)
PM	0.6	0.890	0.598	0.012	0.335	0.334
			(0.180)	(0.014)	(0.015)	(0.015)
Atkinson	0.6	0.856	0.608	0.034	0.500	0.500
			(0.201)	(0.028)	(0.022)	(0.022)

5. Conclusion

Both the derived optimal allocation and the proposed design are contributing factors for increasing the efficiency of clinical trials.

6. Reference

Atkinson, A. C. (1982). Optimum biased coin designs for sequential clinical trials with prognostic factors. *Biometrika* **69,** 61–67.