The Comparison of ACI and MCB Methods for Choosing a Set that Contains the Optimal Dynamic Treatment Regime

Rong Zhou

Abstract

Dynamic treatment regimes (DTRs) are used to operationalize treatment decision-making at different stages by clinicians. The decision rules are based on the time-varying patient characteristics. DTRs can provide more effective decisions compared with once-and-for-all decisions. Here we compare two approaches to identify a set of DTRs that includes the optimal DTR. The methods are: the ACI (Adaptive Confidence Intervals) method by Laber et al. with our modifications and the MCB (Multiple Comparisons with the Best) method by Ertefaie et al. We simulate data from four different scenarios to compare the MCB method and a modified version of the ACI method. By comparing the probabilities that the best DTR is included into the constructed set, and the average set size of each method in four different scenarios, we find that both methods include the true optimal DTR with a specified probability. The MCB method generally has a smaller average set size, indicating that it has more power in excluding inferior DTRs. Thus, we conclude that MCB method performs better in general.

Keyword: DTR, SMART, multiple comparison

1 Introduction

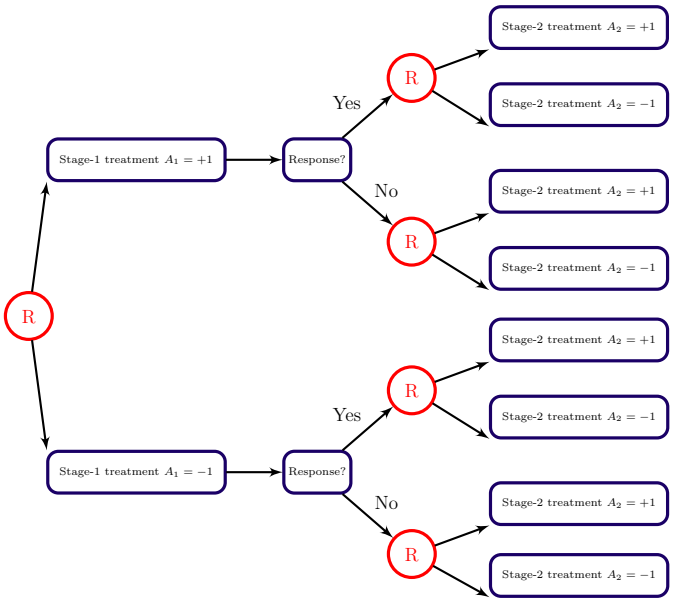
1.1 DTR and SMART

Individualized treatments enhance health care providers' management to different patients. Decisions at different stages are made based on patients' responses to their previous treatments. Since adaptive decisions at different stages are based on accumulative observations, they increase the effectiveness and the efficiency of the treatments compared with once-and-for-all decisions for patients. Dynamic Treatment Regimes (DTRs), also called adaptive interventions [1] or adaptive treatment strategies [5] [6] [10] [13] [14], are sequences of decision rules, which are based on patients' characteristics over time and provide treatment decisions. The rules take information before the current decision-making as inputs and sugggest the treatment decision at this stage. Thus, the rules make it possible for clinicians to consider the combination of factors at different stages. However, the selection of the optimal DTR among multiple ones requires researchers to develop methods and design experiments such that clinicians can provide the best outcome for patients.

In recent years a design known as sequential multiple assignment randomized trials (SMART) [10], has been applied to find the optimal DTR that has the maximum expected outcome among all possible DTRs. The SMART design starts with a randomization of individuals to available treatment options, then based on the intermediate outcome at each stage, individuals are re-randomized to receive the next stage's treatment. This method guarantees that the assignments of treatment among patients are not correlated and that the outcome is conditional on the patients' response overtime until the current stage [16].

SMART trials have three advantages over traditional methods: better validation for finding if one treatment is enhanced by prior treatments, better validation for detecting if an initial treatment is useful for upcoming decisions, and consideration of nonresponse cases at each stage to make patients a good representative of the population [7].

Figure 1: A Simple SMART Design



A simple SMART design as a two-stage treatment is shown as follows: we use a time-ordered trajectory

$$(X_1, A_1, X_2, A_2, Y)$$

to record the process. X_1 is called baseline information, which contains the initial condition of a patient before the first stage treatment. A_1 is the treatment on the first stage, and is coded as binary values {-1,1}. X_2 is the response to the first stage treatment before the second stage treatment, and is coded as binary values {-1,1} with 1 indicating "Yes" and -1 indicating "No". A_2 is the second stage treatment with binary values {-1,1}. Y is the final stage outcome of the regime and is coded such that a higher value is preferred.

1.2 Literature Reivew

So far, different methods have been proposed to help estimate the optimal DTR(s) using the SMART design. Murphy [9] proposed an approach called A-learning to construct and optimize regret functions using the dynamic programming method. The optimization of the regret function at each stage provides the optimal treatment regime. This method was later applied into various studies. Huang et al. [3] used the A-learning method and accelerated failure time models to identify the optimal treatment regime for recurrent diseases. Another method called Q-learning was applied by Murphy [11] and Zhao et al. [15] to obtain the optimal DTR using the dynamic programming technique. Parametric, semi-parametric, or nonparametric Q-functions can be applied to Q-learning and related methods. Backward induction helps find the optimal treatment regime. The Q-learning method is efficient for a single-stage setting if the simple linear regression model is specified, and it is widely used when there are a limited number of treatments at each stage [8]. In addition, Zhao et al. [16] proposed Backward Outcome Weighted learning (BOWL) and Simultaneous Outcome Weighted Learning (SOWL) to find the optimal treatment regime. By assuming optimal regimes are obtained in the following stages, Zhao et al. maximized the optimal decision rule at the current stage. These methods perform well in specific settings. All the techniques

mentioned above focused on finding the best DTR, which might not always be the primary goal in some clinical trials. As will be discussed in the following section, the Multiple Comparisons with the Best (MCB) method developed by Ertefaie et al. [2] identifies a set of DTRs that excludes DTRs with lower expected outcomes among multiple ones, and this provides more flexibility for clinicians to make decisions for their patients. All the methods mentioned above might have been tested in some situations, but they might not work well in all situations. Thus, the comparison between different methods to select a better one for each situation requires further exploration.

1.3 Goal

In this paper, we focus on two different methods introduced by two groups of researchers. There are the ACI (Adaptive Confidence Intervals) method proposed by Laber et al. [4] and the MCB method developed by Ertefaie et al. [2]. These two groups applied different methods that return either the optimal DTR or a set of DTRs. In this study, we used a modified version of the ACI method to obtain a set of DTRs, and the details of the modification can be found in section 2.2. The goals of this paper are: 1) to simulate data from different scenarios to test if the two methods include the optimal DTR(s) in their sets, and 2) to assess how well the two methods perform in excluding non-optimal DTRs.

The rest of this paper is organized as follows: Section 2 discusses the two methods developed by Laber et al. [4] and Ertefaie et al. [2], and the modification of the ACI method; Section 3 talks about simulation design and simulation results; Section 4 is the illustrative data analysis; Section 5 is discussion and conclusion.

2 ACI and MCB Methods

2.1 ACI Method by Laber et al. [4]

To find the optimal DTR, Laber el al. introduced a method of constructing adaptive confidence intervals (ACIs) for the Q-Learning algorithm's parameters in a regression model in 2011. After obtaining smooth, data-dependent upper and lower bounds for each parameter, confidence intervals of those parameters are constructed using the bootstrapping technique. Non-regularity in the Q-Learning method can occur during the maximization of the estimation process since it is a non-smooth and non-linear operation. Based on the setup in our study, non-regularity occurs when the effects of the two treatments are close at the second stage. This can further lead to a biased estimation of the effect of the first stage treatment. Laber et al. are targeted with the problem of the non-regularity of the Q-Learning method. The construction of high quality confidence intervals can efficiently reduce the bias of parameters that lead to huge mean square errors. In this way, although ACIs are still conservative for non-regular generative models, they are asymptotically valid.

Considering a two-stage treatment, the common Q-Learning algorithm functions are described below:

$$Q_2(h_2, a_2) \triangleq \mathbb{E}(Y_2|H_2 = h_2, A_2 = a_2)$$

$$Q_1(h_1, a_1) \triangleq \mathbb{E}(Y_1 + \max_{a_2 \in \{1,-1\}} Q_2(H_2, a_2)|H_1 = h_1, A_1 = a_1);$$

In the above equations, $Q_i(h_i,\ a_i)$, i = 1, 2 are called Q-functions. H_i is the joint expectation over the patient's history, and H_i includes information from the beginning up to the decision for the ith stage treatment A_i. Correspondingly, in our project, $H_1 = X_1$, and H_2 contains X_2.

In order to obtain the means in the functions, Laber et al. used a linear model. The linear model and estimator of each stage is as follows:

$$Q_i(h_i, a_i; \beta_i) = \beta_{i,0}^T h_{i,0} + \beta_{i,1}^T h_{i,1} 1_{a_i=1} \tag{1}$$

$$\hat{\beta}_i \triangleq arg \min_{\beta_i} \mathbb{P}_n(Y_i - Q_i(H_i, A_i; \beta_i))^2 \tag{2}$$

where $i = \{1, 2\}$.

Plug $\hat{\beta}_2$ into the predicted future reward, which is defined as:

$$\tilde{Y_1} \triangleq Y_1 + \max_{a_2 \in \{1,-1\}} Q_2(H_2, A_2; \hat{\beta}_2) = Y_1 + H_{2,0}^T \hat{\beta}_{2,0} + [H_{2,1}^T \hat{\beta}_{2,1}]_+$$

where $[H_{2,1}^T \hat{\beta}_{2,1}]_+$ represents the positive part of $H_{2,1}^T \hat{\beta}_{2,1}$.

By regressing $\tilde{Y_1}$ on H_1 and A_1 using (3), we can obtain $\hat{\beta}_1$.

The optimal DTR $\hat{a} = (\hat{a_1}, \hat{a_2})$ is:

$$\hat{a}_t(h_i) \triangleq arg \max_{a_i \in \{1,-1\}} Q_t(h_i, a_i; \hat{\beta}_i) \tag{3}$$

Laber et al. introduced a method using preliminary hypothesis testing (pretesting) and bootstrapping for individuals in the data to obtain confidence intervals for linear combinations of β_i. In this way, the obtained ACIs asymptotically cover when there is a treatment effect for all patients.

2.2 The Modification of the ACI Method

In order to construct a set of DTRs where the optimal DTR is included, we modified the ACI method. The modifications are as follows:

First of all, three types of hypothesis tests are conducted: the comparison of the two first stage treatments using the confidence interval obtained from the ACI method, the comparison of the two second stage treatments for the responders of the first stage using regular confidence intervals, and the comparison of the two second stage treatments for non-responders of the first stage using regular confidence intervals. For each type of the hypothesis tests, the significance level is set as $\alpha/3$, where $\alpha = 0.05$ in our study. For each type of the hypothesis tests, only one treatment is ruled out if it is significantly worse than the other, and neither is excluded when there is no significant difference between the two treatments. Then, a DTR is included in the recommended set if and only if none of its first stage treatment option, its second stage treatment option for responders, or its second stage

treatment option for non-responders is ruled out. Finally, a set is formed by including all the DTRs based on the above rules.

2.3 MCB Method by Ertefaie et al. [2]

Ertefaie et al. concentrated on methods of multiple comparison of group means, since this method is able to control the Type one error efficiently and hence increases the statistical power. They proposed a method to identify a set of means using Multiple Comparisons with the Best (MCB) technique, which compares the mean of each DTR with the best mean and recognizes DTRs that are significantly below the best, and then returns a set of DTRs that includes the optimal DTR.

Based on the analysis of the Extending Treatment Effectiveness of Naltrexone (EXTEND) study [12], the researchers first estimated the mean outcome under each DTR in the EXTEND study. The Marginal structural model (MSM) was applied in the estimation and two methods were used for the model: inverse probability weighting (IPW) and augmented inverse probability weighting (AIPW). The estimating equations are as follows:

1. Inverse probability weighting (IPW):

Inverse probability weighting (IPW) takes in a v-treatment trajectory τ and a function of β named θ_k. The estimation of the parameters in the model is based on the following equation:

$$\mathbb{P}_n \sum_{k=1}^{K} \dot{m}(\tau,\beta) w_2(V,\bar{A}_2,k)(Y - m(\tau;\beta)) = 0,$$

where $\dot{m}(\tau,\beta) = \partial m(\tau,\beta)/\partial\beta$ and

$$w_2(V,\bar{A}_2,k) = \frac{I_{DTR_{k,1}}(a_1) I_{DTR^v_{k,2}}(a_2)}{p(A_1{=}a_1)p(A_2{=}a_2|A_1{=}a_1,V{=}v)} \text{ for V = v and } \bar{A}_2 = \bar{a_2}$$

The treatment options for stage 2 based on V for DTR k is denoted as $DTR^v_{k,2}(a_2)$

2. Inverse probability weighting (AIPW):

The estimation equation of AIPW is as follows:

$$\mathbb{P}_n \sum_{k=1}^{K} \dot{m}(\tau,\beta)[w_2(V,\bar{A}_2,k)(Y - m(\tau;\beta)) - (w_2(V,\bar{A}_2,k) - w_1(\bar{A}_1,k))(\varphi^k_2(\bar{X}_2) - m(\tau;\beta))$$
$$- (w_1(A_1,k) - 1)(\varphi^k_1(\bar{X}_1) - m(\tau;\beta)) = 0,$$

where $\varphi^k_2(\bar{X}_2) = \mathbb{E}[Y|\bar{A}_2 \in DTR^v_{k,2}(\bar{X}_2)]$, $\varphi^k_1(\bar{X}_1) = \mathbb{E}[\varphi^k_2(\bar{X}_2)|\bar{A}_1 \in DTR^v_{k,1}(\bar{X}_1)]$
and

$$w_1(a_1,k) = \frac{I_{DTR_{k,1}}(a_1)}{p(A_1{=}a_1)} \text{ , for } A_1 = a_1$$

After obtaining the estimated best mean of each DTR, Ertefaie et al. generalized the MCB method to find DTRs that were not significantly different from the set of DTRs that

have the maximum expected outcome. Specifically, they constructed a set of indices (denoted as $\mathcal{B}$) such that $[K] := \arg\max_{k=1,\cdots,K} \theta_k$ belongs to the set with a probability no smaller than a pre-specified level, say, $1-\alpha$. θ_k represents the mean outcome of each DTR with k = 1 to K, and $\hat{\theta}_k$ is obtained by applying MSM and IPW to represent the estimation of each θ_k. This method ensures that the best DTR is contained in the set with the pre-specified probability, and provides more options for healthcare providers to select among DTRs that are within the acceptance range.

3 Simulation

3.1 Simulation Setup

Simulations in this article are constructed based on a two-stage SMART design, and the generative model for final outcome Y is set as follows:

$$Y = \gamma_1 + \gamma_2 X_1 + \gamma_3 A_1 + \gamma_4 X_1 A_1 + \gamma_5 A_2 + \gamma_6 X_2 A_2 + \gamma_7 A_1 A_2 + \epsilon, \epsilon \sim N(0,1)$$

where Y is the final stage outcome; $A_i = \{1, -1\}$ is the ith stage treatment decision for i = 1, 2; $X_i = \{1, -1\}$ is the ith stage outcome for i = 1, 2.

This setup allows different scenarios to be constructed by using different sets of γ vector $(\gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5, \gamma_6, \gamma_7)$ and δ vector (δ_1, δ_2) that determines $X_2|X_1, A_1$ as $X_2|X_1, A_1 \sim$ Bernoulli(expit($\delta_1 X_1 + \delta_2 A_1$)), expit(x) $= e^x/(1+e^x)$. For example, if the γ vector is $(0, 0, 0.1, 0, 0.1, 0, 0)$ and the δ vector is $(0, 0)$, the model becomes $Y = 0.1A_1 + 0.1A_2 + \epsilon$. In order to select the best DTR and obtain the maximized expected final outcome Y, both A_1 and A_2 have to be set as 1. Thus the optimal DTR is obtained: we choose $A_1 = 1$ at the first stage and always choose $A_2 = 1$ at the second stage regardless of X_1.

The ways to determine whether the modified ACI or the MCB is a better method in different scenarios are: 1) by comparing the probabilities that the best DTRs are included into the constructed set, and 2) by comparing the average set size from each method, which is the sum of probabilities of including each of the DTRs for each method. The former tests if the two methods are able to include the optimal DTR with the pre-specified probability, and the latter can be used as an additional gauge to test how good the two methods are in excluding non-optimal DTRs. The method that both reaches the desired probability of containing the optimal DTR and has a relatively smaller average set size is the better method in this setting.

Four different scenarios are constructed in this paper based on the above rules, and the results are presented in the following subsection. Three scenarios with the occurrences of non-regularity are constructed by setting the two treatments at the second stage to be equally best in all or some DTRs. A scenario with no non-regularity is also constructed when there are two optimal DTRs. In the following four scenarios, the δ vector is set to be (0,0), indicating that the probability of a patient responding to a treatment is 0.5.

3.2 Scenarios and Results

3.2.1 Scenario One

In scenario one, all eight DTRs are equally optimal. In this scenario, different sample sizes in each setting are used. In addition, slightly altered γ_3 values are assigned as well, and as γ_3 changes from 0 to positive values, the optimal DTRs changes to when the first stage treatment selection is $A_1 = 1$. In Table 1, We find that the MCB method performs better than the modified ACI method.

Table 1: Results of Scenario One

Number of iteration, Sample size	γ vector setting	Best DTR index	Best DTR Probabilities for the modified ACI	Best DTR Probabilities for the MCB	AVG set size for the modified ACI	AVG set size for the MCB
10000 150	(0,0,0,0,0,0,0,0)	1-8	0.96 0.97 0.96 0.96 0.96 0.96 0.96 0.96	0.96 0.94 0.94 0.94 0.94 0.98 0.97 0.94	7.70	7.52
10000 300	(0,0,0,0,0,0,0,0)	1-8	0.96 0.96 0.96 0.97 0.97 0.96 0.96 0.96	0.94 0.95 0.95 0.95 0.94 0.94 0.94 0.95	7.70	7.56
10000 600	(0,0,0,0,0,0,0,0)	1-8	0.96 0.96 0.96 0.97 0.97 0.96 0.96 0.96	0.94 0.95 0.95 0.95 0.95 0.95 0.95 0.95	7.71	7.57
2000 150	(0,0,0.01,0,0,0,0)	1-4	0.97 0.97 0.96 0.96	0.94 0.95 0.95 0.94	7.70	7.47
2000 150	(0,0,0.05,0,0,0,0)	1-4	0.97 0.97 0.96 0.96	0.95 0.96 0.96 0.96	7.69	7.37
2000 150	(0,0,0.1,0,0,0,0)	1-4	0.97 0.97 0.96 0.96	0.96 0.97 0.97 0.97	7.65	7.03
2000 150	(0,0,0.2,0,0,0,0)	1-4	0.97 0.97 0.96 0.96	0.97 0.97 0.97 0.97	7.33	5.76
2000 150	(0,0,0.3,0,0,0,0)	1-4	0.97 0.97 0.96 0.96	0.97 0.97 0.97 0.97	6.47	4.52

3.2.2 Scenario Two

In scenario two, the final outcomes are different by the second stage decision A_2 if the first stage decision A_1 is 1, but are the same if the first stage decision A_1 is -1. In Table 2, we find that both methods have similar performance, with a slightly better result from MCB.

Table 2: Results of Scenario Two

Iteration, Sample size	γ vector setting	Best DTR index	Best DTR Probabilities for the modified ACI	Best DTR Probabilities for the MCB	AVG set size for the modified ACI	AVG set size for the MCB
10000 150	(0,0,-0.5,0,0.5,0,0.5)	1 5 6 7 8	0.98 0.96 0.96 0.96 0.96	0.97 0.96 0.96 0.96 0.96	4.83	4.82
10000 300	(0,0,-0.5,0,0.5,0,0.5)	1 5 6 7 8	0.98 0.96 0.96 0.96 0.96	0.97 0.96 0.97 0.97 0.97	4.83	4.83
10000 600	(0,0,-0.5,0,0.5,0,0.5)	1 5 6 7 8	0.99 0.97 0.96 0.96 0.96	0.97 0.97 0.97 0.97 0.97	4.84	4.83
2000 150	(0,0,-0.49,0,0.51,0,0.51)	1	0.99	0.98	4.83	4.80
2000 150	(0,0,-0.5,0,0.5,0,0.5)	1 4	0.99 0.97	0.97 0.97	4.82	4.81

3.2.3 Scenario Three

In scenario three, responders to the first stage have the same final outcomes to A_2, but non-responders have different expected final outcomes to A_2. In Table 3, we find that the MCB method performs better than the modified ACI method.

Table 3: Results of Scenario Three

Iteration, Sample size	γ vector setting	Best DTR index	Best DTR Probabilities for the modified ACI	Best DTR Probabilities for the MCB	AVG set size for the modified ACI	AVG set size for the MCB
2000 150	(0,0,0,0,0.5,-0.5,0)	1 3 5 7	0.98 0.97 0.97 0.97	0.97 0.97 0.96 0.96	3.89	3.88
2000 150	(0,0,0.01,0,0.5,-0.5,0)	1 3	0.98 0.97	0.98 0.97	3.89	3.88

2000 150	(0,0,0.05,0,0.5,-0.5,0)	1 3	0.99 0.98	0.98 0.98	3.88	3.85
2000 150	(0,0,0.1,0,0.5,-0.5,0)	1 3	0.99 0.98	0.99 0.98	3.85	3.77
2000 150	(0,0,0.2,0,0.5,-0.5,0)	1 3	0.99 0.98	0.99 0.99	3.60	3.35
2000 150	(0,0,0.3,0,0.5,-0.5,0)	1 3	0.99 0.98	0.99 0.99	3.16	2.79

3.2.4 Scenario Four

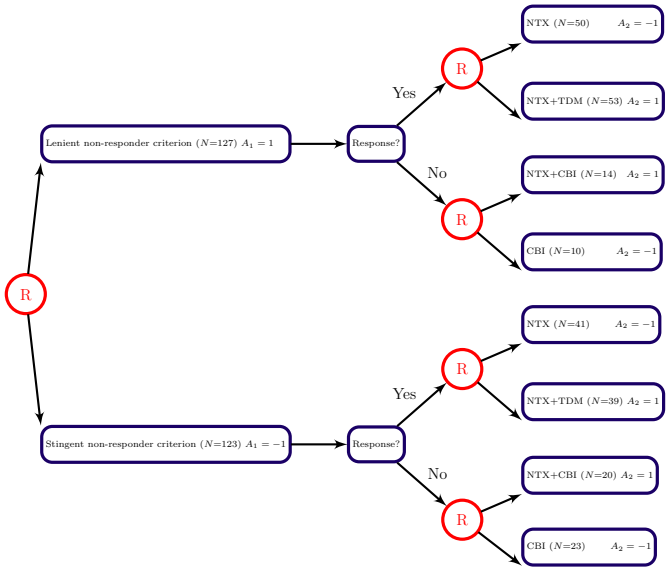
In scenario four, we manipulated the γ vector such that only one or two optimal DTRs will be obtained in this case, and there is no non-regularity. In Table 4, we find that the modified ACI method performs better than the MCB method. Because there is no non-regularity in this scenario, the modified ACI method will not be conservative compared to its performance in the first three scenarios. Also, since there are only two optimal DTRs in this scenario, and the MCB method still uses the critical value for comparing all eight DTRs when it chooses a set in this case, the result of the MCB method will be conservative.

Table 4: Results of Scenario Four

Iteration, Sample size	γ vector setting	Best DTR index	Best DTR Probabilities for the modified ACI	Best DTR Probabilities for the MCB	AVG set size for the modified ACI	AVG set size for the MCB
2000 150	(0,0,-1,0,-1,0,-0.5)	4 8	0.98 0.98	0.99 0.99	1.96	1.98
2000 150	(0,0,-0.99,0,-1,0,-0.5)	4	0.98	0.99	1.96	1.98
2000 150	(0,0,-0.45,0,-1,0,-0.5)	4	0.99	0.99	1.95	1.98
2000 150	(0,0,-0.4,0,-1,0,-0.5)	4	1.00	1.00	1.93	1.96
2000 150	(0,0,-0.51,0,-1,0,-0.5)	8	0.98	0.99	1.96	1.98
2000 150	(0,0,-0.55,0,-1,0,-0.5)	8	0.99	1.00	1.95	1.98
2000 150	(0,0,-0.6,0,-1,0,-0.5)	8	1.00	1.00	1.92	1.96

4 Illustrative Data Analysis

In this project, we use the Extending Treatment Effectiveness of Naltrexone (EXTEND) study [12] to illustrate the comparison of the two methods. The EXTEND study is a 24-week, two-stage study. Naltrexone (NTX) is used in the experiment as a treatment to prevent the relapse of alcoholism. The initial sample size was 302 and 250 patients in total were re-randomized at the second stage. The design is shown below, and there are in total eight DTRs in this experiment. In the EXTEND study, the final outcomes are measured as the percentage of heavy drinking days and the percentage of the drinking days over the last two months of the EXTEND study, and lower percentages represent better outcomes. Detailed information about the EXTEND study can be found in Lei et al. in 2012 [7].



We use the ten imputed datasets from the EXTEND data. The missing data were inputed using standard imputation methods (For details please email wutiansh@umich.edu). For the modified ACI method, the output set of DTRs in each dataset includes all eight DTRs. The MCB method, however, excludes some of the DTRs in some cases. Specifically, among the ten datasets, four of them include eight DTRs, three of them include seven DTRs and the remaining three include six DTRs. As a result, the MCB method in this study can exclude some of the relatively worse DTRs among the eight DTRs, while the modified ACI method fails to exclude DTRs in any of the datasets. In this study, the estimated effects of the eight DTRs are close, indicating that it is likely to be the setting of Scenario One in our simulation study, where MCB has better power.

5 Discussion and Conclusion

Based on the simulation results from the four different scenarios, we find that the MCB method performs better than the modified ACI method in general. Among the four simulation scenarios, the MCB method has a better performance three out of four times compared to the modified ACI method. The modified ACI method performs better in the scenario where there is no non-regularity, but the MCB method has higher performance in the other scenarios.

One possible explanation for the result is that the modified ACI method can possibly be improved by choosing better significance levels for different hypothesis tests. In our study, the significance levels for all three hypothesis tests are set to be $\alpha/3$, where $\alpha = 0.05$. We will further investigate the possibility of combining the two methods so that the new method can improve the performance.

6 Acknowledgments

We thank Dr. David Oslin at University of Pennsylvania for providing the EXTEND study data. We also thank Professor Susan Murphy from the Department of Statistics, University of Michigan for valuable comments. Thank Tianshuang Wu for revision and support.

References

[1] Daniel Almirall, Inbal Nahum-Shani, Nancy E Sherwood, and Susan A Murphy. Introduction to smart designs for the development of adaptive interventions: with application to weight loss research. *Translational behavioral medicine*, 4(3):260–274, 2014.

[2] Ashkan Ertefaie, Tianshuang Wu, Kevin G Lynch, and Inbal Nahum-Shani. Identifying a set that contains the best dynamic treatment regimes. *Biostatistics*, page kxv025, 2015.

[3] Xuelin Huang, Jing Ning, and Abdus S Wahed. Optimization of individualized dynamic treatment regimes for recurrent diseases. *Statistics in medicine*, 33(14):2363–2378, 2014.

[4] Eric B Laber, Min Qian, Daniel J Lizotte, William E Pelham, and Susan A Murphy. Statistical inference in dynamic treatment regimes. revision of univ. of michigan. *Statistics Department Technical Report*, 506, 2011.

[5] Philip W Lavori and Ree Dawson. A design for testing clinical strategies: biased adaptive within-subject randomization. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 163(1):29–38, 2000.

[6] Philip W Lavori and Ree Dawson. Adaptive treatment strategies in chronic disease. *Annual review of medicine*, 59:443, 2008.

[7] H Lei, I Nahum-Shani, K Lynch, D Oslin, and SA Murphy. A "smart" design for building individualized treatment sequences. *Annual review of clinical psychology*, 8, 2012.

[8] Ying Liu, Zeng Donglin, and Wang Yuanjia. Use of personalized dynamic treatment regimes (dtrs) and sequential multiple assignment randomized trials (smarts) in mental health studies. *Shanghai archives of psychiatry*, 26(6):376, 2014.

[9] Susan A Murphy. Optimal dynamic treatment regimes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(2):331–355, 2003.

[10] Susan A Murphy. An experimental design for the development of adaptive treatment strategies. 2005.

[11] Susan A Murphy, David W Oslin, A John Rush, and Ji Zhu. Methodological challenges in constructing effective treatment sequences for chronic psychiatric disorders. *Neuropsychopharmacology*, 32(2):257–262, 2007.

[12] David W Oslin. Managing alcoholism in people who do not respond to naltrexone (extend), 2005.

[13] Peter F Thall, Randall E Millikan, Hsi-Guang Sung, et al. Evaluating multiple treatment courses in clinical trials. *Statistics in medicine*, 19(8):1011–1028, 2000.

[14] Peter F Thall, Hsi-Guang Sung, and Elihu H Estey. Selecting therapeutic strategies based on efficacy and death in multicourse clinical trials. *Journal of the American Statistical Association*, 97(457), 2002.

[15] Ying-Qi Zhao and Eric B Laber. Estimation of optimal dynamic treatment regimes. *Clinical Trials*, 11(4):400–407, 2014.

[16] Ying-Qi Zhao, Donglin Zeng, Eric B Laber, and Michael R Kosorok. New statistical learning methods for estimating optimal dynamic treatment regimes. *Journal of the American Statistical Association*, (just-accepted):00–00, 2014.