MODEL-BASED APPROACH FOR DETERMINING OPTIMAL DYNAMIC TREATMENT REGIMES

by

Bing Yu

A Dissertation

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Doctor of Philosophy

Department of Statistics

West Lafayette, Indiana

December 2021

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Bruce A. Craig, Co-Chair

Department of Statistics

Dr. Michael Y. Zhu, Co-Chair

Department of Statistics

Dr. Xiao Wang

Department of Statistics

Dr. Min Zhang

Department of Statistics

Approved by:

Dr. Jun Xie

ACKNOWLEDGMENTS

First and foremost, I am extremely grateful to my advisors, Professor Bruce Craig and Professor Michael Yu Zhu. Not only has their broad and deep knowledge of statistics brought my research to a higher level, but their tremendous encouragement and continuous support have inspired and motivated me to overcome the many challenges I experienced throughout my Ph.D. study. It has been a long and winding road for me. I would not have made it without their invaluable advice and patience. I feel very lucky to have them as my advisors.

I would also like to thank my committee members, Professor Xiao Wang and Professor Min Zhang for their friendly guidance and insightful suggestions on my research. Many sincere thanks are given to the faculty and staff members in the Department of Statistics, as well as my fellow graduate students. Particularly, I would like to thank Douglas Crabill for his help with computing resources, Professor Hao Zhang for his encouragement, Professor Jun Xie for her support when I first started at Purdue, and Patti Foster for her coordination of my off-campus arrangements.

A special thank you to Professor Hyunyi Cho at The Ohio State University, for her RA support in the last two years while I was on campus. I also would like to thank my current supervisors at work, for their understanding and encouragement.

Last but not the least, I would like to express my gratitude to my husband, my daughter, and my parents, who supported me emotionally and financially, with their unconditional love.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

CCM	Chronic care model
DTR	Dynamic treatment regime
CATE	Conditional average treatment effect
SUTVA	Stable unit treatment value assumption
NUC	No unmeasured confounders
BCAWS	Biased coin adaptive within-subject
SMART	Sequential multiple assignment randomized trial
RL	Reinforcement learning
IPTW	Inverse probability of treatment weight
OWL	Outcome weighted learning
MLE	Maximum likelihood estimator
EM	Expectation–maximization
Q-Q	Quantile-Quantile

ABSTRACT

Dynamic treatment regimes (DTRs) are often considered for the medical care of chronic diseases and complex conditions. They consist of multistage treatment decisions, each based on the individual's health information and their treatment and response history. In this dissertation, we consider this setting with binary responses (i.e., either respond favorably or unfavorably to a treatment) and highlight one type of heterogeneity, specifically the existence of subgroups of patients who respond favorably to only a distinct subset of study treatments.

Currently, most works employ model-free approaches to find the optimal DTR. In contrast, we propose a model-based approach, which focuses more on describing heterogeneity in treatment responses. We first consider the scenario when baseline covariates are not included. A mixture of mixed logit models is proposed along with an EM alogorithm to estimate these subgroup proportions and the probabilities of a favorable response. We describe how an optimal dynamic treatment regime can be determined given the model information. We also discuss the necessary identifiability conditions (i.e., what sets of parameters are necessary for DTR determination).

Then, we extend the proposed model to incorporate baseline covariates. Specifically, we include certain baseline covariates in the logistic model for the probabilities of a favorable response and develop a multivariate Bernoulli model to incorporate the remaining covariates in the determination of subgroup proportions. Furthermore, time effects are considered in the model to allow for a potential overall decline in response effectiveness over time.

In each setting, simulation studies are performed to demonstrate the effectiveness of the proposed method in both parameter and DTR estimation. We also compare our approach with another competing method, Q-learning, and provide the scenarios when our mixture model outperforms Q-learning in terms of finding the optimal DTR.

1. INTRODUCTION

1.1 Overview of Dynamic Treatment Regime

It is widely recognized that a large portion of the variability in treatment response is the result of different individual characteristics, such as demographic information and genomic markers. For example, postmenopausal women with lymph node-negative breast cancer benefit substantially from adjuvant chemotherapy if their cancer is estrogen receptor (ER)-negative, while patients who are ER-positive receive no benefit [1]. Similarly, researchers have found racial and ethnic differences in drug response in a total of 167 new drugs approved by FDA between 2008 and 2013 [2]. Because all treatments are not equally efficacious to all patients, there has been a great interest since the early 1950s in discovering these heterogeneities in treatment response and identifying the subgroups who will benefit [3]. This has led to what today is called personalized medicine, which is a medical model that tailors medical interventions to subgroups of patients based on their individual characteristics.

The purpose of personalized medicine is to develop decision rules that result in individualized therapies that best match (in terms of outcome) each patient in the population [3]. These therapies can either be single-stage or multistage. In the single-stage setting, the goal is to assign the one treatment that maximizes the benefit to each patient. This decision is based on the available baseline information (e.g., medical history, demographics, genetic markers). The multistage setting, on the other hand, typically involves medical care for chronic diseases and complex conditions, where long-term care, such as multistage clinical and behavioral interventions and continuous follow-ups, are recommended. While one could treat each stage separately, attempting to maximize the benefit, it is often more beneficial to consider the sequence of decisions collectively. Determining the set of optimal decision rules in this multistage setting is the focus of this dissertation.

These multistage adaptive treatment strategies are also called dynamic treatment regimes (DTRs) [4], [5]. Under a DTR, a set of decision rules dictate what treatment to provide a patient at each assessment stage, with the consideration of achieving the optimal long-term outcome instead of the seemingly best intermediate outcomes. These rules incorporate the

patient's demographics, clinical information, and treatment history. Thus, a DTR both operationalizes and personalizes the clinical decision process.

For medical settings involving periodic evaluations and the assignment of treatment from a set of treatments at each stage, many DTRs can be considered. It is of practical importance to evaluate and compare these DTRs and select the DTR that has optimal properties.

Consider a T-stage DTR $D = (d_1, ..., d_T)$, where d_t is the decision rule at the t^{th} stage. The action A_t is the treatment assigned at the t^{th} stage based on d_t and Y_t is the response to A_t. The collection of possible sequences of treatments and responses are referred to as the trajectories of D. Each is commonly expressed as:

$$\boldsymbol{X}, A_1, Y_1, ..., A_T, Y_T$$

where $\boldsymbol{X}$ is the set of baseline covariates.

There are numerous different trajectories because of patient heterogeneity and chance variation in the response. Each of these trajectories can be scored according to a utility or desirability function $r(\boldsymbol{X}, A_1, Y_1, ..., A_T, Y_T)$, where a higher score indicates a more favorable overall outcome. The value of a DTR is then defined to be the expected utility score. For a T-stage trajectory, the value is written as:

$$V(D) = E(r(\boldsymbol{X}, A_1, Y_1, ..., A_T, Y_T)).$$

The goal of DTR research is to discover the sequence of decision rules that maximizes this expected utility score.

Even when numeric, patients' responses to treatments are often classified as favorable or unfavorable. For example, in an alcohol addiction management study [6], a favorable response is that a participant has no more than two heavy-drinking days in the two months post treatment. In another study concerning prostate cancer [7], the favorable response is at least an 40% reduction in a prostate-specific antigen (PSA). It is these binary responses to treatments that we focus on in this dissertation.

1.2 Literature Review

An increasing number of research activities on DTRs have emerged in the past decade [6], [8]. In this overview, we discuss the data sources used in DTR research, describe several commonly-used methods aimed at finding the optimal DTR, and review closely connected lines of research that focus on describing the heterogeneity in treatment effects across patients.

1.2.1 Data for DTR studies

In order to compare different DTRs, one has to first quantify the causal relationship between DTRs and the associated values [8]. Potential outcomes, also called counterfactuals, provide a way to formalize this problem. The framework was introduced by Neyman (1923) [9] and then extended by Rubin (1974, 1978) [10], [11] and Robins (1985) [12]. In terms of DTR research, the potential outcome of a DTR is the trajectory observed if the patient had followed that DTR.

For example, suppose that an individual could receive one of two possible DTRs: D and D'. We observe the trajectory associated with the DTR assigned to the patient, while other unobserved trajectories are the counterfactuals. Though it is not possible to evaluate the causal relationship between the DTR and its potential outcome at the individual level, one can still estimate the values of potential trajectories at the population level. The utility score measures the desirability of possible trajectories. Let the values $V(D)$ and $V(D')$ be the expected value of utility scores of the trajectories if this individual had followed DTR D and D' respectively. The optimal DTR can be determined by selecting the DTR that results in the highest value, or by comparing the values of following one DTR instead of the other.

Both observational and experimental studies have been used to evaluate DTRs. In observational studies, subjects are assigned to treatments based on decisions made by physicians. The observational study is suitable for the research of DTR when the axiom of consistency is satisfied, i.e., the potential outcome and the observed outcome are consistent [13]. This requirement is valid under a variety of conditions, such as the stable unit treatment value assumption (SUTVA) [14], and the assumption of no unmeasured confounders (NUC) [13].

SUTVA assumes that a subject's outcome is not influenced by others' treatment assignments, which is usually reasonable. On the other hand, it is hard to satisfy NUC since it requires measuring all sources of confounders. Therefore, the validity of the axiom of consistency is often questionable in observational studies.

On the other hand, randomization can balance both observed and unobserved confounders. Therefore, an experimental study, when appropriately designed, does not suffer from confounding and other biases [8], and should be preferred for evaluating DTRs. Currently, a widely-used experimental study design for evaluating DTRs is the Sequential Multiple Assignment Randomized Trial (SMART). It was first introduced by Lavori and Dawson (2000) [15] and called a biased coin adaptive within-subject (BCAWS) design. Under such a design, each patient is randomly assigned to one of the available treatments at the initial stage, and at subsequent stages, possible re-randomizations are preformed based on the individual's treatment and response history. Figure 1.1 shows an example of two-stage SMART for children affected by ADHD [16]. Randomization happens at the first stage and when patients don't respond to the first treatment in the second stage. Practical design considerations, like dropout, were later discussed by Lavori and Dawson (2004) [4] and a primary framework for analysis of SMART was provided by Murphy (2005) [17].

1.2.2 Statistical Methods for Finding the Optimal DTR

Developing the optimal DTR involves making multistage decisions, each based on the individual's current stage information and history, in order to maximize a future numerical outcome. It resembles the classic problem in reinforcement learning (RL), which is learning what to do–how to map situations to actions, aiming to maximize a numerical cumulative reward in the end [18].

Figure 1.2 illustrates the feedback loop of RL. At each stage, the agent reviews the state of the environment and takes an action based on its policy. The environment then reacts to the action and that results in a reward and the next state. In relation to DTR research, the agent's policy can be viewed as a DTR. The action is the treatment and the environment is similar to the mechanism of how the individuals react to the treatments (i.e.,

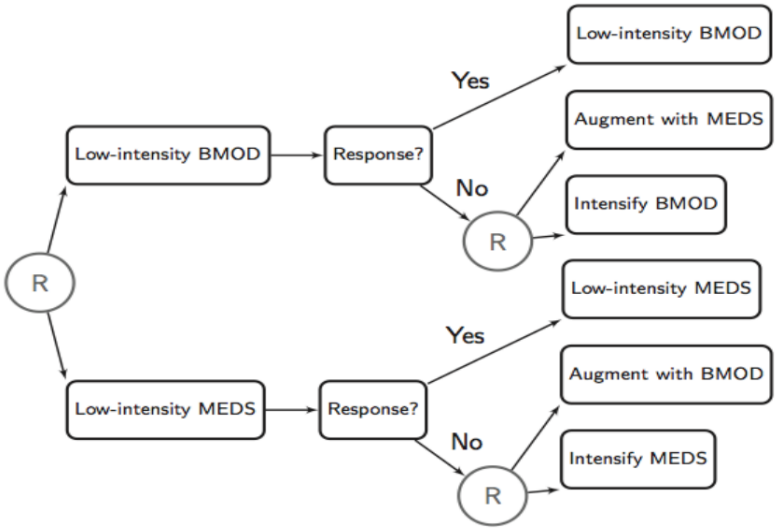

Figure 1.1. Example of a two-stage SMART for children affected by ADHD. Taken from the paper "Robust Hybrid Learning for Estimating Personalized Dynamic Treatment Regimens" [16]

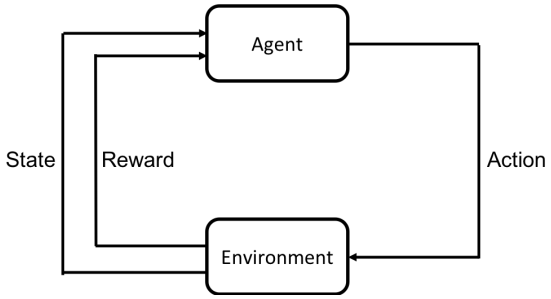

Figure 1.2. The feedback loop of reinforcement learning

the outcomes). Like states and rewards, covariates and previous responses to treatments are taken into account when determining the next treatment. Although there are some distinctions between traditional RL and the research of DTR [8], the probabilistic framework for a general finite-horizon RL is still suitable.

A vast number of different methods have been used to evaluate DTRs and find the optimal one. The majority of these approaches employ model-free methods, which are explicitly trial-and-error learning procedures. The word model-free in RL means that these methods make no attempt to mimic the behavior of the environment, which is similar to not modeling the

response to the treatment assigned in a DTR. Among them, some are inspired by classic approaches from RL and develop the optimal DTR in a backward stage-by-stage manner, such as Q-learning [19]–[23] and A-learning [24]–[26]. Others directly select the optimal DTR. These works either compute the values of all considered DTRs, or estimate the counterfactual expectation of the values and turn it into a classification problem. Only a few model-based approaches have been proposed to describe the mechanism of how an individual reacts to the treatments.

Q-learning and A-learning

Dynamic programming (DP) introduced by Bellman [27] is one of the classic approaches in RL and can be implemented in DTR research without assuming the Markov property. The key idea of DP is to parameterize the decision rule at each stage into a value function that specifies what is good in the long run. Then the problem can be broken down into a backward recursive optimization problem at each stage.

The primary approach that employs DP to construct the optimal DTR is Q-learning. It was originally proposed by Watkins (1989) [19], [20] as an off-policy temporal-difference control algorithm in RL. Murphy (2007) brought Q-learning to the research of DTRs and applied it to a SMART for patients with major depression [28]. The Q-function is defined to measure the quality associated with a treatment at a certain stage given the history up to that decision and then following the optimal regime thereafter, while a value function describes the expected utility score of a patient's history trajectory assuming that optimal decisions are made in the future. Note the value function here is not the same value of a DTR, except in the very beginning when there is no history. Often the Q-function is modeled via regression models [8]. The optimal DTR can be developed recursively by maximizing the Q-functions assuming that the optimal regime will be followed afterwards through a backward iterative fashion.

To formulate the problem, consider data collected from a T-stage SMART design. We still denote each trajectory as $\boldsymbol{X}, A_1, Y_1, ..., A_T, Y_T$. Let lower case letters be the observed

outcomes. The optimal DTR is developed from the last stage first. At the T^{th} stage, the Q-function and the value function can be written as:

$$Q_T(\mathbf{x}, a_1, y_1, ..., y_{T-1}, a_T) = E(r(\mathbf{x}, a_1, y_1, ..., y_{T-1}, a_T, Y_T) \mid \mathbf{x}, a_1, y_1, ..., y_{T-1}, a_T)$$
$$V_T(\mathbf{x}, a_1, y_1, ..., a_{T-1}, y_{T-1}) = \max_{a_T} Q_T(\mathbf{x}, a_1, y_1, ..., y_{T-1}, a_T)$$

$Q_T(\boldsymbol{x}, a_1, y_1, ..., a_T)$ evaluates the quality, i.e., the expected utility score of assigning a_T at the T^{th} stage. Then the optimal decision rule, given $(\boldsymbol{x}, a_1, y_1, ..., a_{T-1}, y_{T-1})$, is:

$$d_T^{opt}(\boldsymbol{x}, a_1, y_1, ..., a_{T-1}, y_{T-1}) = \arg\max_{a_T} Q(\boldsymbol{x}, a_1, y_1, ..., a_{T-1}, y_{T-1}, a_T).$$

For $t = T-1, ..., 1$, the quality associated with treatment a_t at the t^{th} stage can be derived from the value function at the $(t+1)^{th}$ stage as:

$$Q_t(\boldsymbol{x}, a_1, y_1, ..., a_{t-1}, y_{t-1}, a_t) = E(V_{t+1}(\boldsymbol{x}, a_1, ..., a_t, y_t) \mid \boldsymbol{x}, a_1, y_1, ..., a_{t-1}, y_{t-1}, a_t)$$

We have the optimal decision rule $d_t^{opt}(\boldsymbol{x}, a_1, y_1, ..., a_{t-1}, y_{t-1}) = \arg\max_{a_T} Q(\boldsymbol{x}, a_1, y_1, ..., a_{t-1}, y_{t-1}, a_t)$. Then the value function at the t^{th} stage is

$$V_t(\boldsymbol{x}, a_1, y_1, ..., a_{t-1}, y_{t-1}) = \max_{a_t} Q_t(\boldsymbol{x}, a_1, y_1, ..., a_{t-1}, y_{t-1}, a_t)$$

By selecting these optimal decision rules recursively, the optimal DTR is constructed.

Figures 1.3 and 1.4 show an example of Q-learning for a two-stage randomized trial. Suppose two treatments are available at each decision point, a_1 and a_1' at the first stage and a_2 and a_2' at the second stage. Let $\boldsymbol{x}$ be the baseline covariates and $Y_1 = 1$ and $Y_1 = 0$ be the possible responses after receiving the first treatment. The first step is to determine the optimal decision rules at Stage 2. For instance, given $(\boldsymbol{x}, a_1, y_1 = 1)$, $Q_2(\boldsymbol{x}, a_1, y_1 = 1, a_2)$ and $Q_2(\boldsymbol{x}, a_1, y_1 = 1, a_2')$ are compared, and a_2' is selected as the optimal treatment, highlighted in red. Similarly, other optimal treatments are selected, highlighted in red. Then, assuming that the optimal treatment will be assigned at the second stage, a_1 is selected as the optimal

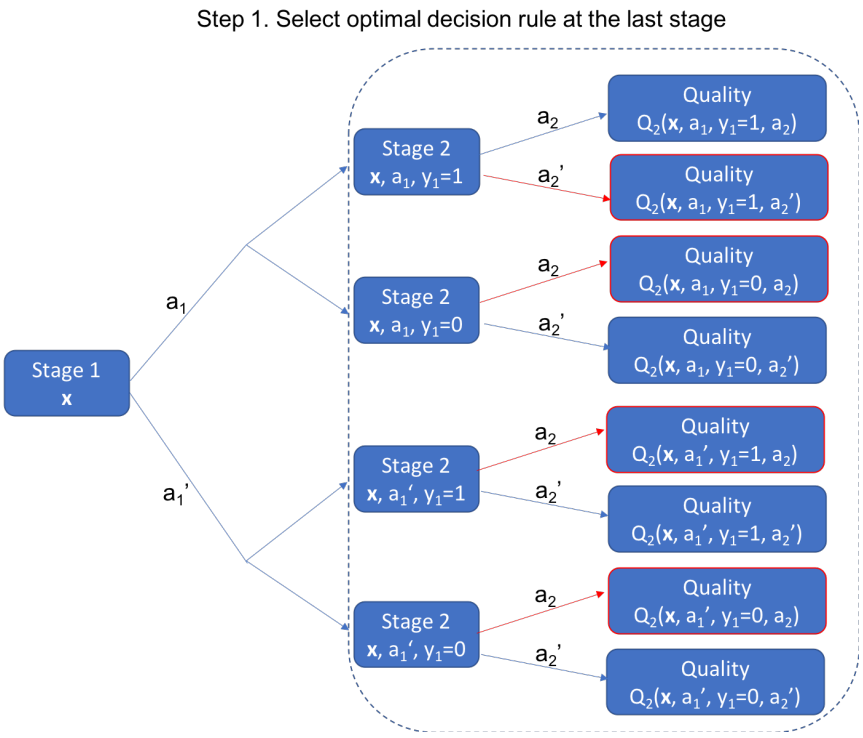

Figure 1.3. An example of the first step in Q-learning for a two-stage problem

treatment given $\boldsymbol{x}$ (Figure 1.4). As a result, the optimal DTR is $d_1(\boldsymbol{x}) = a_1$, and then $d_2(\boldsymbol{x}, a_1, y_1 = 1) = a_2'$ or $d_2(\boldsymbol{x}, a_1, y_1 = 0) = a_2$.

There are many different ways to describe the Q-function, ranging from standard linear regression [28] to non-parametric techniques like kernel regression[21], support vector regression and extremely randomized trees [22]. Nahum-Shani (2012) further discussed how to use Q-learning to analyze different types of more tailored SMART data [23]. In this dissertation, we will primarily use linear regression to model the Q-function when comparing our approach with Q-learning.

Another well-known backward recursive algorithm is advantage-learning or A-learning. Compared with Q-learning, A-learning focuses on the difference between the expected values of outcomes, instead of modeling the expected value of outcomes. Under the potential outcome framework, Robins has proposed several pioneering statistical methods in the domain of modeling the effects of sequential treatments over two decades ago, particularly, G-estimation of structural nested mean models (SNMM) [12], [13], [29], [30] to describe the

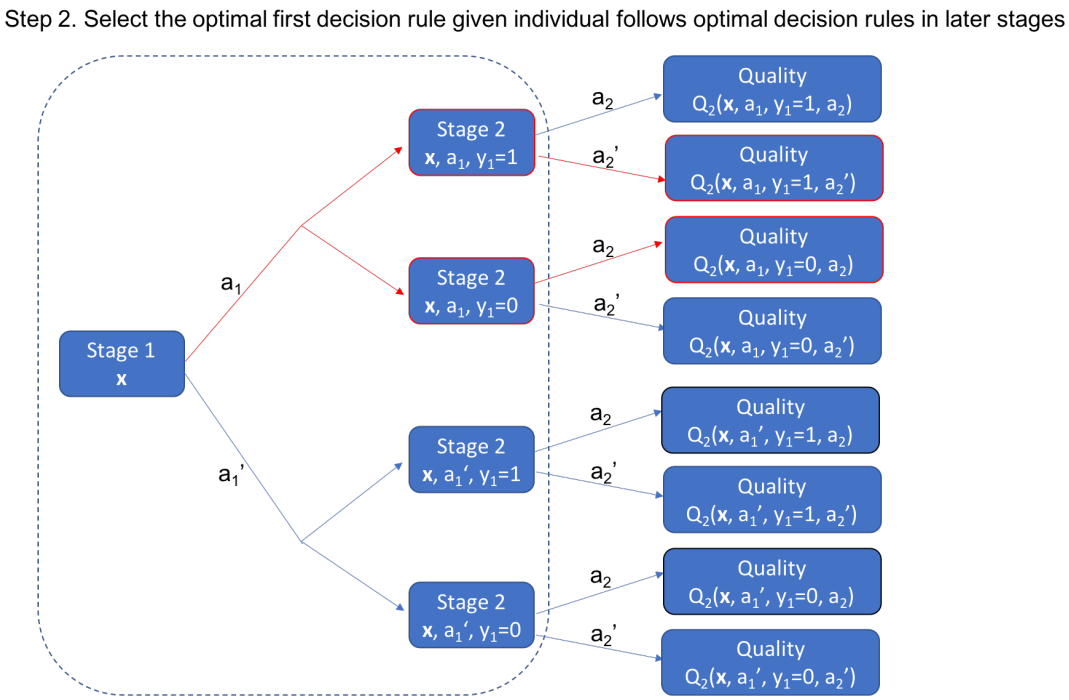

Figure 1.4. An example of the second step in Q-learning for a two-stage problem

causal effect, that is, the contrast between the expected values of outcomes. Based on these works, Murphy (2003) [24] considers a regret function to find the optimal DTR through a semi-parametric method. The regret function represents the loss of the utility incurred if the actual decision at a certain stage is not the optimal one, or in other words, the advantage in response if the optimal decision is given instead of the actual assigned one [25]. Similar to Q-learning, the optimal DTR can be developed by maximizing the potential advantage through a backward iterative fashion. Eventually, the sequence of optimal decision rules obtained makes up the optimal DTR. Later, Robins (2004) proposed using G-estimation to find the parameters of the regret function [26].

Direct Methods

Another common approach is to directly learn the optimal DTR without constructing optimal decision rules stage by stage. Given a set of DTRs, the expected values of these

DTRs can be estimated or compared. The optimal DTR is the one that has the maximum estimated value among the pre-specified set.

In direct methods, the values of DTRs or the contrast of values of DTRs are estimated. Inverse probability of treatment weighting (IPTW) is a commonly-used technique for this purpose. It was initially proposed by Robins (1994) to model the conditional mean of a response when data are missing at random [31] and the missing probabilities are either known or can be modeled [32]. Later on, Robins and his colleagues (2000) developed a new class of causal models, marginal structural model (MSM), and extended IPTW to estimate the parameters of the causal effect of static sequential treatment regimes in observational studies [33]–[35]. Murphy (2001) soon applied it to model the marginal mean response of a dynamic treatment regime [36]. Further extension works of MSM can be found in the literature [37]–[42].

It is desirable to find the estimated values of DTRs through IPTW estimation of MSM, when the set of pre-specified DTRs is small and the number of baseline covariates is low. Otherwise, this type of approach might not be suitable. With more baseline covariates under consideration, Qian and Murphy (2011) proposed a two-step procedure, which estimates the conditional mean for the response using l_1 penalized least squares with a rich linear model first, and then derives the estimated optimal individual treatment rules from the estimated conditional mean [43]. More importantly, this work pointed out the potential connection between the difference in mean responses and the margin in classification problems.

Consequently, research interest arose in estimating the optimal treatment via classification methods. Zhao (2012) cleverly demonstrated that the optimal treatment rule can be estimated within an outcome weighted classification framework, where weights are determined from the clinical outcomes [44]. This work is recognized as outcome weighted learning (OWL), or O-learning. After substituting the $0-1$ loss with the hinge loss [45], the support vector machine (SVM) is employed to determine the estimated optimal treatment rule.

Figure 1.5 illustrates the loss function and the intuition behind it. Suppose two treatments a and a' are considered. R is the clinical outcomes observed. The higher R is, the better the outcome is. Let $f(x)$ be the deterministic function for choosing a treatment, where x is baseline covariates. If $f(x) > 0$ then select a, otherwise select a'. The red line shows

how the loss changes as $f(x)$ changes. On the left where treatment a is assigned, if a is the true optimal treatment indeed, a larger $f(x)$ means selecting a as the optimal treatment, resulting in no loss if $f(x) \geq 1$. A smaller $f(x)$ means that it is highly likely to select a'. The more $f(x)$ is less than 1, the more loss incurs. The same logic applies to the right where a' is assigned. Therefore, minimizing the loss is equivalent to finding a deterministic function f that leads to the optimal treatment.

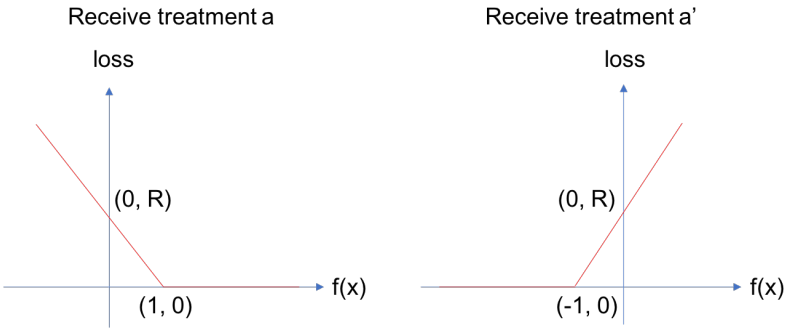

Figure 1.5. Loss function in O-learning

Zhang (2012) proposed a general framework that recasts the problem of estimating an optimal treatment regime into a classification problem wherein the optimal classifier corresponds to the optimal treatment regime, and used the Bayes classifier to estimate the optimal treatment regime by minimizing the expected weighted misclassification error [46].

Based on Zhao's (2012) work for one-stage individualized treatment selection, Zhao (2013) extended OWL to handle a multistage DTR and introduced two novel learning methods, backward outcome weighted learning (BOWL) and simultaneous outcome weighted learning (SOWL), to estimate the optimal DTR [47]. BOWL implements a similar backward fashion as Q-learning, where the optimal decision rule at the final stage is determined by OWL first, and then the optimal decision rules at previous stages are determined recursively with the assumption that the subjects would follow the estimated optimal decision rules thereafter. SOWL, on the other hand, learns the optimal DTRs at all stages simultaneously by substituting the hinge loss function in OWL with a smooth surrogate reward function.

Model-based Methods

Due to the concern over model misspecification [6], only a few model-based approaches have been developed to describe the response to the treatment at each stage. Among these approaches, almost all focus on trials with binary responses. Thall, Millikan, and Sung (2000) [48] suggested a multinomial model and logistic regression for an early example of a SMART design for androgen-independent prostate cancer. Thall, Sung, and Estey (2002) [49] presented a statistical framework, including a family of generalized logistic regression models and an approximate Bayesian method, to select therapeutic strategies based on data from a multi-course acute myelogenous leukemia (AML) trial.

In 2007, Thall [7] proposed a conditional logistic regression model to identify promising treatment regimes for androgen-independent prostate cancer. They modeled $p_{t,\mathrm{j}}$, the probability of a success in the j^{th} stage given treatment t, as a function of the treatment, previous response history and disease volume level (high or low). Their logistic model is

$$logit(p_{t,\mathrm{j}}) = m_t + a_t Y_{t,\mathrm{j-i}} + b_t Z_{t,\mathrm{j}-1} + cI(\text{low disease volume})$$

where $Y_{t,\mathrm{j-i}}$ is the previous stage's binary response and $Z_{t,\mathrm{j}-1}$ is a defined numerical value that quantifies the unfavorable influence of the previous treatment failure before the j^{th} stage. The optimal regime is selected as a combination of the best first-line and second-line therapies.

1.2.3 Subgroup Analysis

Subgroup analysis is of great importance in clinical trials. The goal of subgroup analysis is to evaluate the treatment effect heterogeneity for a specific endpoint in subgroups of patients defined by baseline covariates [50]. Consequently, a subpopulation that is likely to benefit from treatments can be identified according to certain characteristics. Although the interest of subgroup analysis is not in finding the optimal treatment regimes, one can compare treatment effects for patients based on their baseline covariates and identify beneficial subgroups, and therefore find the optimal individualized treatments.

There are two common types of approaches for subgroup analysis: pre-specified analysis and post-hoc analysis. In pre-specified analysis, subgroups are pre-defined before the trial, and the assessment of treatment effects for one or a few specific characteristics are listed as the study objectives. In post-hoc analysis, the process of identifying subgroups happens after the trial is done and the data have been collected. In recent clinical trials, including DTR analysis, subgroups are ususally not defined before the trial. Therefore, it is of more interest in finding the subgroup memberships (i.e., post-hoc approach). Note that at first, subgroup analysis resembles the estimation of optimal one-stage individualized treatment rules [43], [44]. However, the ultimate purpose of subgroup analysis is not treatment selection, but assessing heterogeneity in treatment response. Shahn and Madigan (2014) [51] discuss two types of heterogeneity: continuous heterogeneity and discrete heterogeneity. We will now introduce them and review some methods on how to address each of these types of heterogeneity.

Continuous Heterogeneity

Heterogeneity in treatment responses often refers to the variation of treatment effects across the baseline covariates. Ideally, when the underlying knowledge of the treatment mechanism is known and the baseline covariates are observed, the variation among treatment effects can be well approximated by a smooth function of the baseline covariates and treatment. Shahn and Madigan (2014) referred to this type of heterogeneity as continuous heterogeneity [51].

There is a large amount of literature on statistical methods to model continuous heterogeneity. These approaches include early methods like parametric regression models [52], and generalized linear regression to model the response as a function of treatment, individual baseline covariates, and interactions between treatment and baseline covariates [53]–[55]. In recent years, machine learning approaches have gained popularity in predicting the expected difference between potential outcomes under treatments and control, such as regression tree methods [56]–[58] and boosting methods [59].

Discrete Heterogeneity

Sometimes, even with careful analysis over baseline covariates and treatments, heterogeneity in treatment responses cannot be fully addressed. For example, acute myeloid leukemia (AML) shows considerable genetic heterogeneity but the source of this between-subject variation is currently unknown[60]. This type of heterogeneity, which cannot be approximated by a smoothed function of the observed covariates, is termed discrete heterogeneity. In this setting, the variation among treatment responses depends on some latent subgroup identity, which may or may not be induced by an unmeasurable factor, like a genomic marker. Mixture models are commonly used to model heterogeneity described by latent subgroups.

When subgroup identity is completely independent of all observed covariates, one can assume that the subgroup identity follows a discrete distribution with the sum of all subgroup probabilities equal to 1. The treatment effect of an individual can then be viewed as a weighted sum of the subgroup-specific treatment effects. A large number of works, like the mixed Poisson regression model with covariate dependent rates by Wang (1996) [61], have been proposed to suit this scenario.

When there exist measurable baseline covariates associated with latent subgroup identity, models can be used to adjust the subgroup probabilities. There are many works that adapt logistic regression to model subgroup identity. Wang (1998) [62] proposes mixed logistic regression models for count data, where both subgroup identity and response rate are modeled by logistic regression. Wong and Li (2001) [63] develops the logistic mixture autoregressive model with exogenous variables (LMARX) to handle time series data. The model consists of two Normal distribution models and allows the mixing proportions to change over time through a logistic model. Sobel and Muthén (2012) [64] use a logistic-normal mixture model to describe the complier average treatment effect with the assumption that one subgroup belongs to a zero-effect class (treatment does not affect the outcomes). Shen and He (2015) [65] propose a logistic-normal mixture model where the outcome Y is modeled by a linear function of treatment indicator Z, and the mixing proportions vary through a logistic model on some of the baseline covariates X. Let $\delta \in \{0, 1\}$ be the subgroup indicator. A confirma-

tory statistical test is also developed to test the existence of subgroups. The model can be specified as:

$$Y = Z^T(\beta_1 + \beta_2\delta) + \epsilon$$
$$P(\delta = 1 \mid X) = \frac{\exp(X^T\gamma)}{1 + \exp(X^T\gamma)}$$

where Z contains the intercept and the treatment indicator. One can see that the outcome Y depends on subgroup identity and the received treatment only. As a result, continuous heterogeneity is not considered in the model. Later Shen and Qu [66] extend this model for longitudinal data. Treatment effects are modeled as random effects in the model. Time is also incorporated into the linear function for the response Y.

Shahn and Madigan (2014) [51] provide a general Bayesian framework for modeling treatment effect heterogeneity in experiments with non-categorical outcomes. Compared with Shen and He's (2015) work, this approach incorporates latent subgroup mixture components to capture discrete heterogeneity, as well as interaction terms between baseline covariates and treatment in the regression to capture continuous heterogeneity.

1.3 Our Research

Although heterogeneity has been well-studied in subgroup analysis, there is limited work that evaluates heterogeneity in DTR research. To fill this gap, we investigate how heterogeneity, specifically discrete heterogeneity, impacts the determination of the optimal DTR.

Model-free methods have gained popularity due to their robustness to model misspecification. However, the major drawback is the lack of interpretation of the heterogeneity in treatment responses and how the heterogeneity leads to the differences in the final outcomes. Direct methods concentrate on the values of DTRs only. Even for methods like Q-learning and A-learning, in which the optimal DTRs are constructed stage by stage, the focus is on the expected future return, instead of the treatment effect at the current stage. In addition, some methods have rigid requirements on the trial design and limitation of allowing assessment among a wider range of candidate DTRs. For example, extensions to multiclass

classification in O-learning are needed when more than two treatments are available at a decision point [47]. It calls for additional care and adjustment to handle more tailor-made trials with different endpoints and complicated outcomes.

On the contrary, a model-based approach describes patients' response to treatment at each stage, which, in turn, provides the probabilities of treatment and response trajectories. This information can be used to infer the optimal DTR as well as provides the knowledge of heterogeneity in treatment responses, which can be helpful on future trial design and treatment regime development. Furthermore, it is easy for model-based approaches to accommodate more tailor-made trials with multiple DTRs in consideration. The major criticism of model-based approaches is that they may suffer from model misspecification. However, when the model is reasonable, these approaches can provide both interpretability and generality [18]. In addition, model-based approaches utilize the information from patients who do not receive estimated optimal treatments, especially when the same treatment can be assigned more than once across stages. As a result, model-based approaches can be more sample efficient. These advantages motivate us to adopt a model-based approach to handle multistage DTRs and explore their performance with consideration of heterogeneity.

A full-fledged model-based approach should account for population heterogeneity when modeling the trajectories. In ideal situations, this heterogeneity can be described using all covariates, treatments and interaction terms in the response component of the model. However, in practice, additional heterogeneity exists that is often attributed to a missing class identifier covariate. We want to consider this sort of heterogeneity too.

Thus, we assess two scenarios of the discrete heterogeneity. The first one is when no observable baseline covariates are associated with subgroup identity and the second is when there are baseline covariates associated with the latent class. To do this, we define the latent subgroup structure as follows: for a given treatment, we assume that the whole population can be partitioned into two subgroups–a subgroup of individuals who may respond favorably to treatment and a subgroup of individuals who likely will not. Extending this to K treatments of interest $\{\omega_1, ..., \omega_K\}$, we denote a patient's subgroup through a vector $\boldsymbol{Z} = (Z_{\omega_1}, ..., Z_{\omega_K})$. Here $Z_{\omega_k} = 1$ if the treatment ω_k works for the subject and $Z_{\omega_k} = 0$ otherwise. This means there are 2^K subgroups, each representing a unique collection of in-

dividuals who respond favorably to a different set of treatments. In practice, each subject's subgroup identity is unknown and our interest is to infer the population subgroup structure.

In Chapter 2, we consider the scenario when no baseline covariates are considered. Since a binary treatment response is observed at each stage, we model the responses using the Bernoulli distribution with a logit of a favorable response that depends on the individual's subgroup and the treatment received. To analyze data from a multistage trial, a random effect [67] is included to account for the repeated observations from the same individual. Since subgroup identities are unknown, we propose a mixture of mixed logit models. Thus, the model contains a treatment effect for each of the K treatments, a subject-specific random effect, and the subgroup proportions for the 2^K latent subgroups.

In Chapter 3, we extend our model to incorporate two types of baseline covariates: one type helps determine the subgroup memberships while the other describes heterogeneity in treatment response within subgroups. For the former, the covariates are incorporated into a multivariate Bernoulli model, while for the latter, the covariates and their interactions with treatments are included in the logit model for the response to capture continuous heterogeneity. In addition, two types of time effects are also considered in the logit model for the response.

The rest of the dissertation is organized as follows: In Chapter 2, we introduce our model-based framework to describe discrete heterogeneity in response when there are no baseline covariates. A mixture of mixed logit models is developed and combined with data from a SMART to estimate the subgroup proportions and the probabilities of a favorable response. With this information, an optimal dynamic treatment regime can be determined. An EM algorithm for parameter estimation, as well as consideration of the identifiability conditions, are provided. Simulation studies are performed and compared with Q-learning approach to demonstrate the effectiveness of the proposed method and to determine adequate sample sizes.

In Chapter 3, we extend our model to incorporate two kinds of baseline covariates. We employ a multivariate Bernoulli distribution model to incorporate covariates in the determination of subgroup identity and expand our mixed logistic regression model to incorporate covariates in the determination of treatment response. We further extend our model to incor-

porate time effects and provide discussions on the hypothesis testing of the existence of time effects. As in Chapter 2, an EM algorithm for parameter estimation as well as simulation studies are presented.

In both Chapters 2 and 3, we apply our approaches to data from the MD Anderson advanced prostate cancer trial (2000) with the goal of comparing 12 different treatment regimes [48]. The trial is an early example of using a SMART in cancer research and has been analyzed in several ways, including a likelihood-based method [7], IPTW by Wang (2012) [68], and Q-learning by Huang (2015) [69]. We conclude the dissertation in Chapter 4 with a summary of our contributions along with directions for future research.

2. A SEQUENTIAL MIXTURE OF MIXED LOGIT MODELS FOR DETERMINING OPTIMAL DYNAMIC TREATMENT REGIMES

In this chapter, we introduce our model-based framework and explain how to determine the optimal DTR under the special setting of no baseline covariates. We describe an EM algorithm for estimation and discuss the identifiability of this model and the optimal DTR under a specific SMART design. Simulation studies are conducted to evaluate our model's performance in estimating the values of DTRs and selecting the optimal one. We conclude the chapter with an application of our model to an MD Anderson advanced prostate cancer trial.

2.1 The Model Approach for Determining the Optimal DTR

In this section, we first lay out the probabilistic framework for a multistage DTR, followed by a description of our mixture of mixed logit models. We conclude the section with an explanation of how to compare DTRs and determine the optimal one using our model.

2.1.1 The Mixture Model for Binary Outcomes

Consider a T-stage sequence of treatment assignments (A) and binary responses (Y): $A_1, Y_1, ..., A_T, Y_T$. At each stage, we assume that there are the same K treatments available for assignment. We denote this treatment space as $\mathcal{A} = \{\omega_1, ..., \omega_K\}$ and the response space as $\mathcal{Y} = \{0, 1\}$. At any stage, the sequence of past treatments and responses is called its history. We use $\mathcal{H}_{t-1} = \{(\mathcal{A} \times \mathcal{Y}) \times ... \times (\mathcal{A} \times \mathcal{Y}) = (\mathcal{A} \times \mathcal{Y})^{t-1}\}$ to represent this space for Stage t. This implies $\mathcal{H}_T = \{(\mathcal{A} \times \mathcal{Y})^T\}$ is the trajectory space.

We label a T-stage DTR as $D = (d_1, ..., d_T)$, where d_t is the decision rule at the t^{th} stage. When no baseline covariates are available, the first decision rule d_1 assigns a treatment from $\mathcal{A}$ to all the patients. Each patient then responds to that treatment and decision rule d_2 determines whether the patient will stay on that treatment or switch to a different one. In other words, d_2 is a mapping from (A_1, Y_1) to $\mathcal{A}$. We denote this assignment as

$A_2 = d_2(A_1, Y_1)$. In general, the decision rule d_t is a mapping from the patient history $H_{t-1} = (A_1, Y_1, ..., A_{t-1}, Y_{t-1})$ to $\mathcal{A}$.

We use lower case letters to represent observable outcomes. Thus, at Stage t, we define $\Gamma_{D_{t-1}}$ to be the collection of all possible observable histories under DTR D. More specifically,

$$\Gamma_{D_{t-1}} = \{(a_1, y_1, ..., a_{t-1}, y_{t-1}) \in \mathcal{H}_{t-1} | P_D(a_1, y_1, ..., a_{t-1}, y_{t-1}) > 0\},$$

where P_D is the probability of observing the given history under DTR D. These sets of observable histories $t = 2, ..., T$ are most easily constructed successively, until reaching $\Gamma_D = \{(a_1, y_1, ..., a_T, y_T) \in \mathcal{H}_T | P_D(a_1, y_1, ..., a_T, y_T) > 0\}$, which is the collection of all possible observable trajectories for DTR D.

In order to compute the expected value of a DTR, the probability of each trajectory is needed. At the patient level, these probabilities depend on the latent subgroup $\boldsymbol{z}$ and the patient's random effect s. A trajectory probability conditional on $\boldsymbol{z}$ and s can be broken down into the product of sequential conditional probabilities:

$$P_D(a_1, y_1, ..., a_T, y_T | s, \boldsymbol{z}) = P(y_1 | a_1, s, \boldsymbol{z}) \times \prod_{t=2}^{T} P(y_t | a_1, ..., a_t = d_t(a_1, y_1, ..., y_{t-1}), s, \boldsymbol{z})$$

We impose the following assumptions to simplify this equation:

1. The subgroup identity $\boldsymbol{Z}$ and the subject-specific effect S remain constant across the stages of the trial.

2. A response Y_t is independent of the history H_{t-1} given A_t, $\boldsymbol{Z}$, and S.

Under these assumptions, our model simplifies to:

$$P_D(a_1, y_1, ..., a_T, y_T | s, \boldsymbol{z}) = P(y_1 | a_1, s, \boldsymbol{z}) \prod_{t=2}^{T} P(y_t | a_t = d_t(a_1, y_1, ..., y_{t-1}), s, \boldsymbol{z}) \tag{2.1}$$

The sequential nature of these trajectory probabilities arises from the fact that a_t is determined based on the patient's history and decision rule d_t, i.e., $a_t = d_t(h_{t-1})$. We introduce the indicator function $\mathbb{1}\{d_t(a_1, y_1, ..., a_{t-1}, y_{t-1}) = \omega_k\}$ to represent whether or not treatment ω_k is assigned at the t^{th} stage. There will be only one treatment such that

$\mathbb{1}\{d_t(a_1, y_1, ..., a_{t-1}, y_{t-1}) = \omega_k\} = 1$. Whether the assigned treatment can be effective depends on the latent subgroup. Let $p_t = P(y_t|a_t = d_t(a_1, y_1, ..., y_{t-1}), s, \boldsymbol{z})$. We use the following logit function to describe this relationship:

$$\log(\frac{p_t}{1-p_t}) = \mu + \sum_{k=1}^{K} \mathbb{1}\{d_t(a_1, y_1, ..., a_{t-1}, y_{t-1}) = \omega_k\}\tau_{\omega_k} z_{\omega_k} + s. \tag{2.2}$$

The indicator functions in Equation (2.2) imply that the $Y_1, ..., Y_T$ are correlated. For example, Y_{t+1} is correlated with Y_t since the assignment A_{t+1} depends on Y_t. Because of this correlation, we refer to the collection of these probabilities as a sequential mixed logit model.

To this point, our sequential mixed model is conditional on s and $\boldsymbol{z}$. However, in practice, s and $\boldsymbol{z}$ are unknown. To get the marginal probability of each trajectory, we first integrate out the subject-specific effect:

$$\begin{aligned} P_D(a_1, y_1, ..., a_T, y_T|\boldsymbol{z}) &= \int_s P_D(a_1, y_1, ..., a_T, y_T|s, \boldsymbol{z})\phi(s)ds \\ &= \int_s \prod_{t=1}^{T} P_D(y_t|a_t = d_t(a_1, y_1, ..., a_{t-1}, y_{t-1}), s, \boldsymbol{z})\phi(s)ds \end{aligned}$$

where $\phi(s)$ is the $N(0, \sigma^2)$ density function. We call these the subgroup-specific trajectory probabilities. We then average out the latent subgroup using the population proportions $\boldsymbol{\pi_z}$:

$$P_D(a_1, y_1, ..., a_T, y_T) = \sum_{\boldsymbol{z}} \boldsymbol{\pi_z} P_D(a_1, y_1, ..., a_T, y_T|\boldsymbol{z}),$$

It is these probabilities that represent the DTR D and serve as the probability mass function for computing its value.

Example 2.1 Consider a two-stage DTR involving only two treatments, A and B. This means there are four subgroups, which we proportionally assign as follows:

$Z = (Z_A, Z_B)$	$(1,1)$	$(1,0)$	$(0,1)$	$(0,0)$
$\boldsymbol{\pi_z}$	40%	40%	10%	10%

For the response probability component of our model, we assume that the subject effect $S \sim N(0, 0.1^2)$ with $\mu = -2.9222$. For $s = 0$, this corresponds to a natural response rate

of 5% when assigned an ineffective treatment. The treatment effects and corresponding response rates given different s are:

Parameter	Effect	Response Rate		
		$s=-0.3$	$s=0$	$s=0.3$
τ_A	3.3499	53.1%	60.0%	67.4%
τ_B	4.3307	75.1%	80.0%	84.7%

Now consider the DTR D where $d_1 = B$, $d_2(a_1 = B, y_1 = 1) = B$ and $d_2(a_1 = B, y_1 = 0) = A$. This DTR has four possible trajectories that we label: (B, 1, B, 1), (B, 1, B, 0), (B, 0, A, 1), and (B, 0, A, 0). The conditional logit function for the first stage is

$$\log(\frac{p_1}{1-p_1}) = \mu + \tau_B z_B + s. \tag{2.3}$$

and the conditional logit functions in the second stage are:

$$\begin{aligned} \log(\frac{p_2}{1-p_2}) &= \mu + \tau_B z_B + s, && \text{if } y_1 = 1 \\ \log(\frac{p_2}{1-p_2}) &= \mu + \tau_A z_A + s, && \text{if } y_1 = 0 \end{aligned} \tag{2.4}$$

Table 2.1 displays the subgroup-specific trajectory probabilities as well as the trajectory probabilities. The subject effect S is integrated out using the Gauss-Hermite quadrature.

Table 2.1. The probability of each subgroup-specific trajectory and overall trajectory for Example 2.1

Trajectory	Subgroup-specific trajectory probability				Trajectory probability
	$\boldsymbol{Z}=(1,1)$	$\boldsymbol{Z}=(1,0)$	$\boldsymbol{Z}=(0,1)$	$\boldsymbol{Z}=(0,0)$	
$B,1,B,1$	0.6395	0.0025	0.6395	0.0025	0.3210
$B,1,B,0$	0.1600	0.0477	0.1600	0.0477	0.1038
$B,0,A,1$	0.1199	0.5695	0.0100	0.0477	0.2815
$B,0,A,0$	0.0806	0.3803	0.1905	0.9021	0.2936

2.1.2 Evaluating DTRs

To evaluate and compare DTRs, a utility function $r(.)$ scores each trajectory. This function could focus exclusively on the final response Y_T or involve the entire sequence of responses. The value of the DTR D is defined as the average score of the possible trajectories:

$$V(D) = E_D(r(A_1, Y_1, ..., A_T, Y_T)) \tag{2.5}$$

For our binary outcome setting, there are a finite number of trajectories, so this expectation can be expressed numerically as:

$$\begin{aligned} V(D) &= \sum_{a_1, y_1, ..., a_T, y_T \in \Gamma_d} r(a_1, y_1, ..., a_T, y_T) \times P_D(a_1, y_1, ..., a_T, y_T) \\ &= \sum_{a_1, y_1, ..., a_T, y_T \in \Gamma_d} \{r(a_1, y_1, ..., a_T, y_T) \\ &\quad \times \sum_{\boldsymbol{z}} \{\boldsymbol{\pi}_{\boldsymbol{z}} \int_s \prod_{t=1}^{T} P_D(y_t | a_t = d_t(a_1, ..., y_{t-1}), \boldsymbol{z}, s) \phi(s) ds\}\} \end{aligned} \tag{2.6}$$

If the number of possible DTRs and associated trajectories is small enough, we can directly compute $V(D)$ for each DTR and select the optimal one,

$$D^{opt} = \arg\max_D V(D). \tag{2.7}$$

For Example 2.1, there are eight possible DTRs. Table 2.2 lists the value for each of them given the utility function: $r(a_1, 1, a_2, 1) = 1$, $r(a_1, 1, a_2, 0) = r(a_1, 0, a_2, 1) = 0.5$, and $r(a_1, 0, a_2, 0) = 0$. In this case DTR 7 is optimal with a value of 0.5137.

The number of possible DTR trajectories grows very quickly. For a T-stage DTR with binary responses, since the treatment assignment is deterministic through the DTR, there are 2^T trajectories in Γ_D. Given 2^K subgroups, this means we have to compute at most 2^{K+T} conditional probabilities $P_D(a_1, y_1, ..., a_T, y_T | \boldsymbol{z})$. When K and T are large, directly evaluating $V(D)$ can be computationally expensive. In these cases, one option is to approximate $V(D)$ by Monte Carlo methods. This approach would involve sampling S and $\boldsymbol{Z}$ from their estimated distributions and then generating trajectories based on the DTR. Each resulting

Table 2.2. The values of the eight DTRs for Example 2.1

DTR	d_1	d_2		Value
		$y_1 = 1$	$y_1 = 0$	
1	A	A	A	0.4899
2	A	A	B	0.4976
3	A	B	A	0.4496
4	A	B	B	0.4574
5	B	A	A	0.4574
6	B	A	B	0.3685
7	B	B	A	**0.5137**
8	B	B	B	0.4249

trajectory is scored and the average of the scores is the approximate DTR value. The number of simulated trajectories needed will depend on the desired precision of the DTR value.

To find the optimal DTR, one needs to search over all possible DTRs and choose the one with the maximum value. At the t^{th} stage, there are 2^{t-1} possible histories. For each history, we assume one of the treatments in $\mathcal{A}$ is assigned through d_t. As a result, we can have $K^{2^{t-1}}$ possible decision rules. This means we can have up to $\prod_{t=1}^{T} K^{2^{t-1}} = K^{2^T-1}$ possible DTRs. An exhaustive search using Monte Carlo methods to approximate each DTR's value can be very computationally expensive. The alternative in these cases is to use dynamic programming.

2.1.3 Determining the Optimal DTR through Dynamic Programming

Dynamic programming (DP)[8], [27] does not require estimating all DTR values. Considering that a DTR is a sequence of decision rules, with each decision rule allocating treatments based on treatment and response histories, DP searches for the optimal DTR in a backward fashion.

To illustrate this process stage by stage, we introduce the subset $D^t = (d_t, ..., d_T)$ to be the set of decision rules starting from Stage t. Given an observed history h_{t-1}, we denote $V(D^t|h_{t-1})$ to be the expected outcome from Stage t under D.

Let $V(D^T|h_{T-1}) = \sum_{y_T \in \{0,1\}} r(h_{T-1}, a_T, y_T) P(y_T|a_T = d_T(h_{T-1}))$. For $t = 1, ..., T-1$, the expected outcome can be developed recursively:

$$
\begin{aligned}
V(D^t|h_{t-1}) =& E_D\{r(h_{t-1}, a_t, y_t, ..., a_T, y_T)|h_{t-1}\} \\
=& \sum_{h_{t-1}, a_t, y_t, ..., a_T, y_T \in \Gamma_D} P(a_t, y_t, ..., a_T, y_T|h_{t-1}) \times r(h_{t-1}, a_t, y_t, ..., a_T, y_T) \\
=& \sum_{h_t, a_{t+1}, y_{t+1}, ..., a_T, y_T \in \Gamma_D} \sum_{y_t \in \{0,1\}} P(y_t, ..., a_T, y_T|h_{t-1}, a_t = d_t(h_{t-1})) \\
& \times r(h_{t-1}; a_t, y_t, ..., a_T, y_T) \\
=& \sum_{y_t \in \{0,1\}} \Big\{ P(y_t|h_{t-1}, a_t = d_t(h_{t-1})) \\
& \times \Big\{ \sum_{h_t, a_{t+1}, ..., y_T \in \Gamma_D} r(h_t; a_{t+1}, ..., y_T) P(a_{t+1}, ..., y_T|h_t) \Big\} \Big\} \\
=& \sum_{y_t \in \{0,1\}} P(y_t|h_{t-1}, a_t = d_t(h_{t-1})) V(D^{t+1}|h_{t-1}, a_t = d_t(h_{t-1}), y_t) \\
=& E_D\{V(D^{t+1}|h_{t-1}, a_t = d_t(h_{t-1}), Y_t)\}
\end{aligned}
$$

The optimal expected outcome from Stage t given history h_{t-1} satisfies the Bellman equation[27], and can be written as:

$$
\begin{aligned}
V^{opt}(D^t|h_{t-1}) &= \max_{D^t} V(D^t|h_{t-1}) \\
&= \max_{a_t \in \mathcal{A}} E_D\{V^{opt}(D^{t+1}|h_t, a_t, Y_t)|h_{t-1}, a_t = d_t(h_{t-1})\} \\
&= \max_{a_t \in \mathcal{A}} \sum_{y_t \in \{0,1\}} P(y_t|h_{t-1}, a_t = d_t(h_{t-1})) V^{opt}(D^{t+1}|h_{t-1}, a_t = d_t(h_{t-1}), y_t)
\end{aligned}
$$

The optimal DTR $D^{opt} = (d_1^*, ..., d_T^*)$ can be constructed in a backward fashion by maximizing $V(D^t|h_{t-1})$ at each stage[24]. For all possible h_{T-1}, the optimal decision rule $d_T^*(a_1, ..., y_{T-1})$ can be determined in the beginning by maximizing $V(D^T|h_{T-1})$ over $\mathcal{A}$. At the t^{th} stage ($t > 1$), the subsequent decision rules $d_{t+1}^*, ..., d_T^*$ are already determined. For any h_{t-1}, we can find d_t^* that maximizes the expected utility $V(D^t|h_{t-1})$. Eventually, d_1^* can be determined based on $d_2^*, ..., d_T^*$.

The detailed algorithm is:

Result: The optimal DTR $D^{opt} = (d_1^*, ..., d_T^*)$
$t = T$;
while $t > 0$ **do**
  let $D^t = (d_t, d_{t+1}^*, ..., d_T^*)$;
  For all possible $h_{t-1} = (a_1, y_1, ..., a_{t-1}, y_{t-1})$;
  find $d_t^*(a_1, y_1, ..., a_{t-1}, y_{t-1}) = argmax_{a_t \in \mathcal{A}} V(D^t | h_{t-1})$;
  $t = t - 1$;
end

Algorithm 1: DP solution to find optimal DTR

According to the algorithm, at Stage t, the subsequent decision rules $(d_{t+1}^*, ..., d_T^*)$ and $V^{opt}(D^{t+1}|h_{t-1}, a_t = d_t(h_{t-1}), y_t)$ are known. There are $2^{(t-1)} \times K^{t-1}$ possible histories in $\mathcal{H}_{t-1}$. In order to find d_t^* for each history, K possible treatment choices for d_t and 2 possible responses for y_t are considered. Together, we need to evaluate $2^t \times K^t$ paths at the t^{th} stage. As a result, to find an optimal DTR, we need to evaluate $\sum_{t=1}^{T} K^t 2^t$ treatment and response trajectories. Compared to the exhaustive search, DP can reduce the computational burden when T and K are large. However, DP cannot provide information for the values of DTRs except the optimal one. The sub-optimal DTRs can also be of interest. Therefore, we prefer an exhaustive search when feasible because of the additional information it provides.

We now use Example 2.1 to show how to apply DP to find the optimal DTR. Given (a_1, y_1), shown in the first column in Table 2.3, and the decision rule d_2, shown in the second column, we first compute the conditional probability of all y_2, and evaluate the value. Then, we choose the $a_2 = d_2^*(a_1, y_1)$ that results in $V^{opt}(d_2|a_1, y_1)$.

Given those d_2^*, we compute $V(d_1, d_2^*(a_1, y_1)) = \sum_{y_1 \in \{0,1\}} P(y_1|d_1 = a_1) V^{opt}(d_2|a_1, y_1)$ and select $d_1^* = B$, shown in Table 2.4. The optimal DTR we find through DP is the same as through exhaustive search in Table 2.2.

2.2 Parameter Estimation and DTR Evaluation

Now that we've demonstrated how to evaluate DTRs and determine the optimal one given known model parameter values, we shift to parameter estimation. Intuitively, one could consider collecting data from the DTR of interest, but this can result in identifiability

Table 2.3. Select d_2^* for Example 2.1

(a_1, y_1)	$a_2 = d_2(a_1, y_1)$	y_2	$P(y_2\|a_1, y_1, a_2)$	$V(d_2\|a_1, y_1)$
$(A, 1)$	***A***	1	0.5894	**0.7947**
		0	0.4106	
	B	1	0.4253	0.7126
		0	0.5747	
$(A, 0)$	A	1	0.3942	0.1971
		0	0.6058	
	B	1	0.4244	**0.2122**
		0	0.5756	
$(B, 1)$	A	1	0.4903	0.7452
		0	0.5097	
	B	1	0.7556	**0.8778**
		0	0.2444	
$(B, 0)$	***A***	1	0.4895	**0.2447**
		0	0.5105	
	B	1	0.1805	0.0903
		0	0.8195	

Table 2.4. Select d_1^* for Example 2.1

$d_1 = a_1$	y_1	$a_2 = d_2^*(a_1, y_1)$	$V^{opt}(d_2\|a_1, y_1)$	$V(d_1, d_2^*(a_1, y_1))$
A	1	A	0.7947	0.4976
	0	B	0.2122	
B	1	B	0.8778	**0.5137**
	0	A	0.2447	

issues, which we will discuss later. It is also not an efficient approach to investigate multiple DTRs.

We consider using data generated from a SMART design to search for optimal DTRs. In a SMART design, treatments are assigned to patients according to a certain pre-specified randomization scheme that may or may not be based on their treatment and response histories. In order to fit our model to these data, we modify our mixture of mixed logit model to incorporate this randomization. We then describe the use of the EM algorithm to obtain estimates and then conclude with a discussion on the conditions for identifiability.

2.2.1 Model Specification for Randomized Trials

Consider a T-stage randomized trial involving K treatments. For each subject, there is a sequence of treatment assignments (A) and outcomes (O) labeled $A_1, O_1, ..., A_T, O_T$. The conditional distribution of a subject's sequence given the treatment subgroup and subject-specific effect is:

$$\begin{aligned} P(a_1, o_1, ..., a_T, o_T|s, \boldsymbol{z}) =& P(a_1)P(o_1|a_1, s, \boldsymbol{z}) \\ & \times \prod_{t=2}^{T} P(a_t|a_1, o_1, ..., a_{t-1}, o_{t-1})P(o_t|a_1, o_1, ..., a_t, s, \boldsymbol{z}). \end{aligned} \tag{2.8}$$

In contrast to Equation (2.1), a_t is not determined through d_t but rather through a pre-specified randomization distribution $P(a_t|a_1, o_1, ..., a_{t-1}, o_{t-1})$. Imposing the same model assumptions of constant s and $\boldsymbol{z}$ and conditional independence of the outcomes, Equation (2.8) simplifies to:

$$P(a_1, o_1, ..., a_T, o_T|s, \boldsymbol{z}) = P(a_1)P(o_1|a_1, s, \boldsymbol{z}) \prod_{t=2}^{T} P(a_t|a_1, o_1, ..., a_{t-1}, o_{t-1})P(o_t|a_t, s, \boldsymbol{z}) \tag{2.9}$$

where $P(o_t|a_t, s, \boldsymbol{z})$ is Bernoulli with probability p_t, whose logit is

$$\log(\frac{p_t}{1-p_t}) = \mu + \sum_{k=1}^{K} \mathbb{1}\{A_t = \omega_k\}\tau_{\omega_k} z_{\omega_k} + s. \tag{2.10}$$

For some choices of the randomization distribution, the number of observed stages may vary by subject. For example, a subject's involvement in the trial may end after two consecutive failures or T stages, whichever comes first. For a dataset with N subjects, let T_{i} denote the number of stages for the i^{th} subject. Similarly, let $\boldsymbol{A}^{\mathrm{i}} = (a_1^{\mathrm{i}}, ..., a_{T_{\mathrm{i}}}^{\mathrm{i}})$ and $\boldsymbol{O}^{\mathrm{i}} = (o_1^{\mathrm{i}}, ..., _{T_{\mathrm{i}}}^{\mathrm{i}})$ represent the subject's assignment and outcome vectors, respectively.

Our goal is to estimate the latent subgroup proportions $\boldsymbol{\pi} = (\boldsymbol{\pi}_1, ..., \boldsymbol{\pi}_{2^K})$, treatment parameters μ and $\boldsymbol{\tau} = (\tau_{\omega_1}, ..., \tau_{\omega_K})$, and the subject-effect variance (σ^2). Let $\theta = (\mu, \boldsymbol{\tau}, \boldsymbol{\pi}, \sigma^2)$ represent this collection of parameters. The likelihood of this mixture model is given by:

$$L(\theta) = \prod_{\mathrm{i}=1}^{N} \int \phi(s_\mathrm{i}, \sigma^2) \sum_{z^\mathrm{i}} \{\boldsymbol{\pi}_{z^\mathrm{i}} \prod_{t=1}^{T_\mathrm{i}} \exp(o_t^\mathrm{i} \eta_t^\mathrm{i} - \log(1 + \exp(\eta_t^\mathrm{i})))\} ds_\mathrm{i} \tag{2.11}$$

where $\eta_t^\mathrm{i} = \mu + \sum_{k=1}^{K} \mathbb{1}\{A_t^\mathrm{i} = \omega_k\} \tau_{\omega_k} z_{\omega_k}^\mathrm{i} + s_\mathrm{i}$ and $\phi(s_\mathrm{i}, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{-\frac{s_\mathrm{i}^2}{2\sigma^2}\}$.

We use the following EM algorithm to obtain our parameter estimates. Given these estimates, we can estimate $\hat{P}_D(a_1, y_1, ..., a_T, y_T)$ and thus the value $\hat{V}(D)$ for any DTR D as

$$\hat{P}_D(a_1, y_1, ..., a_T, y_T) = \sum_{\boldsymbol{z}} \hat{\boldsymbol{\pi}}_{\boldsymbol{z}} \int_s \prod_{t=1}^{T} \hat{P}(y_t | a_t = d_t(a_1, y_1, .., a_{t-1}, y_{t-1}), s, \boldsymbol{z}) \phi(s, \hat{\sigma}^2) ds. \tag{2.12}$$

and

$$\hat{V}(D) = \sum_{a_1, y_1, ..., a_T, y_T \in \Gamma_D} r(y_1, ..., y_T) \hat{P}_D(a_1, y_1, ..., a_T, y_T) \tag{2.13}$$

We can use the methods of Sections 2.1.2 and 2.1.3 to evaluate and find the estimated optimal DTR.

2.2.2 EM Algorithm

There is no closed-form MLE formula for the likelihood in Equation (2.11) so we use the EM algorithm[70]. Subgroup identity $\boldsymbol{Z}$ is treated as a latent variable and the subject effect S is integrated out. Several methods could be used to approximate this integration over S, such as Gaussian quadrature and Monte Carlo. In this paper, we adopt Gaussian quadrature.

For convenience, we introduce the vector $\boldsymbol{C} = (C_1, ..., C_{2^K})$ to represent each subject's latent subgroup identity. This involves ordering the proportions $\boldsymbol{\pi}$ so that each element $\boldsymbol{\pi}_l$ agrees with the element C_l. To get the vector $\boldsymbol{C}^\mathrm{i}$ for subject i, we convert the binary K-dimensional vector $\boldsymbol{Z}^\mathrm{i}$ into a decimal number $L = \sum_{k=1}^{K} 2^{(k-1) Z_{\omega_k}}$ and let $C_L^\mathrm{i} = 1$ with all other elements 0.

If $\boldsymbol{C}^{\mathrm{i}}$ were observed, we have the following joint probability for the i^{th} subject's trajectory:

$$P(\boldsymbol{O}^{\mathrm{i}}, \boldsymbol{A}^{\mathrm{i}}, \boldsymbol{C}^{\mathrm{i}}) = P(\boldsymbol{C}^{\mathrm{i}}) \times P(\boldsymbol{O}^{\mathrm{i}}, \boldsymbol{A}^{\mathrm{i}} | \boldsymbol{C}^{\mathrm{i}}) \tag{2.14}$$

and complete-data log-likelihood:

$$\begin{aligned}
\log L(\theta; \boldsymbol{O}, \boldsymbol{A}, \boldsymbol{C}) &= \sum_{\mathrm{i}=1}^{N} \log\{P(\boldsymbol{C}^{\mathrm{i}}) \times P(\boldsymbol{O}^{\mathrm{i}}, \boldsymbol{A}^{\mathrm{i}} | \boldsymbol{C}^{\mathrm{i}})\} \\
&= \sum_{\mathrm{i}=1}^{N} \log\{P(\boldsymbol{C}^{\mathrm{i}})\} + \sum_{\mathrm{i}=1}^{N} \log\{P(\boldsymbol{O}^{\mathrm{i}}, \boldsymbol{A}^{\mathrm{i}} | \boldsymbol{C}^{\mathrm{i}})\} \\
&= \sum_{\mathrm{i}=1}^{N} \log\{\boldsymbol{C}^{\mathrm{i}} \boldsymbol{\pi})\} \\
&+ \sum_{\mathrm{i}=1}^{N} \log\{\int \phi(s_{\mathrm{i}}, \sigma^2) \prod_{t=1}^{T_{\mathrm{i}}} \exp(o_t^{\mathrm{i}} \eta_t^{\mathrm{i}} - \log(1 + \exp(\eta_t^{\mathrm{i}}))) ds_{\mathrm{i}}\}
\end{aligned} \tag{2.15}$$

where $\eta_t^{\mathrm{i}} = \mu + \sum_{k=1}^{K} \mathbb{1}\{A_t^{\mathrm{i}} = \omega_k\} \tau_{\omega_k} Z_{\omega_k}^{\mathrm{i}} + s_{\mathrm{i}}$.

Because the $\boldsymbol{C}^{\mathrm{i}}$s are unobserved, we use the EM algorithm to find the MLEs. Specifically, we iterate between taking the expectation of the complete-data log-likelihood using the conditional distribution of the missing $\boldsymbol{C}^{\mathrm{i}}$s (E-step), and maximizing the expected log-likelihood (M-step).

Suppose $\theta^{(t)}$ is the current set of parameters. In the E-step, the expected value of the complete log-likelihood function is:

$$\begin{aligned}
Q(\theta | \theta^{(t)}) &= E_{\boldsymbol{C}|\theta^{(t)}}\{\log L(\theta; O, A, C)\} \\
&= \sum_{\mathrm{i}=1}^{N} \sum_{l=1}^{2^K} P(C_l^{\mathrm{i}} = 1 | \boldsymbol{O}^{\mathrm{i}}, \boldsymbol{A}^{\mathrm{i}}) \log(\boldsymbol{\pi}_l) \\
&+ \sum_{\mathrm{i}=1}^{N} \sum_{l=1}^{2^K} P(C_l^{\mathrm{i}} = 1 | \boldsymbol{O}^{\mathrm{i}}, \boldsymbol{A}^{\mathrm{i}}) \log\Big\{ \int \phi(s_{\mathrm{i}}, \sigma^2) \{\prod_{t=1}^{T_{\mathrm{i}}} \exp(o_t^{\mathrm{i}} \eta_t^{\mathrm{i}} - \log(1 + \eta_t^{\mathrm{i}}))\} ds_{\mathrm{i}} \Big\}
\end{aligned} \tag{2.16}$$

where

$$P(C_l^{\mathrm{i}} = 1 | \boldsymbol{O}^{\mathrm{i}}, \boldsymbol{A}^{\mathrm{i}}) = \frac{\boldsymbol{\pi}_l^{(t)} \int \phi(s_{\mathrm{i}}, (\sigma^2)^{(t)})) \prod_{t=1}^{T_{\mathrm{i}}} \exp(o_t^{\mathrm{i}} (\eta_t^{\mathrm{i}})^{(t)} - \log(1 + (\eta_t^{\mathrm{i}})^{(t)})) ds}{\sum_{l=1}^{2^K} \boldsymbol{\pi}_l^{(t)} \int \phi(s_{\mathrm{i}}, (\sigma^2)^{(t)}) \prod_{t=1}^{T_{\mathrm{i}}} \exp(o_t^{\mathrm{i}} (\eta_t^{\mathrm{i}})^{(t)} - \log(1 + (\eta_t^{\mathrm{i}})^{(t)})) ds} \tag{2.17}$$

where $(\eta_t^{\mathrm{i}})^{(t)} = \mu^{(t)} + \sum_{k=1}^{K} \mathbb{1}\{A_t^{\mathrm{i}} = \omega_k\} \tau_{\omega_k}^{(t)} Z_{\omega_k}^{\mathrm{i}} + s_{\mathrm{i}}$.

This Q function consists of two parts. The first is:

$$Q_1(\theta|\theta^{(t)}) = \sum_{\mathrm{i}=1}^{N}\sum_{l=1}^{2^K} P(C_l^{\mathrm{i}} = 1|\boldsymbol{O}^{\mathrm{i}}, \boldsymbol{A}^{\mathrm{i}}) \log(\boldsymbol{\pi}_l) \tag{2.18}$$

and the second is

$$Q_2(\theta|\theta^{(t)}) = \sum_{\mathrm{i}=1}^{N}\sum_{l=1}^{2^K} \Big\{ P(C_l^{\mathrm{i}} = 1|\boldsymbol{O}^{\mathrm{i}}, \boldsymbol{A}^{\mathrm{i}}) \log\Big\{ \int \phi(s_{\mathrm{i}}, \sigma^2)\{\prod_{t=1}^{T_{\mathrm{i}}} \exp(o_t^{\mathrm{i}}\eta_t^{\mathrm{i}} - \log(1+\eta_t^{\mathrm{i}}))\}ds_{\mathrm{i}}\Big\}\Big\} \tag{2.19}$$

$Q_1(\theta|\theta^{(t)})$ involves $\boldsymbol{\pi}$ only, while $Q_2(\theta|\theta^{(t)})$ contains μ, $\boldsymbol{\tau}$, and σ^2. Therefore, in the M-step, we maximize them respectively.

To update $\boldsymbol{\pi}_l$, $l = 1, .., 2^K$ we have:

$$\begin{aligned}
\boldsymbol{\pi}_l^{(t+1)} &= arg\max_{\boldsymbol{\pi}_l} Q_1(\theta|\theta^{(t)}) \\
&= arg\max_{\boldsymbol{\pi}_l} \sum_{\mathrm{i}=1}^{N} \log(\sum_{l=1}^{2^K} \boldsymbol{\pi}_l P(C_l^{\mathrm{i}} = 1|\boldsymbol{O}^{\mathrm{i}}, \boldsymbol{A}^{\mathrm{i}})) \\
&= \frac{\sum_{\mathrm{i}=1}^{N} P(C_l^{\mathrm{i}} = 1|\boldsymbol{O}^{\mathrm{i}}, \boldsymbol{A}^{\mathrm{i}})}{N}
\end{aligned} \tag{2.20}$$

As for $Q_2(\theta|\theta^{(t)})$, since it involves the Gaussian distribution, we can use Gauss-Hermite quadrature[71] to approximate the integral. $Q_2(\theta|\theta^{(t)})$ can be approximated as

$$\begin{aligned}
&\sum_{\mathrm{i}=1}^{N}\sum_{l=1}^{2^K} P(C_l^{\mathrm{i}} = 1|\boldsymbol{O}^{\mathrm{i}}, \boldsymbol{A}^{\mathrm{i}}) \log\Big\{ \int \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{-\frac{s_{\mathrm{i}}^2}{2\sigma^2}\}\{\prod_{t=1}^{T_{\mathrm{i}}} \exp(o_t^{\mathrm{i}}\eta_t^{\mathrm{i}} - \log(1+\eta_t^{\mathrm{i}}))\}ds_{\mathrm{i}}\Big\} \\
&\approx \sum_{\mathrm{i}=1}^{N}\sum_{l=1}^{2^K} P(C_l^{\mathrm{i}} = 1|\boldsymbol{O}^{\mathrm{i}}, \boldsymbol{A}^{\mathrm{i}}) \log\Big\{ \frac{1}{\sqrt{\pi}} \sum_{\mathrm{j}=1}^{J} (\prod_{t=1}^{T_{\mathrm{i}}} \exp(o_t^{\mathrm{i}}\tilde{\eta}_{t\mathrm{j}}^{\mathrm{i}} - \log(1+\tilde{\eta}_{t\mathrm{j}}^{\mathrm{i}})))g_{\mathrm{j}}\Big\},
\end{aligned} \tag{2.21}$$

where $\tilde{\eta}_{tq}^{\mathrm{i}} = \mu + \sum_{k=1}^{K} \mathbb{1}\{A_t^{\mathrm{i}} = \omega_k\}\tau_{\omega_k} Z_{\omega_k} + \sqrt{2\sigma^2} d_{\mathrm{j}}$, and g_{j} and d_{j} be quadrature weights and points respectively, and Q is the number of quadrature points.

Since there is a constraint on $\sigma^2 > 0$, we adopt the numeric method L-BFGS-B[72] to maximize $Q_2(\theta|\theta^{(t)})$ over μ, $\boldsymbol{\tau}$, and σ^2. The gradient of $Q_2(\theta|\theta^{(t)})$ is calculated by Richardson's extrapolation[73].

Then the EM algorithm with Gaussian Quadrature is given as follows:

1. Choose starting values $\theta^{(0)}$. Set t=0
2. E-step: Calculate conditional probability $P(C_l^{\mathrm{i}} = 1|\boldsymbol{O}^{\mathrm{i}}, \boldsymbol{A}^{\mathrm{i}})$ according to $\theta^{(t)}$. Construct $Q(\theta|\theta^{(t)})$ as Equation (2.16) and approximate $Q_2(\theta|\theta^{(t)})$ by Equation (2.19).
3. M-step: update $\boldsymbol{\pi}$ by Equation (2.20). Update μ, $\boldsymbol{\tau}$, and σ^2 by maximizing Equation (2.21). Let t=t+1.
4. Repeat until convergence.

Here our stopping criteria is $|L(\theta^{(t+1)}) - L(\theta^{(t)})| < 10^{-8}$ and $|\theta^{(t+1)} - \theta^{(t)}| < 10^{-5}$. Note that the EM algorithm doesn't guarantee the global maximum likelihood. Several sets of random starting values are chosen to find the global MLEs.

2.2.3 Model Identifiability

Our primary goal is to determine the optimal DTR based on the parameters estimated from randomized trial data. However, mixture models sometimes suffer from non-identifiability. For this discussion, we assume that there are M possible trajectories $\{v_1, ..., v_M\}$, where $v_m = (a_1, o_1, ..., a_{T_m}, o_{T_m})$. Under our model, the parameter set θ=(μ, $\boldsymbol{\tau}$, $\boldsymbol{\pi}$, σ^2) is said to be identifiable, if for any two sets θ and θ^*, $P(v_m|\theta) = P(v_m|\theta^*)$, $m = 1, ..., M$, implies $\theta = \theta^*$[74].

Lack of identifiability can arise for mixture models in several ways[75]. Because we assume that there are no empty subgroups (all $\boldsymbol{\pi} > 0$) or identical subgroups in the model. The most likely way is due to generic problems[75]. There are a number of works on identifiability problems in mixture models[76], [77].

We will extend these works so they are suitable for our mixture of trajectories model, and provide sufficient conditions for local identifiability. These conditions can be used to check identifiability and serve as a guide for appropriate randomized trials. Because our primary goal is to determine the optimal DTR based on the parameters estimated from randomized trial data, we will also discuss the impact of non-identifiability on determining the optimal DTR.

Conditions for Identifiability

For any trajectory v_m, let $g_m(\theta) = P(v_m|\theta)$ be a mapping from parameter space Θ to $[0, 1]$. Assume that there are M trajectories of a DTR listed as M mappings $G = (g_1, ..., g_M)$. The number of parameters in θ is $2^K + K + 2$ and there is only one constraint: $\sum_{l=1}^{2^K} \boldsymbol{\pi}_l = 1$. Thus, G is a mapping from $\mathbb{R}^{2^K+K+1}$ to $[0, 1]^M$. Given θ, if there exists a neighborhood B_θ of θ, such that for $\forall \theta' \in B_\theta$, $g_m(\theta) \neq g_m(\theta')$, then we say θ is locally identifiable. According to Goodman[77], by the inverse function theorem, the parameters in the model will be locally identifiable if the rank of Jacobian matrix of G is no less than the number of parameters. Therefore, in our case, θ is locally identifiable, if

$$rank(J_G(\theta)) \geq 2^K + K + 1$$

where $J_G(\theta)$ is the Jacobian matrix of the function G at the point θ.

Non-identifiability and Partial Identifiability

Completely non-identifiability happens when none of the parameters is identifiable. For example, consider a two-stage, two-treatment trial that contains all possible DTRs. Patients are randomly assigned to either A or B in the beginning. At the second stage, they are again randomly assigned to either A or B, regardless of their responses to the first treatment. With consideration of the constrain $\sum_{l=1}^{4} \boldsymbol{\pi}_l = 1$, there are seven parameters in the model: $\theta = (\mu, \tau_A, \tau_B, \boldsymbol{\pi}, \sigma^2)$. We can observe 16 possible treatment and response trajectories from this trial. However, it only has $rank(J_G(\theta)) = 5 < 1 + 2 + 3 + 1 = 7$. Thus, the generic identifiable problem exists and no parameter is identifiable.

It sometimes happens that part of the parameters are identifiable while the others are not. We call this scenario partially identifiable. For example, in Sections 2.3.2, the prostate cancer trial involves four treatments but each patient receives at most two treatments. It is impossible to have all $2^4 = 16$ subgroup proportions identifiable. However, the subgroup structure for any two treatments is identifiable (details are discussed in Section 2.3.2).

The key to our goal of finding the optimal DTR, however, is the fact that for both completely non-identifiable and partially identifiable cases, the trajectory probabilities for the chosen randomized trial can still be estimated uniquely. In the two-stage, two-treatment trial example, although the parameter estimates obtained through the EM algorithm can vary, the estimated trajectory probabilities stay unchanged. In fact, since the trajectories are observed, for any given trajectory, its estimated probability should always be its population proportions, i.e., the ratio of the number of the trajectory observed and the total number of subjects. This means that we can still use our mixture model results to determine the optimal DTR.

In conclusion, for each clinical trial, the conditions of generic identifiability need to be checked case by case. Lack of identifiability impacts the major advantage of our method relative to model-free approaches, specifically, to provide interpretation of the treatment and subgroup structure. Therefore, in order to obtain this underlying knowledge, it is ideal to have a very large number of patients so one obtains a diverse set of trajectories. This implies there is a trade-off between the complexity of the trial and the information contained in the trial. With the goal of determining an optimal DTR, we can come up with a simpler experiment that allows us to obtain adequate parameter estimates to construct DTRs.

2.3 Simulation Studies

In this section, we assess the performance of our model-based approach through some simulation studies. Under several different scenarios, we summarize the accuracy of our model estimates, if identifiable, as well as the expected value of each DTR. More importantly, we assess how well our approach correctly identifies the optimal DTR and compare this accuracy with that of Q-learning.

The simulation studies are conducted under two randomized trial scenarios. For each scenario, we consider sample sizes of 200, 300, 600, 1200, and 2000, and replicate each scenario/sample size combination 200 times. Both our approach and Q-learning are applied to each replicate to identify the optimal DTR.

2.3.1 Two-stage Two-treatment Scenario

Consider the simple two-stage two-treatment trial of Section 2.1.3, where patients are randomly assigned to either A or B at both stages, regardless of their responses. Even though the model parameters are not identifiable (justification in Section 2.2.3), the trajectory probabilities and thus the value of a DTR are.

Table 2.5 summarizes the value of each DTR. Similar to the probabilities, there is very little bias and the precision improves with sample size. Figure 2.1 shows boxplots of the estimated probability of the trajectory (B, 0, A, 1) using our approach for the different sample sizes. We can see that there is very little bias and as the sample size increases, precision increases. Other trajectory probabilities show a similar behavior.

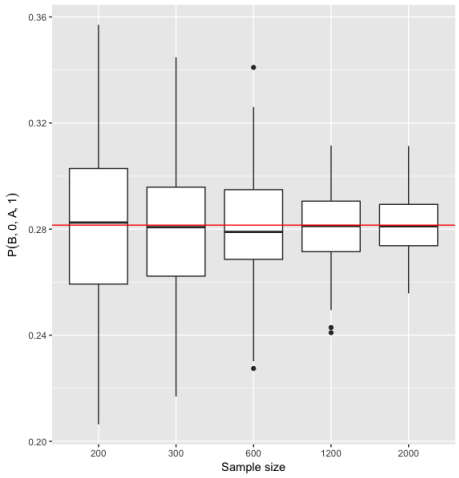

Figure 2.1. Boxplot of estimated probabilities of the trajectory (B, 0, A, 1). The red line represents the true value when s=0.

Table 2.6 summarizes the probabilities of finding the optimal DTR from both our approach and Q-learning. Although nonidentifiability limits our interpretation of each model parameter, it still has advantages in terms of finding the optimal DTR. This is because Q-learning determines the optimal decision rule based only on the information at each stage, while our approach combines information across stages and thus better estimates the probability of each trajectory. Thus, given that our model is correct, even though the model

Table 2.5. Means and standard deviations of estimated values of DTRs

DTR	True value	200	300	600	1200	2000
D_1	0.4899	0.4905 (0.0332)	0.4861 (0.0246)	0.4910 (0.0192)	0.4894 (0.0133)	0.4890 (0.0109)
D_2	0.4976	0.4986 (0.0340)	0.4964 (0.0264)	0.4992 (0.0197)	0.4975 (0.0144)	0.4970 (0.0110)
D_3	0.4496	0.4481 (0.0291)	0.4457 (0.0238)	0.4498 (0.0180)	0.4487 (0.0123)	0.4485 (0.0099)
D_4	0.4574	0.4562 (0.0219)	0.4560 (0.0165)	0.4581 (0.0118)	0.4568 (0.0084)	0.4566 (0.0063)
D_5	0.4574	0.4562 (0.0219)	0.4560 (0.0165)	0.4581 (0.0118)	0.4568 (0.0084)	0.4566 (0.0063)
D_6	0.3685	0.3674 (0.0276)	0.3680 (0.0216)	0.3692 (0.0163)	0.3688 (0.0117)	0.3683 (0.0086)
D_7	0.5137	0.5107 (0.0285)	0.5139 (0.0232)	0.5141 (0.0168)	0.5123 (0.0114)	0.5124 (0.0087)
D_8	0.4249	0.4219 (0.0272)	0.4259 (0.0220)	0.4252 (0.0163)	0.4243 (0.0103)	0.4241 (0.0078)

parameters are not identifiable, our approach has more power to infer the optimal DTR, especially for the larger sample sizes.

Table 2.6. Probabilities of finding the optimal DTR

sample size	Mixture Model	Q-learning
200	57.0%	50.5%
300	67.5%	55.0%
600	77.5%	62.0%
1200	86.5%	69.5%
2000	91.5%	71.0%

2.3.2 MD Anderson Prostate Trial Design

In this subsection, we simulate data under two parameter settings using the design from MD Anderson's advanced prostate cancer trial[7], [48]. Under their design, we have partial identifiability so we evaluate the identifiable parameter estimates as well as the expected values of the DTRs.

The design is a special case of a SMART. For practical considerations, patients receive multiple courses of chemotherapy until a certain criterion for termination is met. As a result, the number of stages a patient undergoes can vary. In the protocol, patients were initially randomized to one of four treatments. A second treatment was randomly assigned to patients at a subsequent stage only if the subject did not respond to the previous treatment. Figure 2.2 summarizes the possible treatment and outcome trajectories of this protocol. The bolded boxes represent the last outcome of each of the seven trajectories. The specific details of the protocol are as follows:

1. Randomly assign a treatment to each participant at the beginning of Stage 1.
2. At the end of a stage, assess whether each subject did (S) or did not (F) respond to treatment.
   (a) If a patient responded, assign the same treatment in the next stage.

(b) If the patient did not respond, randomly assign a different treatment in the next stage

3. Stop a patient once there are two consecutive favorable or a total of two unfavorable responses.

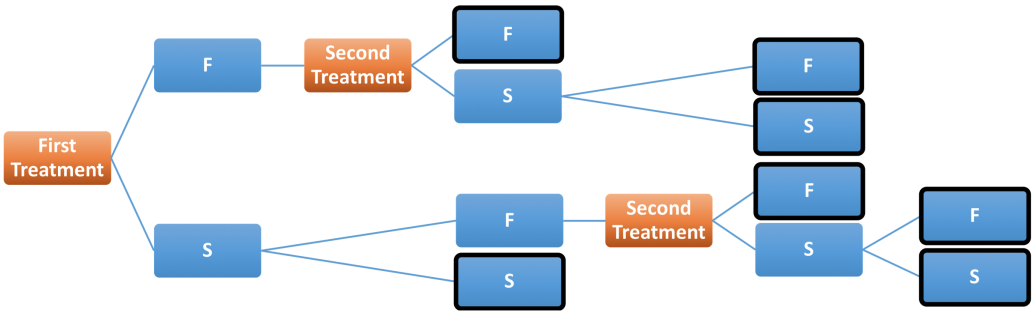

Figure 2.2. Protocol of MD Anderson's advanced prostate cancer trial

In the simulation studies, four treatments are available as the first treatment. Once a failure is observed, patients will be randomly assigned to one of the three remaining treatments. The seven possible sequences of responses are: FF, FSF, FSS, SFF, SFSF, SFSS, and SS. Given that there are 12 ordered pairs of treatments, this means there are 76 possible treatment and response trajectories. To evaluate and compare different DTRs, binary utility functions were applied to score the desirability of each trajectory[68]. The score is 1 if there are two consecutive favorable responses and 0 otherwise. Therefore, among the seven sequences of responses, SS, FSS, and SFSS are scored 1, the others are scored a 0.

Evaluation of Parameter Estimates and Expected Value of DTRs

For these simulations, patients are randomly assigned to one of the treatments (labeled 1, 2, 3, and 4) with equal probability. We also assume the natural response rate is zero. Therefore, the probability of a successful response $p_t = P(O_t = 1|a_t, s, \boldsymbol{z})$ can be written as

$$\log(\frac{p_t}{1-p_t}) = \tau_{a_t} + s, \quad \text{if } z_{a_t} = 1$$
$$p_t = 0, \quad \text{if } z_{a_t} = 0.$$

Here τ_{a_t} is response rate of the treatment a_t. Let 1, 2, 3, and 4 denote the four treatments. There are $2^4 = 16$ subgroups proportions, and 20 model parameters when we include $\boldsymbol{\tau}$ and σ^2. Since the patients receive at most two treatments, there is no trajectory that contains all these parameters, resulting in partial identifiability. Specifically, we are not able to identify all 16 subgroup proportions, but rather only the two treatment subgroups. This still allows us to determine the optimal DTR among those that involve only two treatments. We could reduce $2^4 = 16$ subgroup proportion parameters into 10 parameters, which are $\boldsymbol{\pi}_1$, $\boldsymbol{\pi}_2$, $\boldsymbol{\pi}_3$, $\boldsymbol{\pi}_4$, $\boldsymbol{\pi}_{12}$, $\boldsymbol{\pi}_{13}$, $\boldsymbol{\pi}_{14}$, $\boldsymbol{\pi}_{23}$, $\boldsymbol{\pi}_{24}$, and $\boldsymbol{\pi}_{34}$. Here $\boldsymbol{\pi}_1 = \sum_{\text{i,j},k\in\{0,1\}}$, and $\boldsymbol{\pi}_{\boldsymbol{z}=(1,\text{i,j},k)}$, $\boldsymbol{\pi}_{12} = \sum_{\text{i,j}\in\{0,1\}} \boldsymbol{\pi}_{\boldsymbol{z}=(1,1,\text{i,j})}$. Two generating parameter settings are in shown Tables 2.7 and 2.8.

Table 2.7. Parameter settings of treatment effects

	Setting 1		Setting 2	
	Treatment effect	Response rate	Treatment effect	Response rate
τ_1	1.10	75.0%	1.10	75.0%
τ_2	0.92	71.4%	0.69	66.7%
τ_3	0.69	66.7%	-0.69	33.3%
τ_4	0.41	60.0%	1.39	80.0%

Table 2.8. Parameter settings of subgroup proportions

	$\boldsymbol{\pi}_1$	$\boldsymbol{\pi}_2$	$\boldsymbol{\pi}_3$	$\boldsymbol{\pi}_4$	$\boldsymbol{\pi}_{12}$	$\boldsymbol{\pi}_{13}$	$\boldsymbol{\pi}_{14}$	$\boldsymbol{\pi}_{23}$	$\boldsymbol{\pi}_{24}$	$\boldsymbol{\pi}_{34}$
Setting 1	0.55	0.47	0.51	0.57	0.25	0.28	0.33	0.18	0.29	0.28
Setting 2	0.31	0.56	0.44	0.58	0.09	0.13	0.09	0.18	0.41	0.27

In MD Anderson's study, the second decision rule d_2 is the same for the previous response sequence F and SF, given the first treatment, i.e., $d_2(a_1 = \omega_k, F) = d_2(a_1 = \omega_k, S, a_2 = \omega_k, F)$. Therefore, 12 DTRs were compared. If the values of the two DTRs are the same, we prefer the DTR with a lower expected number of stages. Based on both DTR values and the expected number of stages given DTR, in Setting 1, we assign Treatment 1 first and then switch to Treatment 2 if there is a failure as the optimal DTR. In Setting 2, the optimal DTR is to assign Treatment 4 first and then switch to Treatment 1 if there is a failure.

Note that if we remove the restriction on d_2, there are 36 possible DTRs of interest. In this case, for Setting 2, the optimal DTR is to assign Treatment 4 first. If it fails at the first stage, switch to Treatment 1. If it fails after an initial success, then switch to Treatment 2.

Parameter estimates are evaluated from each replicate. The boxplots of all parameter estimates are included in Appendix A. As in earlier simulation studies, the precision increases with sample size. However, in this case, there appears to be some bias with some parameters (e.g., τ_4 in Setting 1). We conducted additional simulations with larger sample sizes (5000, 10,000, and 20,000) and found these estimates getting closer to the true values albeit slower than the other parameters.

Given the parameter estimates, we can compute the expected value of a DTR and select the optimal one. Table 2.9 shows the estimated values of the optimal DTR. As the sample size increases, the variabilities of the estimated value of the optimal DTR decrease. We also observe that the modified DTR has higher values compared with the optimal DTR chosen from the original protocol.

Table 2.9. Means and standard deviations of estimated values of optimal DTR for binary scores.

Sample size	Setting 1	Setting 2	Setting 2 without restriction
n = 200	0.4812(0.0544)	0.5100(0.0539)	0.5276(0.0551)
n = 300	0.4789(0.0418)	0.5110(0.0436)	0.5302(0.0445)
n = 600	0.4787(0.0282)	0.5096(0.0290)	0.5283(0.0311)
n = 1200	0.4779(0.0208)	0.5143(0.0237)	0.5348(0.0241)
n = 2000	0.4752(0.0162)	0.5140(0.0186)	0.5334(0.0189)
True value	0.4770	0.5128	0.5337

Comparison between our model and Q-learning

We compare our proposed approach with Q-learning. To implement Q-Learning, we separate the data into two parts. One part includes patients who receive two treatments, the other part is patients who receive one treatment. For patients who receive two treatments, given their histories, the optimal decision rule d_2^* is determined first using linear regression to maximize the trajectory utility R. We then create pseudo-outcome $\hat{R}$ by fitting the linear

regression, assuming the patients followed d_2^*. For patients who receive one treatment, i.e., (S, S), pseudo-outcome $\hat{R}$ is set to be 1. Eventually, we fit $\hat{R} \sim A_1$ and determine the optimal d_1^*. The optimal DTR is $d^* = (d_1^*, d_2^*)$.

Table 2.10. Probabilities of finding true optimal DTRs of Setting 1

sample size	mixture model	Q-learning
n = 200	43.5%	29.5%
n = 300	52.5%	33.0%
n = 600	57.0%	36.0%
n = 1200	65.5%	42.5%
n = 2000	73.5%	59.0%

Table 2.11. Probabilities of finding true optimal DTRs of Setting 2

sample size	mixture model	Q-learning
n = 200	50.5%	42.0%
n = 300	53.0%	56.5%
n = 600	56.5%	52.5%
n = 1200	70.0%	63.5%
n = 2000	73.5%	67.0%

Table 2.12. Probabilities of finding true optimal DTRs of Setting 2 without restriction

sample size	mixture model	Q-learning
n = 200	58.5%	14.0%
n = 300	68.5%	13.5%
n = 600	81.0%	25.0%
n = 1200	92.5%	37.5%
n = 2000	93.5%	43.5%

The optimal DTRs are selected for each replicate. We compare our approach with Q-Learning in terms of selecting the optimal DTR. Tables 2.10 - 2.12 shows the probabilities of finding true optimal DTRs for both mixture model and Q-learning.

All accuracies increase as sample size increases and our approach has advantages to find the true optimal DTR. Similar to the case in Section 2.3.1, it is because our approach utilizes

information across stages to determine the stage-wise optimal decision rule backwards, while Q-learning only use information from the current stage.

We also compare the value of optimal DTRs between our approach and Q-learning. Since we cannot calculate the probabilities of trajectories from Q-learning, the value of the optimal DTR are computed through Monte Carlo approach for both our approach and Q-learning. We apply the estimated optimal DTR to a dataset of 10,000 and obtain the mean of outcomes for 10,000 patients followed the estimated optimal DTR. Figures 2.3 and 2.4 show the smoothed histograms of values for estimated DTR of Setting 1 and Setting 2 without restriction.

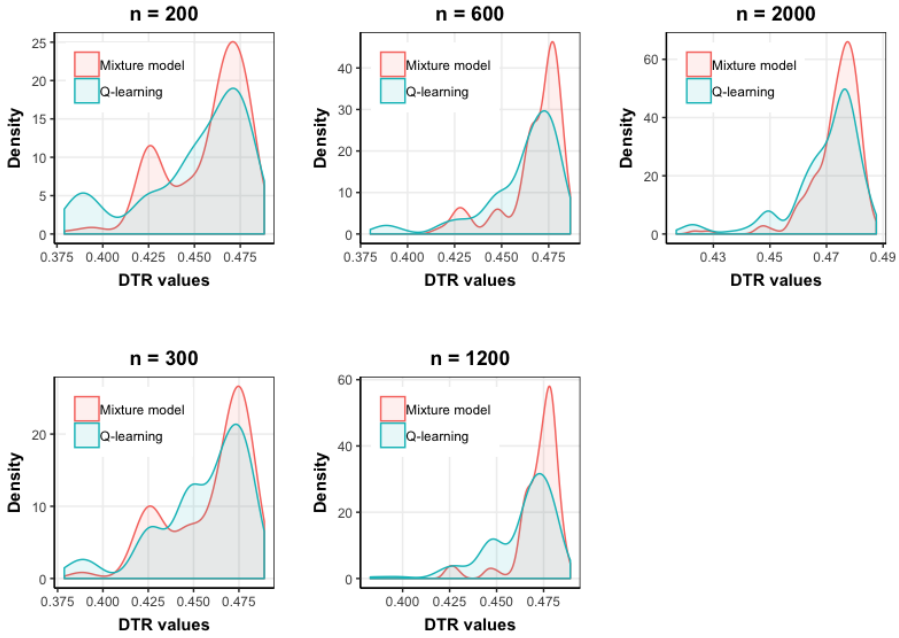

Figure 2.3. Smoothed histograms of values for estimated optimal DTR of Setting 1

In the smoothed histograms, we can see there are several local peaks. For example, in Figure 2.4, sample size = 200, we observe there are peaks around $V(d) = 0.5131$ and $V(d) = 0.5342$. In fact, the value of optimal DTR is 0.5342 while 0.5131 is the value of the second optimal DTR. The peak at 0.5131 is a result of picking the wrong optimal DTR. As the sample size increases, for both approaches, the density around the value of true optimal

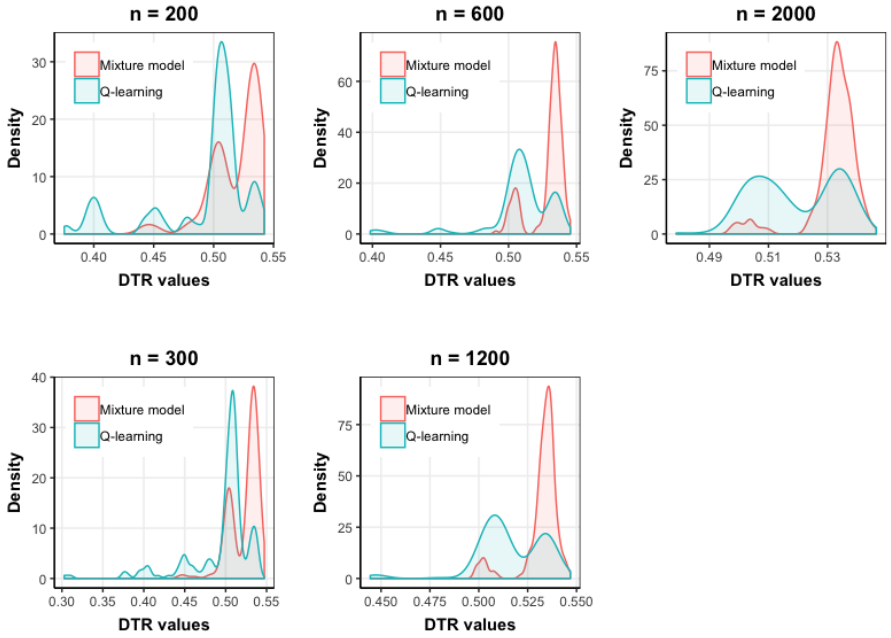

Figure 2.4. Smoothed histograms of values for estimated optimal modified DTR of setting 2 without restriction

DTR increases. Furthermore, our mixture model has a higher density around the value of true optimal DTR.

Table 2.13. Percentage of the proposed mixture model selecting DTRs with higher or equal values in Setting 1

N	Higher than Q-Learning	Equal to Q-Learning	Total
200	32.5%	48.5%	81.0%
300	19.5%	60.0%	79.5%
600	22.0%	60.0%	82.0%
1200	19.0%	67.0%	86.0%
2000	15.0%	75.0%	90.0%

In each replication, we further compare the values of the estimated optimal DTRs selected by the mixture model and Q-Learning. Tables 2.13 and 2.14 present the percentage of the proposed mixture model selecting DTRs with higher or equal values. In both settings, when the sample size increases from 200 to 2000, Q-Learning gradually has more chance to find DTRs with the same values as the mixture model (an increase from 48.5% to 75% in Setting

Table 2.14. Percentage of the proposed mixture model selecting DTRs with higher or equal values in Setting 2 without restriction

N	Higher than Q-Learning	Equal to Q-Learning	Total
200	57.5%	24.0%	81.5%
300	68.0%	21.0%	89.0%
600	60.0%	33.5%	93.5%
1200	57.0%	40.5%	97.5%
2000	51.0%	47.0%	98.0%

1 and from 24% to 47% in Setting 2 without restriction), while the proposed model still outperforms Q-Learning in term of selecting DTRs with higher values.

2.4 Application

We apply our proposed model to analyze real data from MD Anderson advanced prostate cancer trial with the goal of comparing 12 different treatment regimes. The four treatments/regimens were (CVD, KA/VE, TEC, and TEE), and favorable response means a 40% decrease in PSA (prostate-specific antigen) from the baseline. Since the self-cure case of advanced prostate cancer is very rare, we assume the natural response rate is 0.

Although 150 patients participated and received treatments in this trial, only 107 of them followed the protocol. We analyze just these 107 patients. Let 1, 2, 3, and 4 stand for CVD, KA/VE, TEC, and TEE. Table 2.15 shows the estimated subgroup-specific treatment effect and Table 2.16 shows the estimated subgroup proportions.

Table 2.15. Estimated subgroup-specific treatment effects

τ_1	τ_2	τ_3	τ_4
0.397	0.195	0.385	0.384

Table 2.16. Estimated subgroup proportions

π_1	π_2	π_3	π_4	π_{12}	π_{13}	π_{14}	π_{23}	π_{24}	π_{34}
0.43	1.00	1.00	0.65	0.43	0.43	0.43	1.00	0.65	0.65

We can see that KA/VE has the lowest subgroup-specific treatment effect. CVD, TEC, and TEE have similar subgroup-specific treatment effects, with CVD's slightly higher than the other two. However, in terms of subgroup proportions, CVD only makes up 43% of the population, followed by TEE, which is 65%. Both TEC and KA/VE have the whole population as their effective subgroup, i.e., there is no heterogeneity in treatment response to TEC and KA/VE. As for the overlap subgroup proportions, patients who respond to CVD also respond to the other three treatments. Since TEC and KA/VE's effective subgroup are the whole population, we have $CVD \in TEE \in KA/VE = TEC$.

Given the estimates, we calculate the response rates of the treatments on its effective subgroup and the overall population at a single stage, presented in Table 2.17.

Table 2.17. Treatment effect and response rate on favorable subgroup and overall population

	Response rate on subgroup	Overall response rate
τ_1	59.8%	25.7%
τ_2	54.9%	54.9%
τ_3	59.5%	59.5%
τ_4	59.5%	38.7%

The overall response rates are similar to the observed per-course success rate [48], with 25.7% vs 28% for CVD, 54.9% vs 52% for KA/VE, 59.5% vs 57% for TEC, and 38.7% vs 45% for TEE.

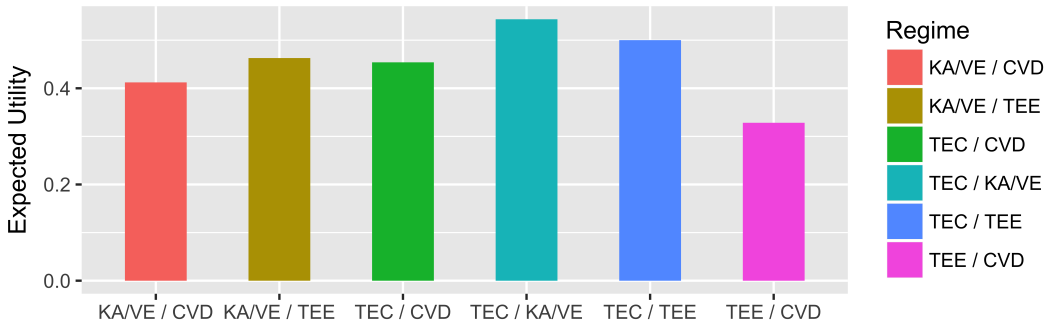

Figure 2.5. Response rates for each pair of treatments

The ultimate goal of this trial is to discover the optimal DTR. We calculate the overall response rate of each pair of treatments and the response rate after the first treatment. Figure 2.5 shows the overall response rates. Both CVD and TEE have pretty high response rates on their effective subgroups. However, due to the small subgroup proportions, the overall response rate is much lower than TEC and KA/VE.

As we can see, the pair of TEC and KA/VE has the highest overall response rate. We prefer assigning the treatment with a higher overall response rate first. Thus, the optimal DTR should be $TEC \rightarrow KA/VE$. This result is consistent with Thall (2007) [7], where TEC followed by KA/VE was found the most promising two-stage strategy by a conditional logistic regression in each course. Wang (2012) [68] also chose this as the optimal DTR under the binary utility score setting.

3. THE MIXTURE OF MIXED LOGIT MODELS WITH BASELINE COVARIATES AND TIME EFFECTS

In Chapter 2, we developed a model-based approach for determining the optimal dynamic treatment regime from binary response data generated under a SMART design. The proposed model explains overdispersion in the response through the inclusion of treatment subgroups and a subject-specific random effect. In this chapter, we expand our approach to include baseline covariates and time effects as additional sources of overdispersion.

Unique to our approach is that we separate the baseline covariates into two distinct groups. The first group contains the covariates that directly explain some of the variation in treatment response. We include them along with the random subject effect in the response component of our model. Examples of these covariates might be age, sex, and various health comorbidities. The other group of covariates that don't directly impact an individual's response, but rather provides information on the individual's latent treatment subgroup. Examples of these covariates are a marker for a genotype that renders certain treatments ineffective or an immune response that is associated with a marker's presence.

We also consider the inclusion of time effects to account for the potential of decreased treatment effectiveness over time. This time-varying covariate directly impacts the treatment response and thus can be considered to be a covariate in the first group. We consider the effect of time both numerically and categorically.

Because the inclusion of these covariates does not alter the basic structure of our model, we only highlight the changes necessary to incorporate them in this chapter. The remainder of Chapter 3 is similar to Chapter 2. We describe our estimation algorithm and approach to determine the optimal DTR given the estimated model. This is followed by simulation studies, including studies comparing our approach to Q-learning.

3.1 The Model-based Approach for Determining the Optimal DTR

For each individual, we now consider a baseline covariate vector $\boldsymbol{X}$ that can be partitioned into $\boldsymbol{X_Z}$, the covariates associated with the latent subgroup identity, and $\boldsymbol{X_R}$, the covariates

directly associated with the response to treatment. We assume that there is no overlap between the covariates in $\boldsymbol{X_Z}$ and $\boldsymbol{X_R}$.

3.1.1 The Mixture Model for Binary Outcomes

Similar to Chapter 2, we consider a T-stage sequence of treatment assignments and responses labeled: $A_1, Y_1, ..., A_T, Y_T$. We again assume that there are the same K treatments available for assignment at each stage and denote the treatment space and response space as $\mathcal{A} = \{\omega_1, ..., \omega_K\}$ and $\mathcal{Y} = \{0, 1\}$, respectively. Given the inclusion of covariates, we now include the baseline covariate space $\mathcal{X} = (\mathcal{X}_Z, \mathcal{X}_R)$, resulting in the space of histories $\mathcal{H}_{t-1} = \{\mathcal{X}, (\mathcal{A} \times \mathcal{Y})^{t-1}\}$ for Stage t and the trajectory space $\mathcal{H}_T = \{\mathcal{X}, (\mathcal{A} \times \mathcal{Y})^T\}$.

When baseline covariates are considered in a T-stage DTR $D = (d_1, ..., d_T)$, the t^{th} stage decision rule d_t becomes a mapping from a patient history $H_{t-1} = (\boldsymbol{X}, A_1, Y_1, ..., A_{t-1}, Y_{t-1})$ to $\mathcal{A}$. We again define $\Gamma_{D_{t-1}}$ to be the collection of all possible histories under DTR D at Stage t. This can be expressed

$$\Gamma_{D_{t-1}} = \{(\boldsymbol{x}, a_1, y_1, ..., a_{t-1}, y_{t-1}) \in \mathcal{H}_{t-1} | P_D(\boldsymbol{x}, a_1, y_1, ..., a_{t-1}, y_{t-1}) > 0\},$$

with the collection of all possible trajectories for DTR D expressed as:

$$\Gamma_D = \{(\boldsymbol{x}, a_1, y_1, ..., a_T, y_T) \in \mathcal{H}_T | P_D(\boldsymbol{x}, a_1, y_1, ..., a_T, y_T) > 0\}$$

In Chapter 2, we computed the probability of each trajectory by first considering the conditional probability of each trajectory based on a patient's latent subgroup vector $\boldsymbol{Z} = (Z_{\omega_1}, ..., Z_{\omega_K})$ and the subject-specific effect S. Since we are including baseline covariates to describe the continuous heterogeneity, this conditional probability now also depends on the baseline covariate vector $\boldsymbol{X_R}$. Note that the covariates $\boldsymbol{X_Z}$ are not needed here because we are conditioning on $\boldsymbol{Z}$.

Imposing the same assumptions considered in Section 2.1.1, the conditional probability of a trajectory given $\boldsymbol{x_R}$, $\boldsymbol{z}$ and s can be broken down into the product of sequential conditional probabilities:

$$\begin{aligned} P_D(a_1, y_1, ..., a_T, y_T|\boldsymbol{x_R}, \boldsymbol{z}, s) =& P(y_1|\boldsymbol{x_R}, \boldsymbol{z}, a_1 = d_1(\boldsymbol{x}), s) \\ & \times \prod_{t=2}^{T} P(y_t|\boldsymbol{x_R}, a_1, y_1, ..., a_t = d_t(\boldsymbol{x}, a_1, ..., y_{t-1}), \boldsymbol{z}, s) \end{aligned}$$

which can be further simplified to:

$$\begin{aligned} P_D(a_1, y_1, ..., a_T, y_T|\boldsymbol{x_R}, \boldsymbol{z}, s) =& P(y_1|\boldsymbol{x_R}, \boldsymbol{z}, a_1 = d_1(\boldsymbol{x}), s) \\ & \times \prod_{t=2}^{T} P(y_t|\boldsymbol{x_R}, a_t = d_t(\boldsymbol{x}, a_1, ..., y_{t-1}), \boldsymbol{z}, s) \end{aligned} \tag{3.1}$$

Each of these Bernoulli probabilities in the product on the right is modeled using an extension of the logit function of Chapter 2. We include baseline covariates $\boldsymbol{x}$ in the indicator function $\mathbb{1}\{d_t(x, a_1, y_1, ..., a_{t-1}, y_{t-1}) = \omega_k\}$ which represents whether treatment ω_k is assigned at the t^{th} stage. We also allow for the interactions between $\boldsymbol{x_R}$ and the treatments in the model. Letting p_t represent the probability that the assigned treatment is effective (i.e., $P(y_t = 1|x_R, a_t = d_t(x, a_1, y_1, ..., y_{t-1}), z, s)$), the logit function is:

$$\begin{aligned} \log(\frac{p_t}{1-p_t}) =& \mu + \boldsymbol{x_R}\boldsymbol{\alpha} \\ & + \sum_{k=1}^{K} \mathbb{1}\{d_t(\boldsymbol{x}, a_1, y_1, ..., a_{t-1}, y_{t-1}) = \omega_k\} z_{\omega_k} \{\tau_{\omega_k} + \boldsymbol{x_R}\boldsymbol{\beta_{\omega_k}}\} + s \end{aligned} \tag{3.2}$$

where $\boldsymbol{x_R}$, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta_{\omega_k}} \in \mathbb{R}^{p_1}$. The vector $\boldsymbol{\alpha}$ quantifies the effect that the covariates have on the logit regardless of treatment while the vector $\boldsymbol{\beta_{\omega_k}}$ quantifies any treatment-specific effects of the covariates.

Because z_{ω_k} is binary, this logit function $\log(\frac{p_t}{1-p_t})$ can also be expressed as:

$$\begin{cases} \mu + \boldsymbol{x_R}\boldsymbol{\alpha} + s, & \text{if } z_{\omega_k} = 0 \\ (\mu + \tau_{\omega_k}) + \boldsymbol{x_R}(\boldsymbol{\alpha} + \boldsymbol{\beta_{\omega_k}}) + s, & \text{if } z_{\omega_k} = 1 \end{cases} \tag{3.3}$$

To get the marginal probability of each trajectory with baseline covariates $\boldsymbol{X}$, as we did in Chapter 2, we first integrate out the subject-specific effect s:

$$\begin{aligned} & P_D(a_1, y_1, ..., a_T, y_T | \boldsymbol{x}, \boldsymbol{z}) \\ & \quad = \int_s P_D(a_1, y_1, ..., a_T, y_T | \boldsymbol{x_R}, \boldsymbol{z}, s) \phi(s) ds \\ & \quad = \int_s \prod_{t=1}^{T} P_D(y_t | \boldsymbol{x_R}, a_t = d_t(\boldsymbol{x}, a_1, y_1, ..., a_{t-1}, y_{t-1}), \boldsymbol{z}, s) \phi(s) ds \end{aligned}$$

where $\phi(s)$ is the $N(0, \sigma^2)$ density function. We then weigh these subgroup-specific trajectory probabilities by the appropriate subgroup probabilities, which now depend on the covariates $\boldsymbol{x_Z}$. Labelling these $P(\boldsymbol{z}|\boldsymbol{x}_z)$, the trajectory probability is:

$$P_D(a_1, y_1, ..., a_T, y_T | X = (\boldsymbol{x_Z}, \boldsymbol{x_R})) = \sum_{\boldsymbol{z} \in \mathcal{Z}} P(\boldsymbol{z}|\boldsymbol{x_Z}) P_D(a_1, y_1, ..., a_T, y_T | \boldsymbol{x}, \boldsymbol{z}). \tag{3.4}$$

We'll now discuss our model component for the subgroup probabilities $P(\boldsymbol{z}|\boldsymbol{x_Z})$.

3.1.2 Multivariate Bernoulli Subgroup Model

For a K-treatment trial, we define the subgroup vector as $\boldsymbol{Z} = (Z_{\omega_1}, ..., Z_{\omega_K})$, where each element Z_{ω_k} denotes whether the patient can respond favorably (1) or not (0) to treatment ω_k. In other words, each element follows a Bernoulli distribution. In Chapter 2, we did not model these Bernoulli random variables because they were the same for each subject. Instead, we considered the multinomial distribution of the 2^K resulting subgroups. However, given baseline covariates, we think that it is more natural and flexible to develop a model for these random variables and therefore adopt a multivariate Bernoulli model [78], [79].

To introduce this model, we first ignore covariates and illustrate the bivariate Bernoulli model for the simplest scenario involving only two treatments and then extend to the multivariate Bernoulli model for K treatments. In each case, we discuss the underlying subgroup structure (i.e., the relationships among the Bernoulli random variables). Our focus will be on the independence and the homogeneous association structures. We then include the covariates into the model and present an example involving four treatments.

Bivariate Bernoulli Model

Given two treatments, the subgroup vector is $\boldsymbol{Z} = (Z_1, Z_2)$. There are four possible subgroups and we label their probabilities $\pi_{z_1 z_2} = P(Z_1 = z_1, Z_2 = z_2)$ such that $\sum_{i=0,1} \sum_{j=0,1} \pi_{ij} = 1$. In Chapter 2, we directly estimated these $\pi_{z_1 z_2}$ but here we'll consider the distribution of $\boldsymbol{Z}$ expressed as follows:

$$\begin{aligned} P(\boldsymbol{Z} = (z_1, z_2)) &= \pi_{11}^{z_1 z_2} \pi_{10}^{z_1(1-z_2)} \pi_{01}^{(1-z_1)z_2} \pi_{00}^{(1-z_1)(1-z_2)} \\ &= \pi_{00} (\pi_{10}/\pi_{00})^{z_1} (\pi_{01}/\pi_{00})^{z_2} ((\pi_{11}\pi_{00})/(\pi_{10}\pi_{01}))^{z_1 z_2} \\ &= \exp\{\log(\pi_{00}) + z_1 \log \frac{\pi_{10}}{\pi_{00}} + z_2 \log \frac{\pi_{01}}{\pi_{00}} + z_1 z_2 \log \frac{\pi_{11}\pi_{00}}{\pi_{10}\pi_{01}}\} \end{aligned} \tag{3.5}$$

From Equation (3.5), we can see that $\boldsymbol{Z} = (Z_1, Z_2)$ belongs to the exponential family with Z_1, Z_2, and $Z_1 Z_2$ as the sufficient statistics and the natural parameters:

$$\begin{aligned} f_1 &= \log \frac{\pi_{10}}{\pi_{00}} \\ f_2 &= \log \frac{\pi_{01}}{\pi_{00}} \\ f_{12} &= \log \frac{\pi_{11}\pi_{00}}{\pi_{10}\pi_{01}} \end{aligned}$$

The parameter f_1 is the log odds of $Z_1 = 1$ conditional on $Z_2 = 0$, f_2 is the log odds of $Z_2 = 1$ conditional on $Z_1 = 0$, and f_{12} is the log odds ratio for $\boldsymbol{Z} = (Z_1, Z_2)$.

The subgroup probabilities $\pi_{z_1 z_2}$ can be expressed in terms of the natural parameters as follows:

$$\begin{aligned} \pi_{00} &= \frac{1}{1 + \exp\{f_1\} + \exp\{f_2\} + \exp\{f_1 + f_2 + f_{12}\}} \\ \pi_{10} &= \frac{\exp\{f_1\}}{1 + \exp\{f_1\} + \exp\{f_2\} + \exp\{f_1 + f_2 + f_{12}\}} \\ \pi_{01} &= \frac{\exp\{f_2\}}{1 + \exp\{f_1\} + \exp\{f_2\} + \exp\{f_1 + f_2 + f_{12}\}} \\ \pi_{11} &= \frac{\exp\{f_1 + f_2 + f_{12}\}}{1 + \exp\{f_1\} + \exp\{f_2\} + \exp\{f_1 + f_2 + f_{12}\}} \end{aligned}$$

and the marginal probabilities are:

$$P(Z_1 = 1) = \frac{\exp\{f_1\} + \exp\{f_1 + f_2 + f_{12}\}}{1 + \exp\{f_1\} + \exp\{f_2\} + \exp\{f_1 + f_2 + f_{12}\}}$$
$$P(Z_2 = 1) = \frac{\exp\{f_2\} + \exp\{f_1 + f_2 + f_{12}\}}{1 + \exp\{f_1\} + \exp\{f_2\} + \exp\{f_1 + f_2 + f_{12}\}}$$

One major advantage of the multivariate Bernoulli distribution is its capability of modeling the association between the Bernoulli RVs. For example, in the bivariate setting, the covariance between Z_1 and Z_2 is

$$\begin{aligned}
cov(Z_1, Z_2) &= E(Z_1 - E(Z_1))(Z_2 - E(Z_2)) \\
&= E(Z_1 - P(Z_1 = 1))(Z_2 - P(Z_2 = 1)) \\
&= E(Z_1 Z_2) - P(Z_1 = 1)E(Z_2) - P(Z_2 = 1)E(Z_1) + P(Z_1 = 1)P(Z_2 = 1) \\
&= \boldsymbol{\pi}_{11} - P(Z_1 = 1)P(Z_2 = 1) \\
&= \frac{\exp\{f_1 + f_2 + f_{12}\} - \exp\{f_1 + f_2\}}{(1 + \exp\{f_1\} + \exp\{f_2\} + \exp\{f_1 + f_2 + f_{12}\})^2}
\end{aligned}$$

When $f_{12} = 0$, the numerator is 0 implying the RVs are uncorrelated. Dai (2013) [79] delves further into these relationships and in Lemma 2.1, confirms that Z_1 and Z_2 are independent if and only if $f_{12} = 0$.

Multivariate Bernoulli Model

Extending this to the general K treatment setting, there are now 2^K subgroups with probabilities, denoted $\boldsymbol{\pi}_{z_1 z_2, \ldots z_K}$. We can express the density of $\boldsymbol{Z}$ as

$$\begin{aligned}
P(Z_1 = z_1, ..., Z_K = z_K) =& \boldsymbol{\pi}_{0\ldots0}^{\{\prod_{k=1}^{K}(1-z_k)\}} \times \boldsymbol{\pi}_{10\ldots0}^{\{z_1 \prod_{k=2}^{K}(1-z_k)\}} \times \boldsymbol{\pi}_{010\ldots0}^{\{(1-z_1)z_2 \prod_{k=3}^{K}(1-z_k)\}} \times \ldots \\
& \times \boldsymbol{\pi}_{1\ldots1}^{\{\prod_{k=1}^{K} z_k\}} \\
=& \prod_{z_1,\ldots,z_K \in \{0,1\}^K} \boldsymbol{\pi}_{z_1 \ldots z_K}^{\prod_{\{z_k|z_k=1\}} z_k \{\prod_{\{z_k|z_k=0\}}(1-z_k)\}} \\
=& \exp\Big(\sum_{z_1,\ldots,z_K \in \{0,1\}^K} \prod_{\{z_k|z_k=1\}} z_k \Big\{ \prod_{\{z_k|z_k=0\}} (1 - z_k) \Big\} \log \boldsymbol{\pi}_{z_1 \ldots z_K} \Big) \quad (3.6)
\end{aligned}$$

Same as in the two-treatment example, this distribution belongs to the exponential family. The sufficient statistics are the products of these Bernoulli RVs, i.e., $Z_{\mathrm{j}_1}...Z_{\mathrm{j}_r}$, for any $1 \leq \mathrm{j}_1 < ... < \mathrm{j}_r \leq K$. There are $(2^K - 1)$ sufficient statistics in total. To get the natural parameters, we first expand the entire part inside the exponential function in Equation (3.6) and then simplify by $z_{\mathrm{j}_1}...z_{\mathrm{j}_r}$'s. Specifically, to get the term $z_{\mathrm{j}_1}...z_{\mathrm{j}_r}$ from $\prod_{\{z_k|z_k=1\}} z_k\{\prod_{\{z_k|z_k=0\}}(1 - z_k)\} \log \boldsymbol{\pi}_{z_1...z_K}$, z_k has to be 0 for any $k \notin \{\mathrm{j}_1, ..., \mathrm{j}_r\}$. For any $\mathrm{j}_l \in \{\mathrm{j}_1, ..., \mathrm{j}_r\}$, if $z_{\mathrm{j}_l} = 1$, we can have z_{j_l} as a positive product in $z_{\mathrm{j}_1}...z_{\mathrm{j}_r}$. On the other hand, if $z_{\mathrm{j}_l} = 0$, z_{j_l} will bring a negative sign to $z_{\mathrm{j}_1}...z_{\mathrm{j}_r}$. Therefore, the natural parameter of the sufficient statistic $z_{\mathrm{j}_1}...z_{\mathrm{j}_r}$ can be expressed as:

$$
\begin{aligned}
& f_{\mathrm{j}_1...\mathrm{j}_r} \\
&= \sum_{1\leq \mathrm{j}_1 \leq ... \leq \mathrm{j}_r \leq K} (-1)^{\prod_{l=1}^{r} \mathbb{1}_{z_{\mathrm{j}_l}=0}} \log P(Z_{\mathrm{j}_1} = z_{\mathrm{j}_1}, ..., Z_{\mathrm{j}_r} = z_{\mathrm{j}_r}, \text{and others are all zero})) \\
&= \log \frac{\prod P(\text{even \# zeros among } z_{\mathrm{j}_1}, ..., z_{\mathrm{j}_r} \text{ and others are all zero})}{\prod P(\text{odd \# zeros among } z_{\mathrm{j}_1}, ..., z_{\mathrm{j}_r} \text{ and others are all zero})}
\end{aligned}
$$

Then $P(Z_1 = z_1, ..., Z_K = z_K)$ can be written as:

$$
P(Z_1 = z_1, ..., Z_K = z_K) = \exp\{\log(\boldsymbol{\pi}_{0...0}) + \sum_{r=1}^{K} \sum_{1\leq \mathrm{j}_1 \leq ... \leq \mathrm{j}_r \leq K} z_{\mathrm{j}_1}...z_{\mathrm{j}_r} f_{\mathrm{j}_1,...,\mathrm{j}_r}\} \tag{3.7}
$$

To simplify the notation, we follow Dai (2013) [79] and denote the quantity S as:

$$
S^{\mathrm{j}_1\mathrm{j}_2...\mathrm{j}_r} = \sum_{1\leq \mathrm{j}_s \leq r} f_{\mathrm{j}_s} + \sum_{1\leq \mathrm{j}_s < \mathrm{j}_k \leq r} f_{\mathrm{j}_s\mathrm{j}_k} + ... + f_{\mathrm{j}_1\mathrm{j}_2...\mathrm{j}_r} \tag{3.8}
$$

and

$$
b(f) = \log\{1 + \sum_{r=1}^{K} \sum_{1\leq \mathrm{j}_1 < ... < \mathrm{j}_r \leq K} \exp\{S^{\mathrm{j}_1\mathrm{j}_2...\mathrm{j}_r}\}\}. \tag{3.9}
$$

After parameter transformation, we get:

$$
\boldsymbol{\pi}_{z_{\mathrm{j}_1}=z_{\mathrm{j}_2}=...=z_{\mathrm{j}_r}=1 \text{ others are } 0} = \frac{\exp(S^{\mathrm{j}_1\mathrm{j}_2...\mathrm{j}_r})}{\exp(b(f))} \tag{3.10}
$$

and

$$\boldsymbol{\pi}_{z_{0\ldots0}} = \frac{1}{\exp(b(f))}. \tag{3.11}$$

Same as for the bivariate Bernoulli model, it is of great interest to understand the subgroup structure of $\boldsymbol{Z}$. Equations (3.10) and (3.11) present the multivariate Bernoulli model with all nonzero natural parameters, commonly referred to as the saturated model. The number of natural parameters increases exponentially with K. This can complicate estimation of the saturated model. Furthermore, the underlying structure may be simpler than the saturated model structure or approximated well using a simpler structure. The two we consider here are the independence and homogeneous association structures.

The independence structure was introduced in the bivariate case. It states that subjects responding to one treatment is uncorrelated to how they respond to other treatments. In other words, the subgroup indicators Z_1,...,Z_K are element-wise independent, and the probability of a subject responding favorably to a set of treatments is simply the product of the probability of responding favorably to each treatment. According to Theorem 3.1 in Dai (2013) [79], subgroup vector $\boldsymbol{Z} = (Z_1, ..., Z_K)$ are independent element-wise if and only if

$$f_{\mathrm{j}_1\ldots,\mathrm{j}_r} = 0, \text{ for } 1 \leq \mathrm{j}_1 < .. < \mathrm{j}_r \leq K, r \geq 2.$$

In other words, only f_1, f_2,...,f_K are non-zero natural parameters.

The homogeneous association structure incorporates an underlying relationship between pairs of subgroups. It is also known as the structure with no non-zero second or higher order interactions. This results in the association between any two treatment subgroup indicators Z_{i} and Z_{j} being the same, regardless of other treatment subgroup indicators. Based on Theorem 3.2 in Dai (2013) [79], each pair of $\boldsymbol{Z} = (Z_1, ..., Z_K)$ has homogeneous association if and only if:

$$f_{\mathrm{j}_1\ldots,\mathrm{j}_r} = 0, \text{ for } 1 \leq \mathrm{j}_1 < .. < \mathrm{j}_r \leq K, r \geq 3.$$

The Inclusion of Covariates

As discussed earlier, the multivariate Bernoulli distribution belongs to the exponential family. Therefore, we can model each of the natural parameters (f's) as a linear function of the p_2 covariates, $\boldsymbol{X_Z} = (1, x_{Z,1}, x_{Z,1}, ..., x_{Z,p_2})$. Letting λ belong to the power set of $\{1, 2, ..., K\}$, the natural parameter f_λ can be written $f_\lambda = \gamma_{\lambda,0} + \gamma_{\lambda,1} x_{Z,1} + ... + \gamma_{\lambda,p_2} x_{Z,p_2} = \boldsymbol{x_Z \gamma_\lambda}$, where $\boldsymbol{\gamma_\lambda} = (\gamma_{\lambda,0}, ..., \gamma_{\lambda,p_2})$ is the coefficient vector of f_λ. For the saturated model, there are $(2^K - 1)$ natural parameters meaning that this model requires $(2^K - 1) \times (p_2 + 1)$ parameters. The number of parameters can be reduced by considering a simpler association structure. For example, the independence structure involves only $K \times (p_2 + 1)$ parameters.

The set of binary regressions can be written:

$$\begin{aligned}
&\log \frac{P(Z_{\mathrm{j}_1} = Z_{\mathrm{j}_2} = ... = Z_{\mathrm{j}_r} = 1 \text{ others are } 0 | \boldsymbol{x_z})}{P(z_1 = 0, ..., z_K = 0 | \boldsymbol{x_z})} \\
&\quad = S^{\mathrm{j}_1 \mathrm{j}_2 ... \mathrm{j}_r} \\
&\quad = \sum_{1 \le \mathrm{j}_s \le r} f_{\mathrm{j}_s} + \sum_{1 \le \mathrm{j}_s < \mathrm{j}_k \le r} f_{\mathrm{j}_s \mathrm{j}_k} + ... + f_{\mathrm{j}_1 \mathrm{j}_2 ... \mathrm{j}_r} \\
&\quad = \boldsymbol{x_z} (\sum_{1 \le \mathrm{j}_s \le r} \boldsymbol{\gamma_{\mathrm{j}_s}} + \sum_{1 \le \mathrm{j}_s < \mathrm{j}_k \le r} \boldsymbol{\gamma_{\mathrm{j}_s \mathrm{j}_k}} + ... + \boldsymbol{\gamma_{\mathrm{j}_1 \mathrm{j}_2 ... \mathrm{j}_r}}),
\end{aligned} \tag{3.12}$$

and the subgroup probabilities $P(\boldsymbol{z}|\boldsymbol{x_z})$ are:

$$\begin{aligned}
&P(Z_{\mathrm{j}_1} = Z_{\mathrm{j}_2} = ... = Z_{\mathrm{j}_r} = 1 \text{ others are } 0 | \boldsymbol{x_z}) \\
&\quad = \frac{\boldsymbol{x_z} (\sum_{1 \le \mathrm{j}_s \le r} \boldsymbol{\gamma_{\mathrm{j}_s}} + \sum_{1 \le \mathrm{j}_s < \mathrm{j}_k \le r} \boldsymbol{\gamma_{\mathrm{j}_s \mathrm{j}_k}} + ... + \boldsymbol{\gamma_{\mathrm{j}_1 \mathrm{j}_2 ... \mathrm{j}_r}})}{1 + \sum_{r=1}^{K} \boldsymbol{x_z} (\sum_{1 \le \mathrm{j}_s \le r} \boldsymbol{\gamma_{\mathrm{j}_s}} + \sum_{1 \le \mathrm{j}_s < \mathrm{j}_k \le r} \boldsymbol{\gamma_{\mathrm{j}_s \mathrm{j}_k}} + ... + \boldsymbol{\gamma_{\mathrm{j}_1 \mathrm{j}_2 ... \mathrm{j}_r}})}
\end{aligned} \tag{3.13}$$

and

$$\begin{aligned}
&P(Z_1 = 0, ..., Z_K = 0 | \boldsymbol{x_z}) \\
&\quad = \frac{1}{1 + \sum_{r=1}^{K} \sum_{1 \le \mathrm{j}_1 < ... < \mathrm{j}_r \le K} P(z_{\mathrm{j}_1} = z_{\mathrm{j}_2} = ... = z_{\mathrm{j}_r} = 1 \text{ others are } 0 | \boldsymbol{x_z})}
\end{aligned} \tag{3.14}$$

We want to point out that this set of binary logistic regression models resembles multinomial logistic regression. In fact, when the multivariate Bernoulli model is saturated, it is

equivalent to multinomial logistic regression. Our model in Chapter 2 can be viewed as the saturated multivariate Bernoulli model without baseline covariates.

Now we consider the example of MD Anderson's advanced prostate cancer trial that is explained in Section 2.3.2. There are four possible treatments of consideration. We denote the subgroup vector as $\boldsymbol{Z} = (Z_1, Z_2, Z_3, Z_4)$. Because of the complexity of the saturated model, even with one covariate, the number of parameters can be as large as 30. To avoid overfitting and reduce the complexity, we focus on two simpler models, the independence model and the homogeneous association model, and assume that these reduced relationship models can adequately approximate the probabilities of belonging to the beneficial subgroups of the one or two assigned treatments. For the independence model, we have 4 non-zero natural parameters f_1, f_2, f_3, f_4, and 10 non-zero natural parameters f_1, f_2, f_3, f_4, f_{12}, f_{13}, f_{14}, f_{23}, f_{24}, and f_{34} for the homogeneous association model. Table 3.1 presents the set of binary regressions for both models with baseline covariates included.

3.1.3 Evaluating and Determining the Optimal DTR

Unlike in Chapter 2 where the trajectory probabilities are the same for the entire population, here the trajectory probabilities are different given different baseline covariates $\boldsymbol{X}$. Now that we've described our model components for the subgroup probabilities $P(\boldsymbol{z}|\boldsymbol{X_Z})$ and for each of the responses, we can compute the collection of trajectory probabilities given the baseline covariates $\boldsymbol{X}$. Each of these is obtained by summing over all subgroups:

$$P_D(a_1, y_1, ..., a_T, y_T|\boldsymbol{X} = (\boldsymbol{x_Z}, \boldsymbol{x_R})) = \sum_{\boldsymbol{z}} P(\boldsymbol{z}|\boldsymbol{x_Z}) P_D(a_1, y_1, ..., a_T, y_T|\boldsymbol{x}, \boldsymbol{z})$$

These probabilities make up the probability mass function of T-stage trajectories given $\boldsymbol{X}$ and the chosen DTR D.

Table 3.1. Binary regressions for independence and homogeneous association models

Binary regressions	Independence model	Homogeneous association model
$\log \frac{P(Z_{\mathrm{j}}=1 \text{ and others are } 0 \mid \boldsymbol{x_z})}{P(Z=(0,0,0,0) \mid \boldsymbol{x_z})}$	$\boldsymbol{x_z}\boldsymbol{\gamma}_{\mathrm{j}}$	$\boldsymbol{x_z}\boldsymbol{\gamma}_{\mathrm{j}}$
$\log \frac{P(Z_{\mathrm{j}}=Z_k=1 \text{ and others are } 0 \mid \boldsymbol{x_z})}{P(Z=(0,0,0,0) \mid \boldsymbol{x_z})}$	$\boldsymbol{x_z}(\boldsymbol{\gamma}_{\mathrm{j}} + \boldsymbol{\gamma}_k)$	$\boldsymbol{x_z}(\boldsymbol{\gamma}_{\mathrm{j}} + \boldsymbol{\gamma}_k + \boldsymbol{\gamma}_{\mathrm{j}k})$
$\log \frac{P(Z_{\mathrm{j}}=Z_k=Z_l=1 \text{ the other is } 0 \mid \boldsymbol{x_z})}{P(Z=(0,0,0,0) \mid \boldsymbol{x_z})}$	$\boldsymbol{x_z}(\boldsymbol{\gamma}_{\mathrm{j}} + \boldsymbol{\gamma}_k + \boldsymbol{\gamma}_l)$	$\boldsymbol{x_z}\big(\boldsymbol{\gamma}_{\mathrm{j}} + \boldsymbol{\gamma}_k + \boldsymbol{\gamma}_l + \boldsymbol{\gamma}_{\mathrm{j}k} + \boldsymbol{\gamma}_{\mathrm{j}l} + \boldsymbol{\gamma}_{kl}\big)$
$\log \frac{P(Z=(1,1,1,1) \mid \boldsymbol{x_z})}{P(Z=(0,0,0,0) \mid \boldsymbol{x_z})}$	$\boldsymbol{x_z}(\boldsymbol{\gamma_1} + \boldsymbol{\gamma_2} + \boldsymbol{\gamma_3} + \boldsymbol{\gamma_4})$	$\boldsymbol{x_z}\big(\boldsymbol{\gamma_1} + \boldsymbol{\gamma_2} + \boldsymbol{\gamma_3} + \boldsymbol{\gamma_4} + \boldsymbol{\gamma_{12}} + \boldsymbol{\gamma_{13}} + \boldsymbol{\gamma_{14}} + \boldsymbol{\gamma_{23}} + \boldsymbol{\gamma_{24}} + \boldsymbol{\gamma_{34}}\big)$

Similar to Section 2.1.2, a utility function $r(a_1, y_1, ..., a_T, y_T)$ is defined to evaluate each trajectory. The value of the DTR D given $\boldsymbol{x}$ is:

$$\begin{aligned}
V(D|\boldsymbol{x}) =& E_D(r(A_1, Y_1, ..., A_T, Y_T)|\boldsymbol{x}) \\
=& \sum_{a_1,y_1,...,a_T,y_T \in \Gamma_d} r(a_1, y_1, ..., a_T, y_T) P_D(a_1, y_1, ..., a_T, y_T|\boldsymbol{x}) \\
=& \sum_{a_1,y_1,...,a_T,y_T \in \Gamma_d} r(a_1, y_1, ..., a_T, y_T) \times \{\sum_{\boldsymbol{z}} P(\boldsymbol{z}|\boldsymbol{x_Z}) P_D(a_1, ..., y_T|\boldsymbol{x}, \boldsymbol{z})\} \\
=& \sum_{a_1,y_1,...,a_T,y_T \in \Gamma_d} r(a_1, y_1, ..., a_T, y_T) \\
& \times \sum_{\boldsymbol{z}} \{P(\boldsymbol{z}|\boldsymbol{x_Z}) \int_s \prod_{t=1}^{T} P_D(y_t|\boldsymbol{x_R}, a_t = d_t(\boldsymbol{x}, a_1, y_1, ..., a_{t-1}, y_{t-1}), \boldsymbol{z}, s)\phi(s)ds\}
\end{aligned} \tag{3.15}$$

Within the scope of the data used to estimate the model, our model allows for the determination of the individualized optimal DTR for any $\boldsymbol{x}$, which is defined as:

$$D_{\boldsymbol{x}}^{opt} = \arg\max_D V(D|\boldsymbol{x}) \tag{3.16}$$

Example 3.1

Suppose we observe two baseline covariates under the 4-treatment MD Anderson prostate trial design introduced in Section 2.3.2. We'll assume one covariate is in X_Z and one is in X_R, with both ranging from -3 to 3. For the treatment response, similar as in Section 2.3.2, we assume that the natural response rate is zero. As a result, the probability of a successful response $p_t = P(Y_y = 1|X_R, a_t, \boldsymbol{z}, s)$ can be expressed

$$\begin{aligned}
\log(\frac{p_t}{1-p_t}) = \tau_{a_t} + x_R\beta_{a_t} + s, & \qquad \text{if } z_{a_t} = 1 \\
p_t = 0, & \qquad \text{if } z_{a_t} = 0
\end{aligned} \tag{3.17}$$

For the subgroup probability component, we consider the independence structure. Tables 3.2 and 3.3 summarize the parameter settings.

Table 3.2. Treatment response parameter settings for Example 3.1

τ_1	β_1	τ_2	β_2	τ_3	β_3	τ_4	β_4	σ^2
0.8	-0.5	0.3	0.6	1.1	0.4	0.6	1	0.01

Table 3.3. Subgroup parameter settings for Example 3.1

$\gamma_{1,0}$	$\gamma_{1,1}$	$\gamma_{2,0}$	$\gamma_{2,1}$	$\gamma_{3,0}$	$\gamma_{3,1}$	$\gamma_{4,0}$	$\gamma_{4,1}$
0.3	1.5	1	-3	-1	1	0.5	-1

Table 3.4. Possible DTRs found optimal in Example 3.1

DTR	Detailed DTR
1	Trt 1 first, switch to Trt 2 if observe a failure to the first treatment
2	Trt 1 first, switch to Trt 3 if observe a failure to the first treatment
3	Trt 1 first, switch to Trt 4 if observe a failure to the first treatment
4	Trt 2 first, switch to Trt 3 if observe a failure to the first treatment
5	Trt 2 first, switch to Trt 4 if observe a failure to the first treatment
6	Trt 3 first, switch to Trt 4 if observe a failure to the first treatment

We investigated all 12 DTRs described in Chapter 2. To assess the value of each DTR, we use the same utility that we used in Example 2.1–a trajectory is scored 1 if two consecutive successes are observed and 0 otherwise. Possible DTRs that were found to be optimal are listed in Table 3.4.

Figure 3.1 displays the optimal DTRs over the ranges of X_Z and X_R. There are five regions, each involving a different pair of treatments. Treatments 1, 2, or 3 are assigned initially, followed by either Treatment 3 or 4.

3.2 Parameter Estimation and DTR Evaluation

In practice, we need to estimate the model parameters. We again consider data collected from a SMART design, where treatments are assigned to patients according to a certain pre-specified randomization scheme that does not depend on the baseline covariates. Similar to what we did in Chapter 2, we detail the model modifications needed to accommodate this randomization and then describe our EM algorithm to obtain the estimates.

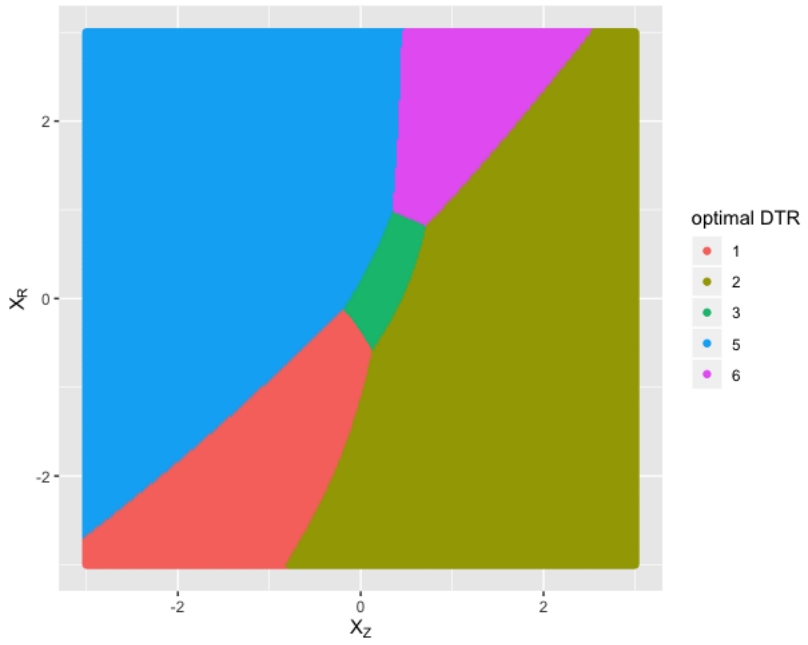

Figure 3.1. Optimal DTR given X of Example 3.1

3.2.1 Model Specification for Randomized Trials

Consider an individual in a T-stage randomized trial involving K possible treatments. For each subject, we have baseline covariates $\boldsymbol{X} = (\boldsymbol{X_Z}, \boldsymbol{X_R})$, and a sequence of treatment assignments (A) and responses (Y) labeled $A_1, Y_1, ..., A_T, Y_T$. Note that we do not distinguish the responses in a randomized trial and DTR as we did in Chapter 2 when we used Y for a DTR and O for the randomized trial.

The conditional distribution of a sequence given the treatment subgroup vector $\boldsymbol{z}$ and subject-specific random effect s and covariates $\boldsymbol{x_R}$ is:

$$\begin{aligned} &P(a_1, y_1, ..., a_T, y_T | \boldsymbol{x_R}, \boldsymbol{z}, s) \\ &\quad = P(a_1)P(y_1|\boldsymbol{x_R}, a_1, \boldsymbol{z}, s)\prod_{t=2}^{T} P(a_t|a_1, y_1, ..., a_{t-1}, y_{t-1})P(y_t|\boldsymbol{x_R}, a_1, y_1, ..., a_t, \boldsymbol{z}, s). \end{aligned}$$

In contrast to Equation (3.1), the treatment assignment a_t is determined by a pre-specified randomization distribution $P(a_t|a_1, y_1, ..., a_{t-1},\ y_{t-1})$. Under our model assumptions, this conditional distribution simplifies to:

$$\begin{aligned} &P(a_1,y_1, ..., a_T, y_T|\boldsymbol{x_R}, \boldsymbol{z}, s) \\ &\quad = P(a_1)P(y_1|\boldsymbol{x_R}, a_1, \boldsymbol{z}, s) \times \prod_{t=2}^{T} P(a_t|a_1, y_1, ..., a_{t-1}, y_{t-1})P(y_t|\boldsymbol{x_R}, a_t, \boldsymbol{z}, s), \end{aligned} \tag{3.18}$$

with each response being Bernoulli with probability p_t, whose logit function is:

$$\log(\frac{p_t}{1-p_t}) = \mu + x_R\alpha + \sum_{k=1}^{K} \mathbb{1}\{A_t = \omega_k\} z_{\omega_k}\{\tau_{\omega_k} + x_R\beta_{\omega_k}\} + s$$

where $\mathbb{1}\{A_t = \omega_k\}$ indicates that treatment ω_k is assigned at the t^{th} stage.

Now consider a group of N individuals. Depending on the choice of trial design, the number of stages could vary across individuals, so we use T_{i} denote the number of stages for the i^{th} subject. Similarly, $\boldsymbol{A}^{\mathrm{i}} = (a_1^{\mathrm{i}}, ..., a_{T_{\mathrm{i}}}^{\mathrm{i}})$ and $\boldsymbol{Y}^{\mathrm{i}} = (y_1^{\mathrm{i}}, ..., y_{T_{\mathrm{i}}}^{\mathrm{i}})$ are the assignment and outcome vectors, respectively. Lastly, we denote baseline covariates $\boldsymbol{X}^{\mathrm{i}} = (\boldsymbol{X}_{\boldsymbol{R}}^{\mathrm{i}}, \boldsymbol{X}_{\boldsymbol{Z}}^{\mathrm{i}})$ and the latent subgroup vector $\boldsymbol{Z}^{\mathrm{i}} = (Z_1^{\mathrm{i}}, ..., Z_K^{\mathrm{i}})$.

Our goal is to estimate the latent subgroup parameters $\boldsymbol{\gamma}$, treatment effects $\boldsymbol{\beta}$ and $\boldsymbol{\tau}$, and subject effect variance σ^2. Let $\boldsymbol{\theta} = (\mu, \boldsymbol{\gamma}, \boldsymbol{\alpha}, \boldsymbol{\tau}, \boldsymbol{\beta}, \sigma^2)$ represent this collection of model parameters. The likelihood of this mixtures model is given by:

$$L(\boldsymbol{\theta}) = \prod_{\mathrm{i}=1}^{N} \{\sum_{\boldsymbol{Z}^{\mathrm{i}}} P(\boldsymbol{Z}^{\mathrm{i}}|\boldsymbol{X}_{\boldsymbol{Z}}^{\mathrm{i}}) \int \prod_{t=1}^{T_{\mathrm{i}}} \exp(y_t^{\mathrm{i}}\eta_t^{\mathrm{i}} - \log(1+\exp(\eta_t^{\mathrm{i}})))\phi(s)ds\} \tag{3.19}$$

where $\eta_t^{\mathrm{i}} = \mu + \boldsymbol{X_R\alpha} + \sum_{k=1}^{K} \mathbb{1}\{A_t = \omega_k\} z_{\omega_k}\{\tau_{\omega_k} + \boldsymbol{X_R\beta_{\omega_k}}\} + s$.

We use the EM algorithm, described in the next subsection, to obtain our parameter estimates. Having obtained our estimates, we can estimate the probabilities of all possible trajectories $P_D(a_1, y_1, ..., a_T, y_T|\boldsymbol{x})$ for any DTR D as

$$\sum_{\boldsymbol{z}} P(\boldsymbol{Z_k^{\mathrm{i}}} = \boldsymbol{z}|\boldsymbol{x_Z}) \int \prod_{t=1}^{T} \hat{P}(y_t|a_t = d_t(a_1, y_1, .., a_{t-1}, y_{t-1}), \boldsymbol{x_R}, \boldsymbol{z})\phi(s)ds,$$

and subsequently estimate the value of individualized DTR D:

$$\hat{V}(D|\boldsymbol{x}) = \sum_{a_1,y_1,...,a_T,y_T \in \Gamma_D} r(y_1,...,y_T)\hat{P}_D(a_1,y_1,...,a_T,y_T|\boldsymbol{x}) \tag{3.20}$$

We call

$$\hat{D}^{opt}_{\boldsymbol{x}} = \arg\max_D \hat{V}(D|\boldsymbol{x}) \tag{3.21}$$

as our estimated optimal DTR given baseline covariates $\boldsymbol{X} = \boldsymbol{x}$.

Similar to Chapter 2, the optimal DTR here can be found through an exhaustive search or dynamic programming. Considering that there is a limited number of DTRs of interest, given any $\boldsymbol{x}$, we use the model estimates to compute the values of all DTRs and find the one with the maximum value.

3.2.2 EM Algorithm

Because there are no closed-form MLE formulas for the likelihood in Equation (3.19), we use the EM algorithm to obtain estimates. As part of this algorithm, we need to integrate over the random subject effect S. We again adopt Gaussian quadrature to do this.

Similar to our approach in Chapter 2, we convert the subgroup vector $\boldsymbol{Z}$ into a vector $\boldsymbol{C}$ of length 2^K, each element representing one of the 2^K treatment subgroups. For the ith subject, the vector $\boldsymbol{C}^{\mathrm{i}} = (C^{\mathrm{i}}_1,...,C^{\mathrm{i}}_{2^K})$ will have one element equal to 1 and 0's otherwise. Unfortunately, which element is 1 is unknown to us and must be estimated. We use the same conversion method and thus ordering of the 2^K subgroups, as we did in Chapter 2 (see Section 2.2.2). We complement each vector $\boldsymbol{C}^{\mathrm{i}}$ with a subgroup proportion vector $\boldsymbol{\pi}$. This vector is simply a function of the covariates $\boldsymbol{X}^{\mathrm{i}}_{\boldsymbol{Z}}$ and the latent subgroup parameters $\boldsymbol{\gamma}$.

If $\boldsymbol{C}^{\mathrm{i}}$ were observed, the joint probability for the trajectory of the ith subject is

$$P(\boldsymbol{Y}^{\mathrm{i}},\boldsymbol{A}^{\mathrm{i}},\boldsymbol{C}^{\mathrm{i}},\boldsymbol{X}^{\mathrm{i}}|\boldsymbol{\theta}) = P(\boldsymbol{C}^{\mathrm{i}}|\boldsymbol{X}^{\mathrm{i}}_{\boldsymbol{Z}},\boldsymbol{\theta}) \times P(\boldsymbol{Y}^{\mathrm{i}},\boldsymbol{A}^{\mathrm{i}}|\boldsymbol{C}^{\mathrm{i}},\boldsymbol{X}^{\mathrm{i}}_{\boldsymbol{R}},\boldsymbol{\theta})$$

The complete-data log-likelihood for N subjects is

$$\begin{aligned}
&\log L(\boldsymbol{\theta}|\boldsymbol{Y},\boldsymbol{A},\boldsymbol{C},\boldsymbol{X}) \\
&\quad= \sum_{\mathrm{i}=1}^{N} \log\{\{P(\boldsymbol{C}^{\mathrm{i}}|\boldsymbol{X}^{\mathrm{i}}_{\boldsymbol{Z}},\boldsymbol{\theta}) \times P(\boldsymbol{Y}^{\mathrm{i}},\boldsymbol{A}^{\mathrm{i}}|\boldsymbol{C}^{\mathrm{i}},\boldsymbol{X}^{\mathrm{i}}_{\boldsymbol{R}},\boldsymbol{\theta})\}\} \\
&\quad= \sum_{\mathrm{i}=1}^{N} \{\log\{P(\boldsymbol{C}^{\mathrm{i}}|\boldsymbol{X}^{\mathrm{i}}_{\boldsymbol{Z}},\boldsymbol{\theta})\} + \log\{P(\boldsymbol{Y}^{\mathrm{i}},\boldsymbol{A}^{\mathrm{i}}|\boldsymbol{C}^{\mathrm{i}},\boldsymbol{X}^{\mathrm{i}}_{\boldsymbol{R}},\boldsymbol{\theta})\}\} \\
&\quad= \sum_{\mathrm{i}=1}^{N} \log\{P(\boldsymbol{C}^{\mathrm{i}}|\boldsymbol{X}^{\mathrm{i}}_{\boldsymbol{Z}},\boldsymbol{\theta}) + \log \int \prod_{t=1}^{T_{\mathrm{i}}} \exp(y^{\mathrm{i}}_t \eta^{\mathrm{i}}_t - \log(1+\exp(\eta^{\mathrm{i}}_t)))\phi(s)ds\}
\end{aligned}$$

where $\eta^{\mathrm{i}}_t = \mu + \boldsymbol{X}^{\mathrm{i}}_{\boldsymbol{R}}\boldsymbol{\alpha} + \sum_{k=1}^{K} \mathbb{1}\{A_t = \omega_k\} z_{\omega_k}\{\tau_{\omega_k} + \boldsymbol{X}^{\mathrm{i}}_{\boldsymbol{R}}\boldsymbol{\beta}_{\boldsymbol{\omega_k}}\} + s$.

Given that the $\boldsymbol{C}^{\mathrm{i}}$'s are missing, we use the EM algorithm to find the MLEs by iteratively taking the expectation of the complete-data log-likelihood using the conditional distribution of the missing $\boldsymbol{C}^{\mathrm{i}}$'s (E-step), and maximizing the expected log-likelihood (M-step). Suppose $\boldsymbol{\theta}^{(t)}$ is the current set of parameters. In the E-step, the expected value of the complete log-likelihood function is:

$$\begin{aligned}
Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}) &= E_{\boldsymbol{C}|\boldsymbol{\theta}^{(t)}}\{\log L(\boldsymbol{\theta}|\boldsymbol{Y},\boldsymbol{A},\boldsymbol{C},\boldsymbol{X})\} \\
&= \sum_{l=1}^{2^K} P(C^{\mathrm{i}}_l = 1|\boldsymbol{Y}^{\mathrm{i}},\boldsymbol{A}^{\mathrm{i}},\boldsymbol{X}^{\mathrm{i}},\boldsymbol{\theta}^{(t)})\{\sum_{\mathrm{i}=1}^{N} log\{P(\boldsymbol{Y}^{\mathrm{i}},\boldsymbol{A}^{\mathrm{i}},\boldsymbol{C}^{\mathrm{i}},\boldsymbol{X}^{\mathrm{i}}|\boldsymbol{\theta})\}\} \\
&= \sum_{l=1}^{2^K} P(C^{\mathrm{i}}_l = 1|\boldsymbol{Y}^{\mathrm{i}},\boldsymbol{A}^{\mathrm{i}},\boldsymbol{X}^{\mathrm{i}},\boldsymbol{\theta}^{(t)})\{\sum_{\mathrm{i}=1}^{N} \log P(C^{\mathrm{i}}_l = 1|\boldsymbol{X}^{\mathrm{i}}_{\boldsymbol{Z}},\boldsymbol{\theta}) \\
&\quad + \sum_{\mathrm{i}=1}^{N} \log P(\boldsymbol{Y}^{\mathrm{i}},\boldsymbol{A}^{\mathrm{i}}|\boldsymbol{C}^{\mathrm{i}},\boldsymbol{X}^{\mathrm{i}}_{\boldsymbol{R}},\boldsymbol{\theta})\} \\
&= \sum_{l=1}^{2^K} P(C^{\mathrm{i}}_l = 1|\boldsymbol{Y}^{\mathrm{i}},\boldsymbol{A}^{\mathrm{i}},\boldsymbol{X}^{\mathrm{i}},\boldsymbol{\theta}^{(t)})\{\sum_{\mathrm{i}=1}^{N} \log P(C^{\mathrm{i}}_l = 1|\boldsymbol{X}^{\mathrm{i}}_{\boldsymbol{Z}},\boldsymbol{\theta}) \\
&\quad + \sum_{\mathrm{i}=1}^{N} \log \int \prod_{t=1}^{T_{\mathrm{i}}} \exp(y^{\mathrm{i}}_t \eta^{\mathrm{i}}_t - \log(1+\exp(\eta^{\mathrm{i}}_t)))ds\}
\end{aligned}$$

Here the conditional probability $P(C^{\mathrm{i}}_l = 1|\boldsymbol{Y}^{\mathrm{i}},\boldsymbol{A}^{\mathrm{i}},\boldsymbol{X}^{\mathrm{i}})$ given $\boldsymbol{\theta}^{(t)}$ can be calculated as:

$$P(C^{\mathrm{i}}_l = 1|\boldsymbol{Y}^{\mathrm{i}},\boldsymbol{A}^{\mathrm{i}},\boldsymbol{X}^{\mathrm{i}},\boldsymbol{\theta}^{(t)}) = \frac{\boldsymbol{\pi}^{(t)}_l \int \prod_{t=1}^{T_{\mathrm{i}}} \exp(y^{\mathrm{i}}_t (\eta^{\mathrm{i}}_t)^{(t)} - \log(1+(\eta^{\mathrm{i}}_t)^{(t)}))ds}{\sum_{l=1}^{2^K} \boldsymbol{\pi}^{(t)}_l \int \prod_{t=1}^{T_{\mathrm{i}}} \exp(y^{\mathrm{i}}_t (\eta^{\mathrm{i}}_t)^{(t)} - \log(1+(\eta^{\mathrm{i}}_t)^{(t)}))\phi(s)ds} \quad (3.22)$$

where

$$(\eta_t^{\mathrm{i}})^{(t)} = \mu^{(t)} + \boldsymbol{X}_{\boldsymbol{R}}^{\mathrm{i}}\boldsymbol{\alpha}^{(\boldsymbol{t})} + \sum_{k=1}^{K} \mathbb{1}\{A_t = \omega_k\} z_{\omega_k}\{\tau_{\omega_k}^{(t)} + \boldsymbol{X}_{\boldsymbol{R}}^{\mathrm{i}}\boldsymbol{\beta}_{\boldsymbol{\omega_k}}^{(\boldsymbol{t})}\} + s$$

and

$$\boldsymbol{\pi}_l^{(t)} = P(C_l^{\mathrm{i}} = 1 | \boldsymbol{X}_{\boldsymbol{Z}}^{\mathrm{i}}, \boldsymbol{\theta}^{(\boldsymbol{t})}).$$

We can see that the Q function consists of two additive parts. The first is:

$$Q_1(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}) = \sum_{l=1}^{2^K}\{P(C_l^{\mathrm{i}} = 1|\boldsymbol{Y}^{\mathrm{i}}, \boldsymbol{A}^{\mathrm{i}}, \boldsymbol{X}^{\mathrm{i}}, \boldsymbol{\theta}^{(t)}) \sum_{\mathrm{i}=1}^{N} \log P(C_l^{\mathrm{i}} = 1|\boldsymbol{X}_{\boldsymbol{Z}}^{\mathrm{i}}, \boldsymbol{\theta})\} \tag{3.23}$$

and the second is:

$$\begin{aligned} &Q_2(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}) \\ &= \sum_{l=1}^{2^K}\{P(C_l^{\mathrm{i}} = 1|\boldsymbol{Y}^{\mathrm{i}}, \boldsymbol{A}^{\mathrm{i}}, \boldsymbol{X}^{\mathrm{i}}, \boldsymbol{\theta}^{(\boldsymbol{t})}) \sum_{\mathrm{i}=1}^{N} \log \int \{\prod_{t=1}^{T_{\mathrm{i}}} \exp(y_t^{\mathrm{i}}\eta_t^{\mathrm{i}} - \log(1+\eta_t^{\mathrm{i}}))\}\phi(s)ds\} \end{aligned} \tag{3.24}$$

$Q_1(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})$ contains the subgroup parameters $\boldsymbol{\gamma}$ only, while $Q_2(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})$ contains the response parameters $\boldsymbol{\beta}$, $\boldsymbol{\tau}$, and σ^2. Consequently, in the M-step, the maximization of $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})$ over $\boldsymbol{\theta}$ can be separated into maximizing $Q_1(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})$ over $\boldsymbol{\gamma}$ and maximizing $Q_2(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})$ over ($\boldsymbol{\beta}$, $\boldsymbol{\tau}$, σ^2). Since $Q_1(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})$ has a closed form, we can derive the first gradient function and use the Broyden-Fletcher-Goldfarb-Shanno (BFGS)[80]–[83] algorithm for maximization.

On the other hand, $Q_2(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})$ contains the integration of a Gaussian distribution. Similar to the approach in Section 2.2.2, we adopt Gauss-Hermite quadrature to approximate the integral and use numerical methods to calculate the gradient. Then we maximize the approximation of $Q_2(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})$ by the L-BFGS-B algorithm, restricting $\sigma^2 > 0$. The gradient of $Q_2(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})$ is calculated through numerical methods.

The EM algorithm is:

1. Choose starting values $\boldsymbol{\theta}^{(0)} = (\boldsymbol{\gamma}^{(0)}, \boldsymbol{\mu}^{(0)}, \boldsymbol{\alpha}^{(0)}, \boldsymbol{\tau}^{(0)}, \boldsymbol{\beta}^{(0)}, (\sigma^2)^{(0)})$. Set t=0

2. E-step: calculate conditional probability $P(C_l^{\mathrm{i}} = 1|Y_{\mathrm{i}}, A_{\mathrm{i}}, \boldsymbol{X}_{\boldsymbol{R}}^{\mathrm{i}}, \boldsymbol{X}_{\boldsymbol{Z}}^{\mathrm{i}})$ according to $\boldsymbol{\theta}^{(t)}$. Construct $Q_1(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})$ and $Q_2(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})$.

3. M-step: maximize $Q_1(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})$ over $\boldsymbol{\gamma}$ and $Q_2(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})$ over $(\boldsymbol{\mu}, \boldsymbol{\alpha}, \boldsymbol{\tau}, \boldsymbol{\beta}, \sigma^2)$ respectively. Update $\boldsymbol{\theta}$. Let t=t+1.

4. Repeat until convergence.

Our stopping criteria is $|\boldsymbol{\theta}^{(t+1)} - \boldsymbol{\theta}^{(t)}| < 10^{-4}$ and $|L(\boldsymbol{\theta}^{(t+1)}) - L(\boldsymbol{\theta}^{(t)})| < 10^{-8}$. We run the algorithm multiple times, starting from different sets of initial values, to search for the global maximum likelihood. Compared with Chapter 2, there are more parameters for estimation. For better convergence, each set of starting values is chosen carefully. In the following simulation studies, the starting values are randomly generated within a small range of the true values.

3.3 Inclusion of Time Effects

So far in this chapter, we've assumed that for each individual, the response rates to the same treatment are the same over the stages. However, in practice, the body and disease may adjust to a treatment or treatments, thereby reducing the likelihood of a response to a treatment in later stages. Therefore, it is important to consider incorporating time effects to accommodate the potential changes in the response rate over stages. In this section, we expand the response component of our model to incorporate time effects. The modification in the EM algorithm is also provided. Without loss of generality, we also assume that the time effect is reflected as a decline in the response rate.

There are many different approaches one can take to describe a decrease in the response rates over time. The one we consider in this dissertation is that the body and disease adjust to protect themselves from any treatment and so there is a natural decline in effect over time. The body and disease adjustments may impact treatments differently, so we consider this decline to be treatment-specific. In terms of modeling the time effects, we consider this decline both numerically (i.e., linear decline on the logit scale) or treating time as a categorical factor.

3.3.1 The Time Effect Models

When treating time numerically, time enters the logit model as a numerical covariate. Because we allow it to be treatment-specific, there is a different rate parameter δ_{ω_k} for each treatment. As a result, the logit function can be written as:

$$\begin{aligned}\log(\frac{p_t}{1-p_t}) =&\mu + \boldsymbol{X_R}\boldsymbol{\alpha} + \sum_{k=1}^{K} \mathbb{1}\{d_t(\boldsymbol{X_R}, a_1, y_1, ..., a_{t-1}, y_{t-1}) = \omega_k\} z_{\omega_k}\{\tau_{\omega_k} + \boldsymbol{X_R}\boldsymbol{\beta_{\omega_k}}\} \\ &+ (t-1)\delta_{\omega_k} + s \end{aligned} \tag{3.25}$$

where $t = 1, 2, ..., T$. We expect each of the δ_{ω_k} to be negative, implying that the log odds of responding favorably in the current stage would reduce by δ_{ω_k} each stage of the study.

When treating time categorically, there is now a different decline for each treatment and stage. For stages $t > 1$, the logit function now includes a decline parameter $\delta_{\omega_k,t}$ and can be expressed as:

$$\mu + \boldsymbol{X_R}\alpha + \sum_{k=1}^{K} \mathbb{1}\{d_t(\boldsymbol{X_R}, a_1, y_1, ..., a_{t-1}, y_{t-1}) = \omega_k\} z_{\omega_k}\{\tau_{\omega_k} + \boldsymbol{X_R}\boldsymbol{\beta_{\omega_k}}\} + \delta_{\omega_k,t} + s, \tag{3.26}$$

where $t = 2, ..., T$ and $\delta_{\omega_k,t}$ is the time effect at time t for treatment ω_k. In contrast to treating time numerically, this approach allows for non-linear changes over time. It can accommodate scenarios when the body dramatically shuts down to the treatments but at the expense of including $(t-2)K$ more model parameters.

3.3.2 Estimation for Time Effects

Just as in Section 3.2, the EM algorithm is applied for parameter estimation. We consider categorical time effects for illustrative purposes. Let $\boldsymbol{\delta} = (\delta_{\omega_k,2}, ..., \delta_{\omega_k,T})$ be the time effect parameters. In E-step, $P(C_l^{\mathrm{i}} = 1|\boldsymbol{Y}^{\mathrm{i}}, \boldsymbol{A}^{\mathrm{i}}, \boldsymbol{X}^{\mathrm{i}}, \boldsymbol{\theta}^{(t)})$ is calculated via Equation (3.22), where we use Equation (3.26) for $(\eta_t^{\mathrm{i}})^{(t)}$. In the M-step, $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})$ is maximized by maximizing $Q_1(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})$ over $\boldsymbol{\gamma}$ and $Q_2(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})$ over $(\boldsymbol{\beta}, \boldsymbol{\tau}, \boldsymbol{\delta}, \sigma^2)$ separately. Thus, we can update subgroup parameter $\boldsymbol{\gamma}$ in the same way. As for $Q_2(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})$, parameters are updated together using the L-BFGS-B algorithm as before.

3.3.3 Hypothesis Testing for Time Effects

Letting $\boldsymbol{\delta}$ be the vector of time effects. The null hypothesis and the alternative hypothesis are:

$$H_0 : \boldsymbol{\delta} = \mathbf{0}$$

$$H_a : \text{not all } \delta \text{ are 0's}$$

We consider both Wald's test and the likelihood ratio versions of the test. For Wald's test, we first find a $Q \times P$ matrix R such that $R\boldsymbol{\theta}' = \boldsymbol{\delta}'$. Then the null hypothesis can also be written as $R\boldsymbol{\theta}' = \mathbf{0}'$. Let $\hat{\boldsymbol{\theta}}$ be the MLE under the alternative hypothesis and $\hat{V}$ be the asymptotic covariance matrix, which can be approximated by the second partial derivatives of $L(\hat{\boldsymbol{\theta}})$ (i.e., Hessian matrix). There are several built-in functions in R we can use to calculate it through numerical methods. One can also consider bootstrapped standard errors. However, the bootstrap method requires that the EM algorithm is applied numerous times [84], and therefore, is very time-consuming. Once we get the covariance matrix $\hat{V}$, the test statistic $W = \hat{\boldsymbol{\delta}}(R'\hat{V}R/N)^{-1}\hat{\boldsymbol{\delta}}' \rightarrow \chi^2$. The degree of freedom depends on the number of time effect parameters.

For the likelihood ratio test, we must estimate the MLE under both H_0 ($\boldsymbol{\delta} = 0$) and H_a ($\boldsymbol{\delta} \neq 0$). Similar to Wald's test, we first obtain $\hat{\boldsymbol{\theta}}$ under H_a through the EM algorithm. Next, we assume that there is no time effects ($\boldsymbol{\delta} = 0$) and apply the EM algorithm again to get $\hat{\boldsymbol{\theta}}^* = (\hat{\mu}^*, \hat{\boldsymbol{\gamma}}^*, \hat{\alpha}^*, \hat{\boldsymbol{\tau}}^*, \hat{\boldsymbol{\beta}}^*, (\hat{\sigma^2})^*)$. The test statistic is $\lambda_{LR} = -2(L(\hat{\boldsymbol{\theta}}^*) - L(\hat{\boldsymbol{\theta}}))$ and also asymptotically follows the χ^2 distribution. Compared with Wald's test, the likelihood ratio test involves using the EM algorithm to fit the data twice, while Wald's test only needs to use the EM algorithm once.

3.4 Simulation Studies

Simulation studies are conducted to evaluate the performance of our model-based approach. Data are simulated under the same MD Anderson design that was used in Chapter 2 (see Section 2.3.2 for a description). We consider that there are two baseline covariates.

One is a response covariate X_R and the other is a subgroup covariate X_Z. They each follow standard Normal distributions $N(0,1)$. There are four possible treatments so the subgroup vector as $\boldsymbol{Z}$ is of length four.

This section consists of three subsections. In the first two, we perform simulation studies under the independence association structure (Section 3.4.1) and the homogeneous association structure (Section 3.4.2) assuming no time effects. A summary of these two multivariate Bernoulli subgroup models can be found in Table 3.1 in Section 3.2.3. Our response model for both subsections is expressed by Equation (3.17) in Section 3.1.3.

In both subsections, we consider sample sizes of 200, 600, 1200, and 2000 subjects. For each sample size, 200 datasets are generated from the true model and then fitted by our mixture model. The accuracy of our model estimates is summarized using means and standard deviations. We also investigate the probabilities of a favorable response and the probabilities of being in the beneficial subgroups given the parameter estimates, since these probabilities are later used to determine the optimal DTRs. Lastly, we compare the estimated optimal DTRs selected by our model and Q-learning in these aspects:

1. The mean values of estimated optimal DTRs over the entire population.

2. The probabilities of selecting the true optimal DTR over the entire population.

3. Pairwise comparisons of the estimated optimal DTRs between two methods within the same dataset.

In the last subsection, we conduct simulation studies considering both numerical and categorical time effects. We restrict our attention to only a sample size of $N = 1200$. As before, 200 datasets are generated for each model and we fit each of these datasets using a model with and without time effects. We compare these two fits to quantify how much bias will incur when time effects are ignored. Then, we compare our model with Q-learning in terms of finding the optimal DTR. Hypothesis testing is also conducted to detect the existence of time effects.

3.4.1 Independence Association Structure

Given one response covariate and one subgroup covariate, there are 9 parameters in the treatment response component of the model, and 8 parameters in the subgroup component of our model. The values used to simulate the data are listed in the second column of Table 3.5.

Table 3.5. Means and standard deviations of parameter estimates for 4-treatment independence model when N = 2000

Parameters	True value	Mean	Std. Dev.
$\gamma_{1,0}$	0.30	0.3247	0.1691
$\gamma_{1,1}$	1.50	1.5469	0.2109
$\gamma_{2,0}$	1.00	1.0848	0.2986
$\gamma_{2,1}$	-3.00	-3.1207	0.4734
$\gamma_{3,0}$	-1.00	-0.9920	0.1171
$\gamma_{3,1}$	1.00	1.0081	0.1252
$\gamma_{4,0}$	0.50	0.5308	0.1732
$\gamma_{4,1}$	-1.00	-1.0424	0.1610
τ_1	0.80	0.8035	0.1218
β_1	-0.50	-0.5198	0.0979
τ_2	0.30	0.2938	0.0893
β_2	0.60	0.6299	0.0876
τ_3	1.10	1.1070	0.1629
β_3	0.40	0.4118	0.1377
τ_4	0.60	0.6023	0.1155
β_4	1.00	1.0178	0.1182
σ	0.1	0.2014	0.2484

Also included in Table 3.5 are the means and the standard deviations of the 200 estimates when sample size $N = 2000$. For all subgroup and response parameters, the means are close to true values with fairly small standard deviations. The means and the standard deviations for the other sample sizes are summarized in the Appendix B (see Table B.1.). As the sample size increases, the precision increases. We observe large variation when $N = 200$. When N is 600 or greater, the estimations are relatively decent and variation becomes smaller.

In addition to parameter estimates, we also looked at the probabilities and examined how the marginal subgroup probability changes given different X_Z, as well as how the response

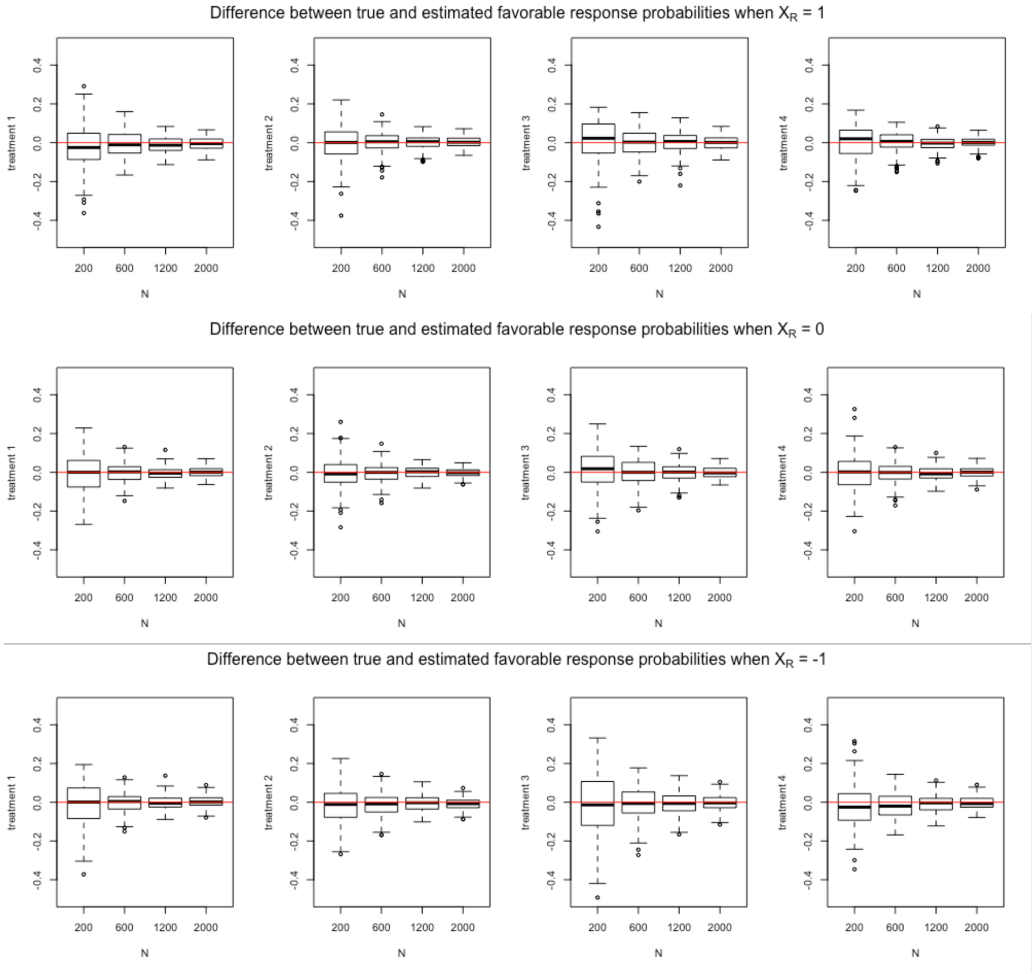

Figure 3.2. Boxplots of differences between the true and the estimated subgroup probabilities given $X_Z = -1, 0, 1$ for independence model. The red line is difference = 0.

probability on its favorable treatment subgroup changes given X_R. These probabilities are computed based on parameter estimates and used later for computing the value of the DTR. They have a more direct impact on the estimated value of DTR.

For each set of parameter estimates, we calculate the subgroup probabilities $P(Z_k = 1|X_Z)$ and the response probabilities $P(Y = 1|X_R, A_t = k, Z_k = 1, s = 0)$ given $X_Z = -1, 0, 1$ and $X_R = -1, 0, 1$, for $k = 1, ..., 4$. Then we compare the estimated probabilities with the true probabilities. Figures 3.2 and 3.3 show the boxplots of the difference between the estimated and the true probabilities (estimated - true) of each subgroup probabilities and the response probabilities on each treatment, respectively, under different sample sizes. When

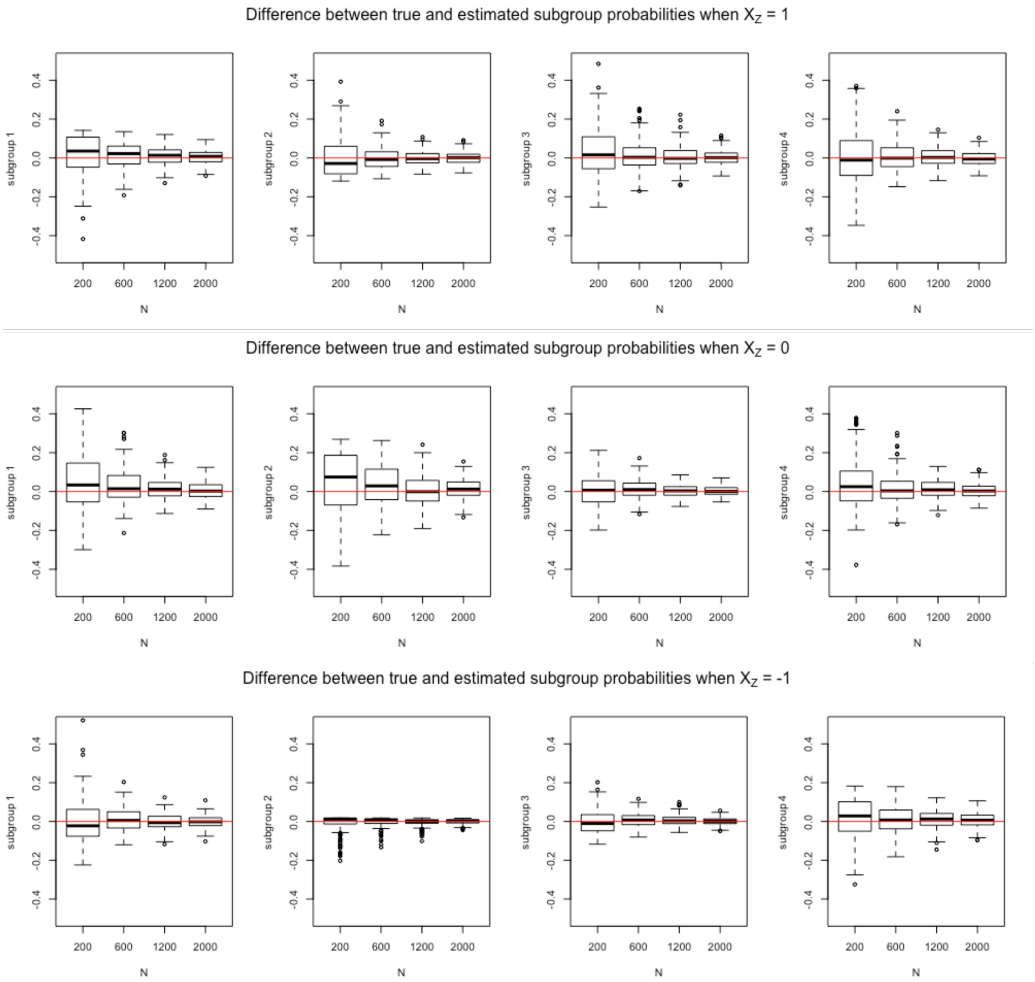

Figure 3.3. Boxplots of differences between the true and the estimated probabilities of a favorable response the treatment on its favorable subgroup given $X_R = -1, 0, 1$ for independence model. The red line is difference $= 0$.

the sample size is small ($N = 200$ and $N = 600$), the range of these probabilities is relatively large. We also notice that some medians of these probabilities are slightly shifted away from the true value when $N = 200$. This could be due to the lack of observations on all possible trajectories. As the sample size increase, bias decreases and precision increases.

Comparison with Q-learning

Given the baseline covariates $\boldsymbol{x}$, our approach selects the estimated optimal DTR $D_{\boldsymbol{x}}^{opt}$ through an exhaustive search, comparing the values of all possible DTRs. As for Q-learning,

similar to Chapter 2 (see Section 2.3.2), linear regression models are adopted to fit the Q-functions stagewise. We now include both baseline covariates X_R and X_Z as the predictors into the model, as well as the interaction terms between covariates and treatments. For patients who receive two treatments, we model the Q-function $Q_2(\boldsymbol{x}, a_1, y_1, a_2)$ first. Based on the parameter estimates, we find the optimal decision rule $d_2^*(\boldsymbol{x}, a_1, y_1)$ by comparing the estimated Q-functions over all possible treatments. We then use all patients' data to fit the linear model for $Q_1(\boldsymbol{x}, a_1)$ and determine the optimal rule $d_1^*(\boldsymbol{x})$. The optimal DTR is $d^* = (d_1^*, d_2^*)$.

The above estimated optimal DTRs selected by both methods are at the individual level, i.e., based on a specific $\boldsymbol{x}$. Our interest is in our model's capability of selecting the DTR with a higher value, if followed by the overall population. Therefore, rather than checking the value of estimated optimal DTR for a given $\boldsymbol{x}$, we focus on the true average value of the DTR for the entire population. For the purpose of comparison, suppose that the distribution of covariates and true parameter values are known. We can then average over the covariate distribution. Letting f be the density function of the baseline covariates, we compute the value of D as follows:

$$V(D) = \int V(D|\boldsymbol{x}) f(\boldsymbol{x}) d\boldsymbol{x}.$$

For these simulations, this value is computed using the true parameter values.

It is intuitive to consider Monte Carlo methods to approximate the population value of the estimated optimal DTRs. One can generate a large set of $\boldsymbol{X}$ from its density function f and compute the average of $V(D|\boldsymbol{x})$. Since we assume that X_Z and X_R follow independent standard Normal distributions, we can use the Gauss-Hermite quadrature to estimate the value for both approaches. For each replicate, we first generate a set of 400 two-dimensional Gauss quadrature nodes as $\boldsymbol{X} = (X_Z, X_R)$ and obtain their corresponding weights. Then for each $\boldsymbol{x} = (x_Z, x_R)$ and DTR D, we compute its value $V(D|\boldsymbol{x})$ based on the true parameters. Finally, by calculating the weighted sum of $V(D|\boldsymbol{x})$ over $\boldsymbol{x}$, we obtain the expected value.

We summarize the results for both our mixture model and Q-learning using means and standard deviations of the estimated optimal DTR values (Table 3.6). The true optimal DTR value is 0.5552. As the sample size increases, for both methods, the value of the estimated

optimal DTR increases and approaches the true optimal DTR value. After a sample size of 600, the values of estimated DTRs in both methods are close to the true optimal DTR value. This implies that the true optimal DTR has been selected in most cases. For all sample sizes, the mean and standard deviation support our approach over Q-learning, especially when the sample size is very small. This is most likely due to the fact that Q-learning determines the optimal decision rule based only on the information from each stage, while our model-based approach combines information across stages. Thus, given that our model is approximately correct, our approach has more power to infer the optimal DTR, especially for very small sample sizes.

Table 3.6. Means and standard deviations of estimated optimal DTR values for independence model

sample size	Mixture Model		Q-learning	
	Mean	Std. Dev.	Mean	Std. Dev.
N = 200	0.5451	0.0074	0.4934	0.0348
N = 600	0.5521	0.0026	0.5277	0.0182
N = 1200	0.5533	0.0015	0.5397	0.0099
N = 2000	0.5540	0.0011	0.5442	0.0077

Figure 3.4 shows the smoothed histograms of values for the estimated optimal DTR. As the sample size increases, for both approaches, the peak gets closer to the value of true optimal DTR but there is a much larger range of estimated optimal DTR values for Q-learning than our model-based approach.

Table 3.7. Means and standard deviations of probabilities of finding the optimal DTRs for independence model

sample size	Mixture Model		Q-learning	
	Mean	Std. Dev.	Mean	Std. Dev.
n = 200	74.74%	10.98%	40.87%	18.69%
n = 600	85.42%	8.38%	60.51%	16.28%
n = 1200	88.15%	7.29%	68.70%	12.78%
n = 2000	90.48%	7.13%	72.56%	11.77%

Another way to look at this is to consider the frequency of each approach finding the optimal DTR. These values are also estimated by numerical integration. For each replicate,

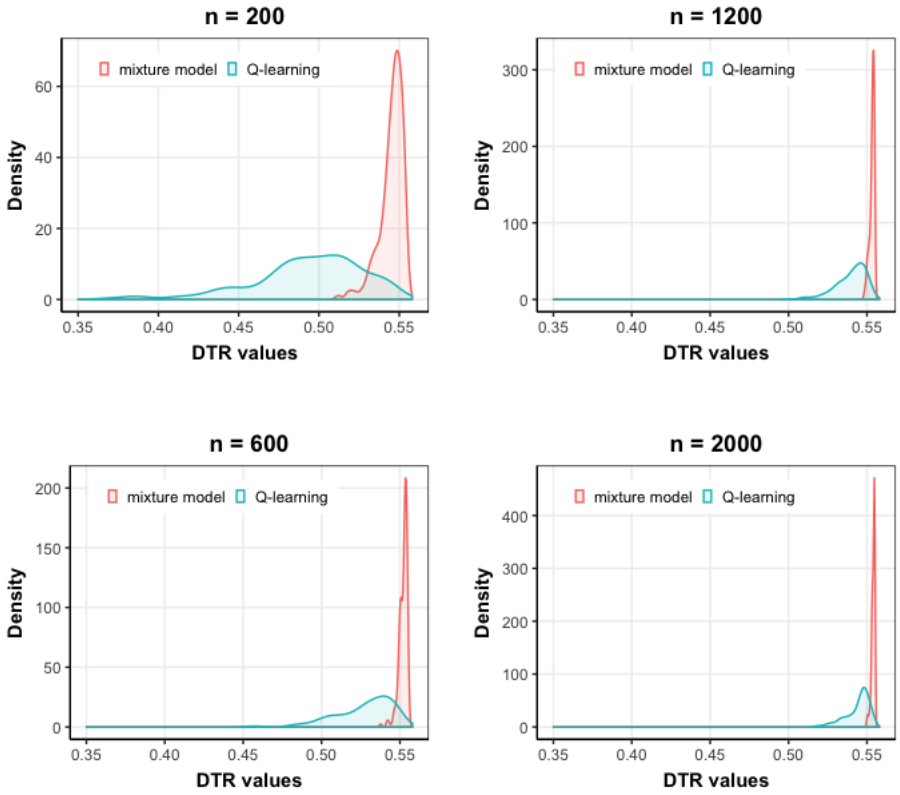

Figure 3.4. Smoothed histograms of values for estimated optimal DTR for independence model

we generate a set of 400 two-dimensional Gauss quadrature nodes as $\boldsymbol{X} = (X_Z, X_R)$ and corresponding weights. For a given $\boldsymbol{x} = (x_Z, x_R)$, we find the estimated optimal DTR D and check whether it is the true optimal DTR or not. We then sum up the weighted frequency of selecting the true optimal DTR to get the probability of selecting the true optimal DTR over the population. Table 3.7 summarizes the means and standard deviations of these proportions under the different sample sizes for both methods. The probability increases as the sample size increases. We observe a significant increase for both methods in the probability when N increases from 200 to 600. Our method has better chances to find the optimal DTR on the overall population under all sample sizes. In fact, the mean and standard deviation of the proportion in our model when $N = 200$ are close to those of Q-learning when $N = 2000$. This is consistent with what we see in Figure 3.4.

Table 3.8. Percentages of selecting DTRs with higher values for independence model

N	Percentage
200	99.0%
600	98.0%
1200	98.5%
2000	99.5%

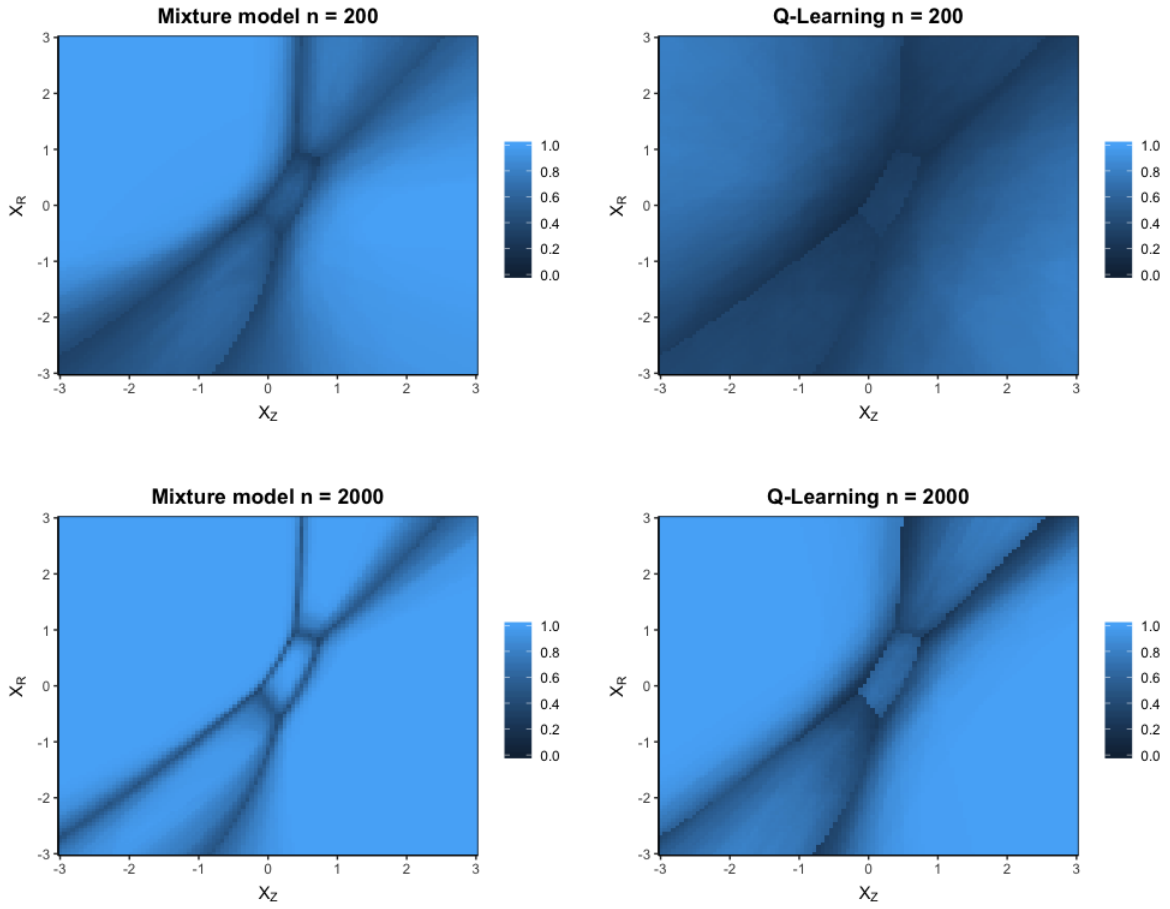

Figure 3.5. Heatmap of probabilities of selecting the true optimal DTR when N = 200 and 2000 by both mixture model and Q-learning for independence model

Finally, we compare the values of the estimated optimal DTRs for each replication and determine the proportion of times that our model outperforms Q learning. Table 3.8 summarizes the result. We see that the proposed model chooses a better DTR than Q-learning almost all the time, regardless of sample size.

This can also be seen visually in Figure 3.5. It shows a heatmap of these proportions of finding the true optimal DTR over different value of $\boldsymbol{x}$. The lighter the blue is, the higher chance to find the true optimal DTR. Both methods show significant improvement when the sample size increases from 200 to 2000 but it is also clear that our method provides substantial improvements.

Comparing these figures with Figure 3.1, we see that the deep blue area areas represent the areas near the boundaries where two different DTRs have the same value. This means that the value of the second best DTR is likely very close to the value of the optimal DTR. Consequently, selecting the second best DTR as the optimal DTR in these regions won't result in a significant drop in the estimated population value. Q-learning, even when $N = 2000$, has much wider ranges around these boundaries, explaining in general why Q-learning does more poorly.

3.4.2 Homogeneous Association Structure

We now expand the association structure to the more complicated homogeneous association, which includes 10 non-zero second-order natural parameters instead of just the four first-order ones. Consequently, there are now 29 parameters, 20 for the subgroup component to go along with the 9 for the response component.

Table 3.9. Means and standard deviations of response parameter estimates for 4-treatment homogeneous association model when N = 2000

Parameters	True value	Mean	Std. Dev.
τ_1	0.80	0.7844	0.1544
β_1	-0.50	-0.5328	0.1347
τ_2	0.30	0.2790	0.1857
β_2	0.60	0.6225	0.1507
τ_3	1.10	1.1106	0.1147
β_3	0.40	0.4049	0.0906
τ_4	0.60	0.6038	0.1179
β_4	1.00	1.0175	0.1176
σ	0.1	0.2168	0.2782

Tables 3.9 and 3.10 summarize the parameter settings used to generate the data and the resulting means and standard deviations of the 200 estimates when sample size $N = 2000$. The means of the parameter estimates are close to the true values with very small standard deviations. Compared with the means and standard deviations of the independent structure in Table 3.5, we observe larger standard deviations of subgroup parameters in homogeneous association structure.

Table 3.10. Means and standard deviations of subgroup parameter estimates for 4-treatment homogeneous association model when N = 2000

Parameters	True value	Mean	Std. Dev.
$\gamma_{1,0}$	-0.6000	-0.6282	1.6587
$\gamma_{1,1}$	1.5000	1.5794	1.5838
$\gamma_{2,0}$	-1.0000	-1.1420	1.7147
$\gamma_{2,1}$	1.3000	1.6053	1.3630
$\gamma_{3,0}$	0.9000	1.4504	1.2738
$\gamma_{3,1}$	1.3000	1.6440	1.7896
$\gamma_{4,0}$	0.6000	1.0589	1.2792
$\gamma_{4,1}$	-1.5000	-1.6717	1.7642
$\gamma_{12,0}$	0.5	0.7883	1.5340
$\gamma_{12,1}$	-1	-1.2829	1.9526
$\gamma_{13,0}$	0.2	0.1512	1.5343
$\gamma_{13,1}$	-1.2	-1.3751	1.1891
$\gamma_{14,0}$	0	-0.1352	1.4182
$\gamma_{14,1}$	0	0.3751	1.8368
$\gamma_{23,0}$	0	-0.2291	1.7623
$\gamma_{23,1}$	0	0.1033	1.9614
$\gamma_{24,0}$	-0.5	-0.6973	1.5568
$\gamma_{24,1}$	-1	-1.2954	1.3044
$\gamma_{34,0}$	-0.4	-0.8386	1.5516
$\gamma_{34,1}$	1.4	1.4541	1.8672

As in our other simulation studies, we also compute the subgroup probability and the response probability on its favorable treatment subgroup given different X_Z and X_R, respectively. For each set of parameter estimates, we calculate the marginal subgroup probabilities $P(Z_k = 1|X_Z)$ and the response probabilities $P(Y = 1|X_R, A_t = k, Z_k = 1, s = 0)$, for $k = 1, 2, 3, 4$ given $X_Z = -1, 0, 1$ and $X_R = -1, 0, 1$. Figures 3.6 and 3.7 show the boxplots of the differences between the true and the estimated subgroup probabilities, and the differ-

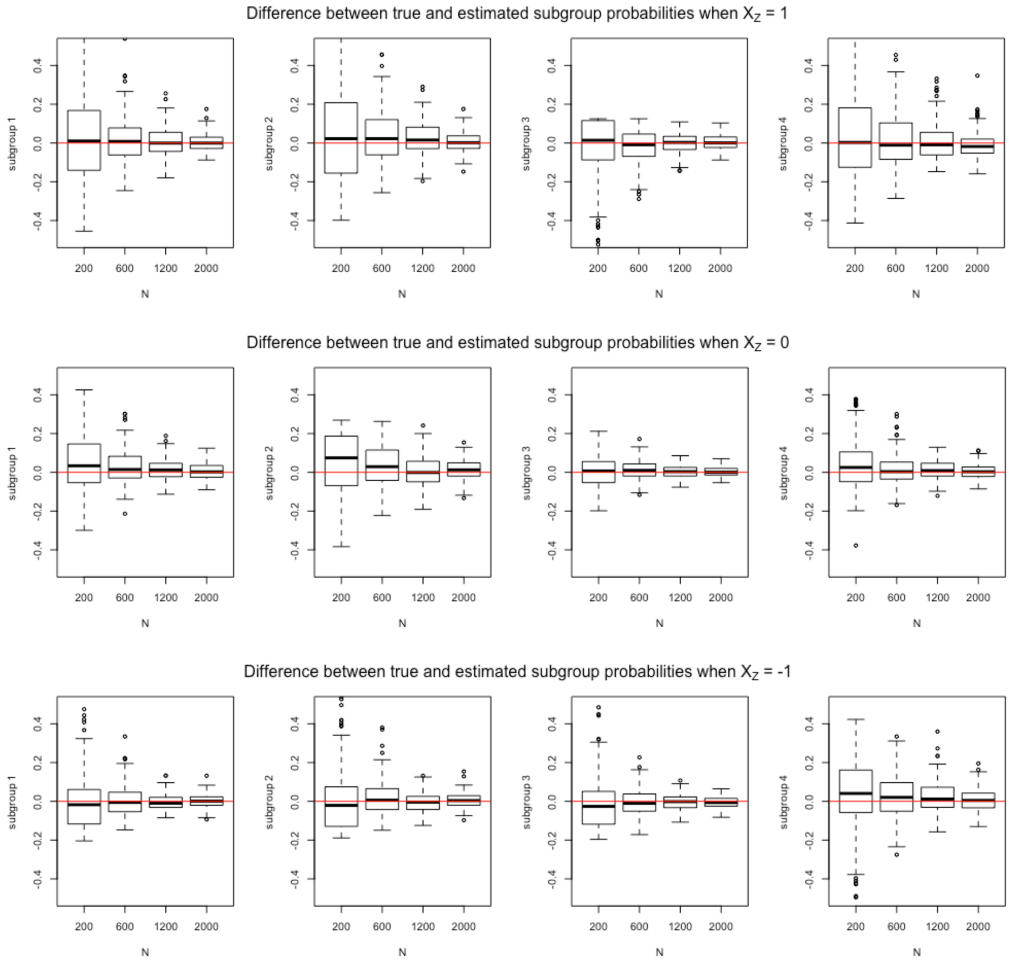

Figure 3.6. Boxplots of differences between the true and the estimated subgroup probabilities given $X_Z = -1, 0, 1$ for homogeneous association model. The red line is difference $= 0$.

ences between the true the estimated response probabilities on each treatment respectively. In both figures, as sample size increases, bias decreases and precision increases. We see a larger range and more outliers in Figure 3.6, which is the reflection of larger variation in subgroup parameter estimates, especially when the sample size is small. Similarly, when comparing with the boxplots subgroup probabilities under the independent model structure (Figure 3.2), we see larger range in Figure 3.6. This is likely due to the more complex subgroup structure under the homogeneous association structure.

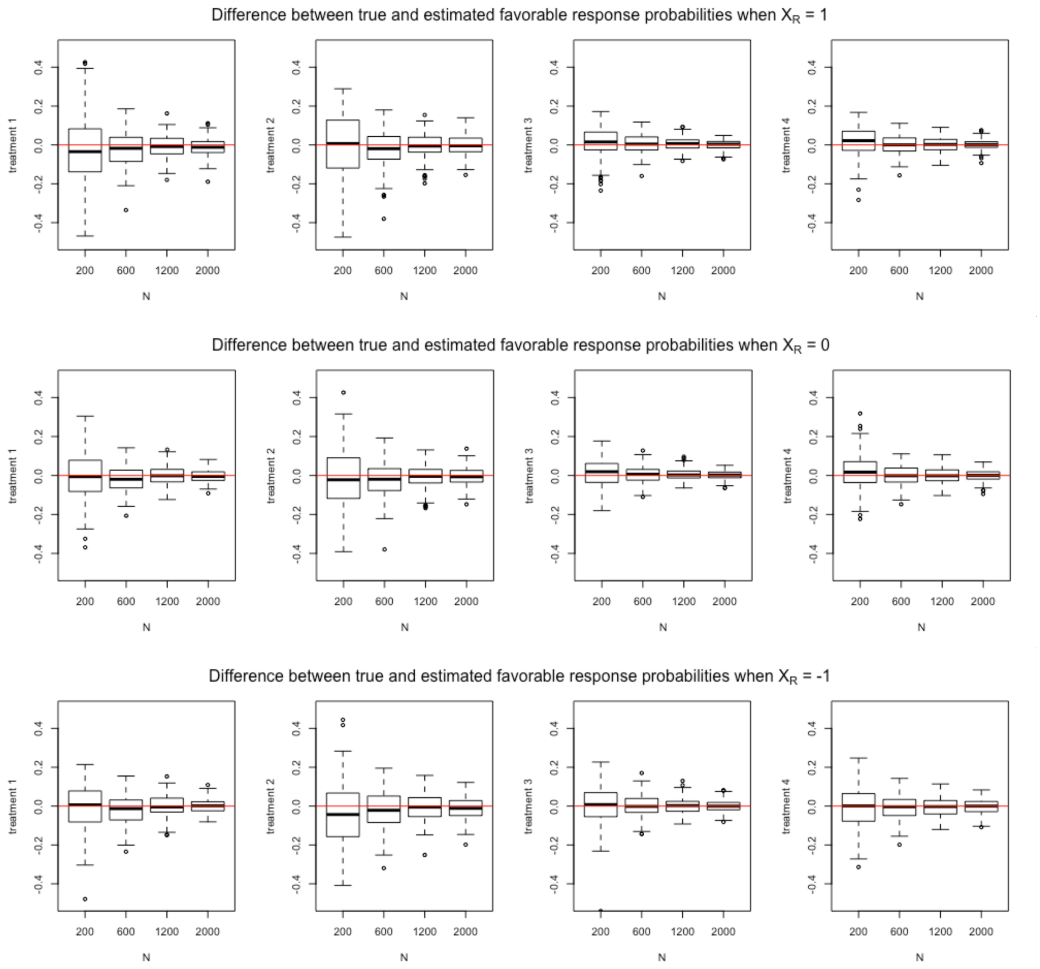

Figure 3.7. Boxplots of differences between the true and the estimated probabilities of a favorable response the treatment on its favorable subgroup given $X_R = -1, 0, 1$ for homogeneous association model. The red line is difference $= 0$.

Comparison with Q-learning

Like we did in the independence association structure, we compare our proposed approach with Q-learning. Table 3.11 summarizes the means and standard deviations of the estimated optimal DTR values. The true optimal DTR value is 0.5397. For both methods, the value increases with the sample size. A substantial jump occurs when the sample size increases from 200 to 600. For a large sample size, both are close to the true optimal DTR value. Figure 3.8 shows the smoothed histograms of values of estimated optimal DTRs. Since

our approach utilizes information across stages, our approach can select DTRs with higher values and outperforms Q-learning under different sample sizes. Furthermore, we find that the performance of our model when $N = 600$ is similar to Q-learning when $N = 2000$.

Table 3.11. Means and standard deviations of estimated optimal DTR value for homogeneous association model

sample size	Mixture Model		Q-learning	
	Mean	Std. Dev.	Mean	Std. Dev.
n = 200	0.5198	0.0142	0.4988	0.0293
n = 600	0.5343	0.0046	0.5216	0.0151
n = 1200	0.5375	0.0017	0.5306	0.0062
n = 2000	0.5386	0.0009	0.5320	0.0059

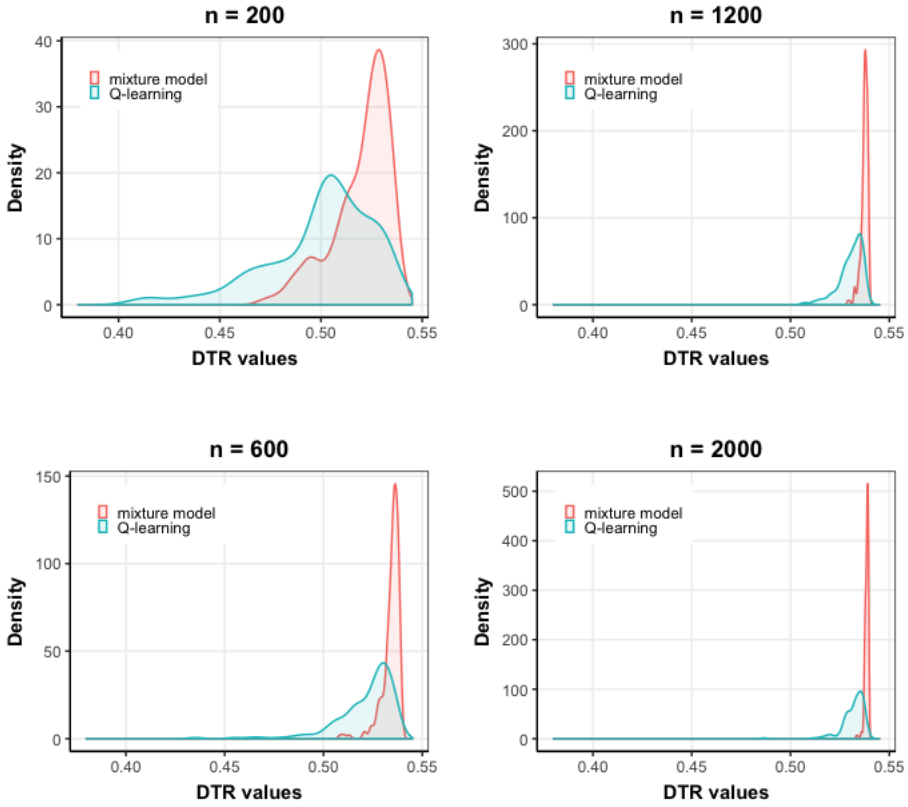

Figure 3.8. Smoothed histograms of values for estimated optimal DTR for homogeneous association scenario

Table 3.12. Means and standard deviations of probabilities of finding the optimal DTRs for homogeneous association model

sample size	Mixture Model		Q-learning	
	Mean	Std. Dev.	Mean	Std. Dev.
n = 200	66.00%	11.28%	49.14%	17.38%
n = 600	79.78%	7.75%	65.61%	13.44%
n = 1200	86.25%	6.46%	74.30%	8.05%
n = 2000	90.25%	5.47%	76.71%	7.65%

We estimated the proportion of times that we find the optimal DTR. The results are summarized in Table 3.12. The chance of selecting the true optimal DTR by our model when $N = 600$ is similar to Q-learning when $N = 2000$.

Table 3.13. Percentages of selecting DTRs with higher values for homogeneous association model

N	Percentage
200	77.5%
600	88.5%
1200	95.0%
2000	96.5%

In each replication, we compare the values of these estimated optimal DTRs for both approaches and find the percentage when the proposed model selects DTRs with a higher value than Q-learning. Table 3.13 presents the results. In this case, there is a much large difference between the two methods. It suggests that Q-learning may be struggling with the correlated subgroup structure of the data.

We further investigate the heatmap of probabilities of finding the true optimal DTR given specific baseline covariate $\boldsymbol{x} = (x_Z, x_R)$, shown in Figure 3.9. The boundaries where the true optimal DTR changes from one to another are clearly present under both methods. For our method, there is an obvious improvement when the sample size increases from 200 to 2000. On the contrary, for Q-learning, significant improvement happens in the upper center, but not in the upper left and upper right. It even gets worse in some areas around the boundaries. It's likely the result of the more complicated subgroup structure. Q-learning cannot capture

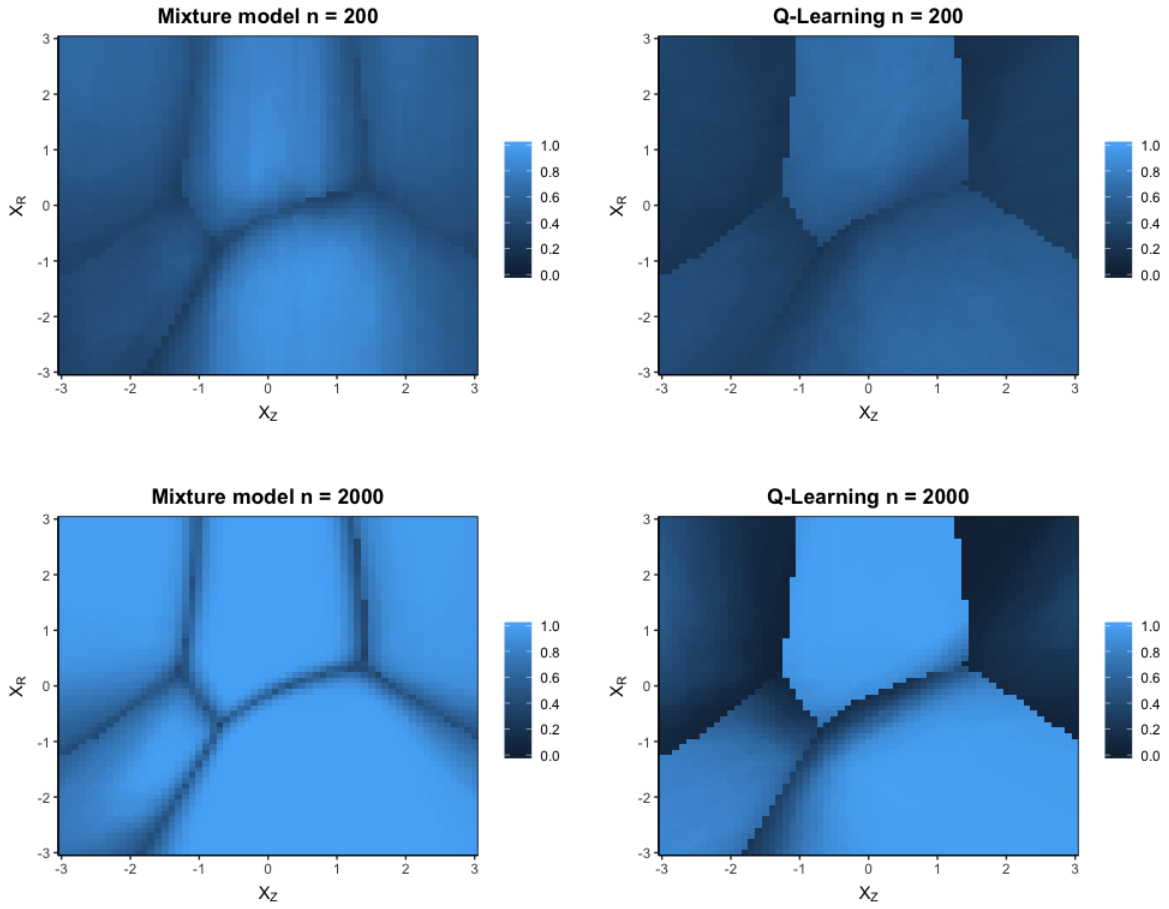

Figure 3.9. Heatmap of probabilities of selecting the true optimal DTR when N = 200 and 2000 by both mixture model and Q-learning for homogeneous association model

the underlying relationships among subgroups while our approach takes account of them. This result is consistent with Table 3.13.

3.4.3 Time Effects

For both numerical and categorical time effects models, the independence association structure parameters that were used for the subgroup model component as Section 3.4.1. As for the response model, since the inclusion of time effects increases the number of parameters, we assume no interactions between the treatments and X_R. For both types of time effects, simulation studies are conducted with sample size $N = 1200$ only.

Numerical Time Effects

Under the assumptions of zero natural response rate, the probability of a successful response $p_t = P(Y_t = 1|X_R, a_t, \boldsymbol{z}, s)$ with numerical time effects in Equation (3.25) now can be expressed as:

$$\log(\frac{p_t}{1-p_t}) = \tau_{a_t} + x_R\beta + \delta_{a_t}(t-1) + s, \qquad \text{if } z_{a_t} = 1$$
$$p_t = 0, \qquad \text{if } z_{a_t} = 0$$

where $t = 1, 2, 3, 4$. There are 10 parameters in the logistic regression part: $(\tau_1, \tau_2, \tau_3, \tau_4, \beta, \delta_1, \delta_2, \delta_3, \delta_4, \sigma^2)$. The true values are:

Parameters	τ_1	τ_2	τ_3	τ_4	β	δ_1	δ_2	δ_3	δ_4	σ^2
True value	0.8	0.3	1.1	0.6	-0.5	-0.3	-0.1	-0.15	-0.2	0.01

According to this parameter setting, Treatment 1's effect diminishes over time the most. Treatment 4 ranks second, followed by Treatment 3. Treatment 2 is the most resilient over time.

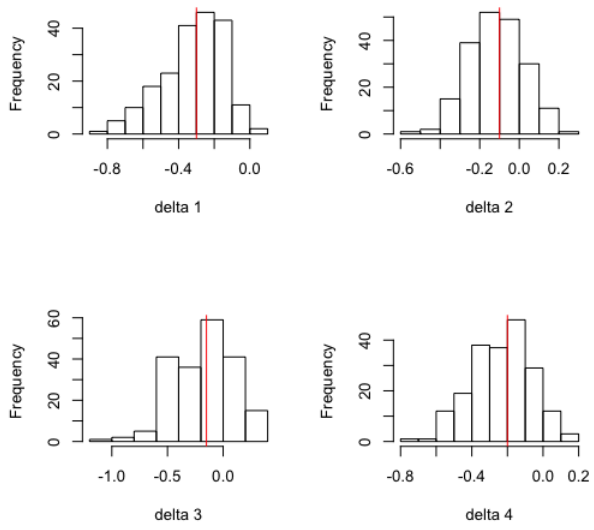

Figure 3.10. Histograms of estimated numerical time effects. The red line represents the true value.

Figure 3.10 shows the histogram of the four treatment-specific numerical time effects, where the red lines indicate the true values. As we can see, the estimates are fairly close to the true values. Thus, our model is capable of estimating the time effects adequately.

We then fit the data using the model without incorporating time effects as in Section 3.4.1 ($\boldsymbol{\delta} = \mathbf{0}$), and compare treatment parameter estimates between models with and without time effects. Figure 3.11 presents the histograms of the estimates of each treatment effect, where the red lines indicate true parameters. Obvious biases among estimates τ_1, τ_2, τ_3, and τ_4 are observed from the model without time effects. Particularly, the differences in treatment effect estimates between the two models increase when the time effects of the treatments intensify. For example, the average difference of estimated treatment effect τ_1 between models with time effects and without time effects is 0.4032, but only 0.1205 in τ_2, where time effects are -0.3 and -0.1 respectively. This is because when the time effects are missed, the model underestimates treatment effects to compensate for the lower success rates in later stages.

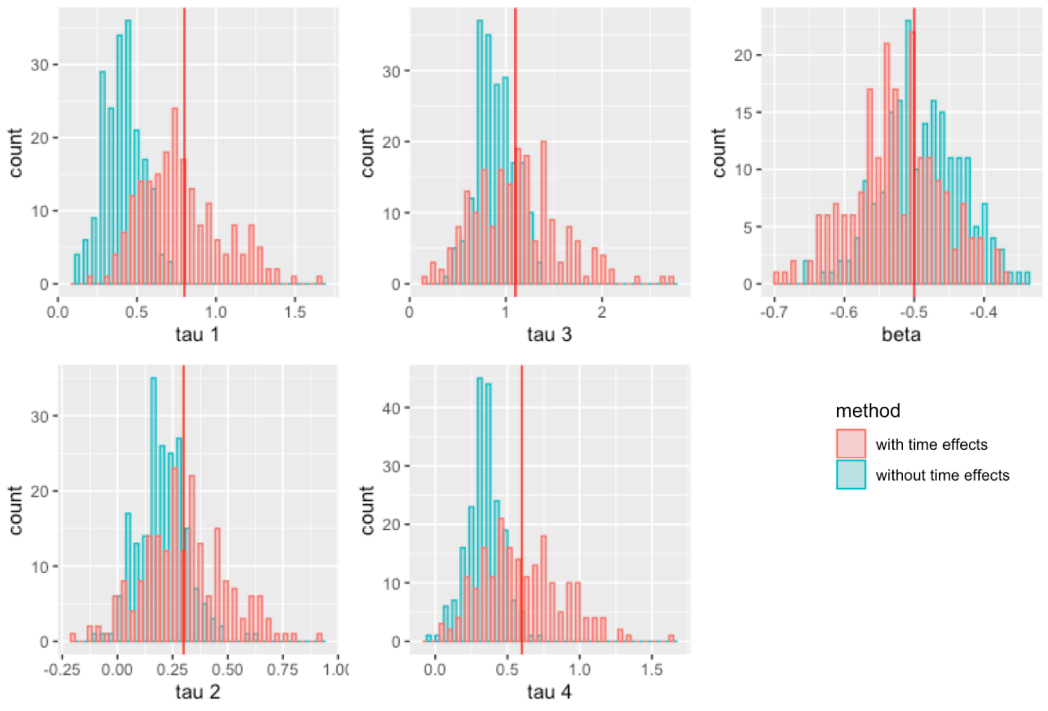

Figure 3.11. Histograms of estimated treatment effects w/o time effects for data with numerical time effects. The red line represents the true value.

We compare these results for the models, with and without time effects with Q-learning in finding the optimal DTRs. Q-learning is implemented in the same way as described in Section 3.4.1. Time effects are not included in the linear regressions for Q-functions. For each replicate, we compute the value of estimated optimal DTR and the probabilities of finding the true optimal in the overall population. Tables 3.14 and 3.15 present the mean and standard deviations of these results. As we expect, our mixture model with correct time effects has the highest probability to find the optimal DTR for the overall population and achieves the highest values of estimated optimal DTRs. Not surprisingly Q-learning outperforms our approach when our mixture model does not include time effects. This is because Q-learning's stage-wise approach to learning allows for treatment effects to vary across stages.

Table 3.14. Means and standard deviations of estimated optimal DTR values when data are generated with numerical time effect. True optima DTR value = 0.4767

Model	Mean	Std. Dev.
Mixture model with time effects	0.4727	0.0059
Mixture model without time effects	0.4426	0.0042
Q-learning	0.4526	0.0163
Mixture model after hypothesis testing	0.4608	0.0141

Table 3.15. Means and standard deviations of probabilities of finding true optimal DTRs when data are generated with numerical time effect

Model	Mean	Std. Dev.
Mixture model with time effects	64.39%	24.65%
Mixture model without time effects	23.88%	6.27%
Q-learning	43.82%	22.06%
Mixture model after hypothesis testing	52.97%	23.93%

We conduct hypothesis testing on the existence of time effects. Since we already fit the model with and without time effects, we can apply the likelihood ratio test and directly compute the test statistic λ_{LR} described in Section 3.3.3. Given $\alpha = 0.05$, it turns out that in 111 out of 200 (55.5%) datasets, the null hypothesis is rejected. For these datasets where

there is significant evidence that time effects exist, we can use the mixture model with time effects to fit the data. As for the rest, we use the model without time effects to fit the data. The mean and standard deviation of the estimated optimal DTRs and the probabilities of finding true optimal DTRs after conducting hypothesis testing are also included in Tables 3.14 and 3.15. We can see that the proposed model performs slightly better than Q-learning after checking for time effects.

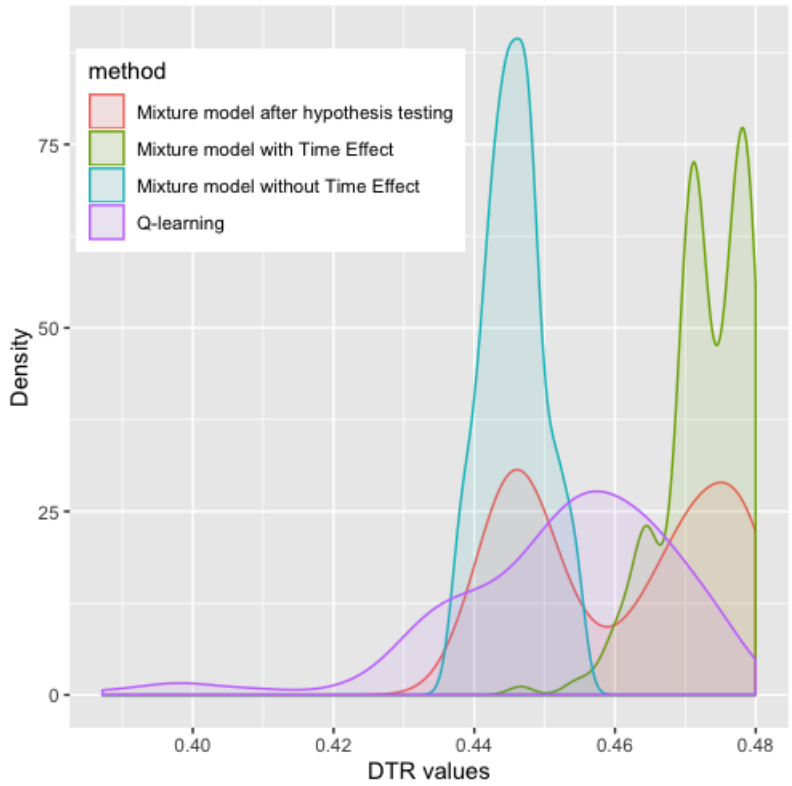

Figure 3.12. Smoothed histograms of estimated optimal DTR values in numerical time effect scenario

Figure 3.12 shows the smoothed histograms of the values of estimated optimal DTRs from different approaches when the data are generated with numerical time effects. When the model is correctly specified, the values of estimated optimal DTRs are close to the value of the true optimal DTR. When time effects are ignored, our model is likely to underperform relative to Q-learning. As a result, the values of those estimated optimal DTRs are lower. As for the mixture model after hypothesis testing, we can see that the distribution is bimodal.

This is because we select DTRs with higher values when the existence of time effects is detected, and vice versa when the time effects are ignored.

Table 3.16. Percentage of finding a better DTR than Q-learning when numerical time effects exist

Method	Percentage
Mixture model with time effects	89.0%
Mixture model without time effects	26.5%
Mixture model after hypothesis testing	67.0%

We further compare the values of the estimated optimal DTRs from each replication. Table 3.16 presents the percentage of finding a better DTR than Q-learning does. After we conduct the hypothesis testing, the percentage of finding a better DTR increases from 26.5% to 67.0%.

Categorical Time Effects

The response model with categorical time effects was described Section 3.3.2 (Equation 3.26). In this simulation study, to avoid overfitting, we also assume that the categorical time effects are not treatment-specific, i.e., all $\delta_{\omega_k,t} = \delta_t$. Along with other assumptions, $p_t = P(Y_t = 1|X_R, a_t, \boldsymbol{z}, s)$ can be written as:

$$\begin{aligned} \log(\frac{p_t}{1-p_t}) &= \tau_{a_t} + x_R\beta + \delta_t \mathbb{1}\{t \geq 1\} + s, && \text{if } z_{a_t} = 1 \\ p_t &= 0, && \text{if } z_{a_t} = 0 \end{aligned}$$

where $t = 1, 2, 3, 4$ and δ_1=0. There are 9 parameters in the logistic regression part: (τ_1, τ_2, τ_3, τ_4, $\beta, \delta_2, \delta_3, \delta_4, \sigma^2$). The true values of them are:

Parameters	τ_1	τ_2	τ_3	τ_4	β	δ_2	δ_3	δ_4	σ^2
True value	0.8	0.3	1.1	0.6	-0.5	-0.1	-0.3	-0.6	0.01

With this setting, the time effect changes nonlinearily and patients' situations will be devastated in later stages.

Same as in Section 3.4.1, our model is capable of estimating the time effects accurately as shown in Figure 3.13. We also observe biased estimates of treatment effects from the model without time effects in Figure 3.14. However, since time effects are not treatment-specific, the differences are similar among the treatments.

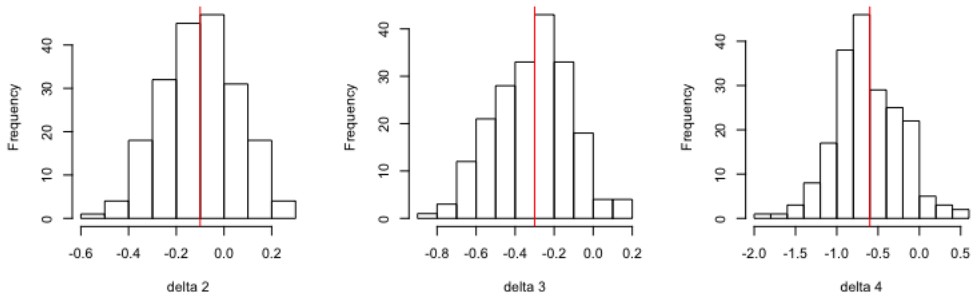

Figure 3.13. Histograms of estimated categorical time effects. The red line represents the true value.

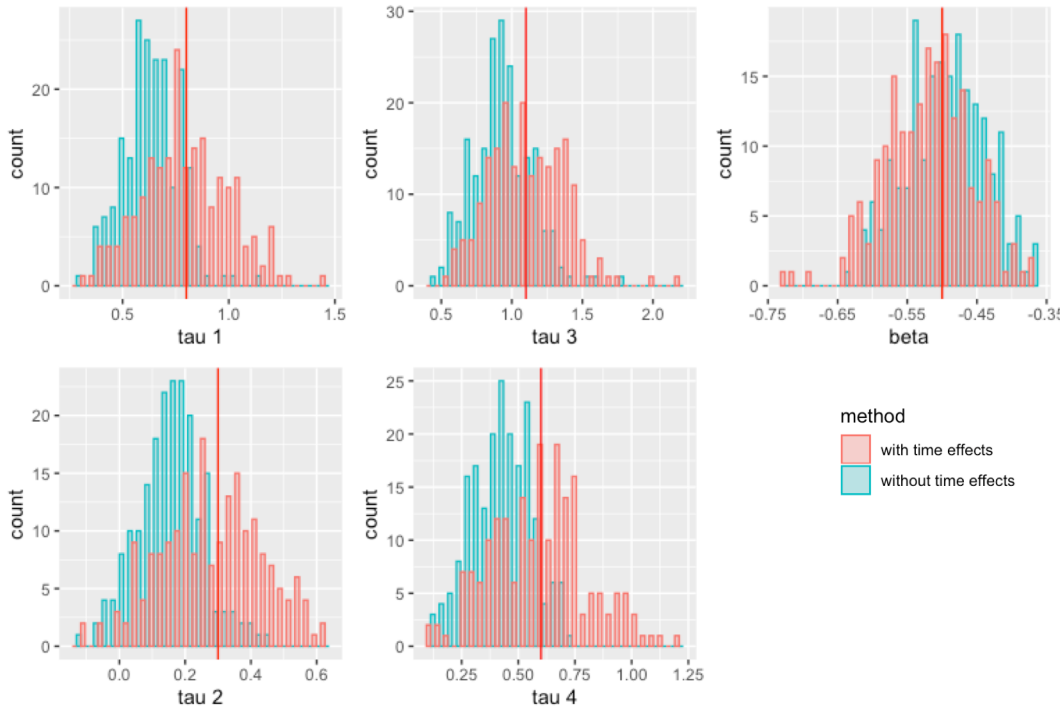

Figure 3.14. Histograms of estimated treatment effects w/o time effects for data with categorical time effects. The red line represents the true value.

We also perform the likelihood ratio test on the existence of the time effects. In 90 out of 200 (45%) datasets, the null hypothesis is rejected. As we did previously, we fit the model with categorical time effects if there is significant evidence of the existence of time effects, and the model without time effects, otherwise. The results are compared among the mixture models with and without time effects, Q-learning, and the mixture model after checking for time effects. Tables 3.17 and 3.18 present the mean and standard deviations of the values of estimated optimal DTRs and the probabilities of finding the true optimal DTR. Our conclusions are the same as in Section 3.4.1. When time effects are correctly specified in the model, our approach outperforms Q-learning. Q-learning has advantages over our approach when time effects are ignored in our response model component. However, after hypothesis testing on time effects, our model performs slightly better.

Table 3.17. Means and standard deviations of estimated optimal DTR values when data are generated with categorical time effect. True optima DTR value = 0.4882.

Model	Mean	Std. Dev.
Mixture model with time effects	0.4864	0.0020
Mixture model without time effects	0.4593	0.0044
Q-learning	0.4679	0.01389
Mixture model after hypothesis testing	0.4712	0.01411

Table 3.18. Means and standard deviations of probabilities of finding true optimal DTRs when data are generated with categorical time effects.

Model	Mean	Std. Dev.
Mixture model with time Effects	72.57%	15.01%
Mixture model without time Effects	45.35%	7.01%
Q-learning	37.30%	16.03%
Mixture model after hypothesis testing	58.05%	18.13%

Figure 3.15 shows the smoothed histograms of these values of estimated optimal DTRs. Again, the model with time effects yields higher values of DTRs. There are improvements for the mixture model after conducting hypothesis testing for time effects, compared with the model without time effects.

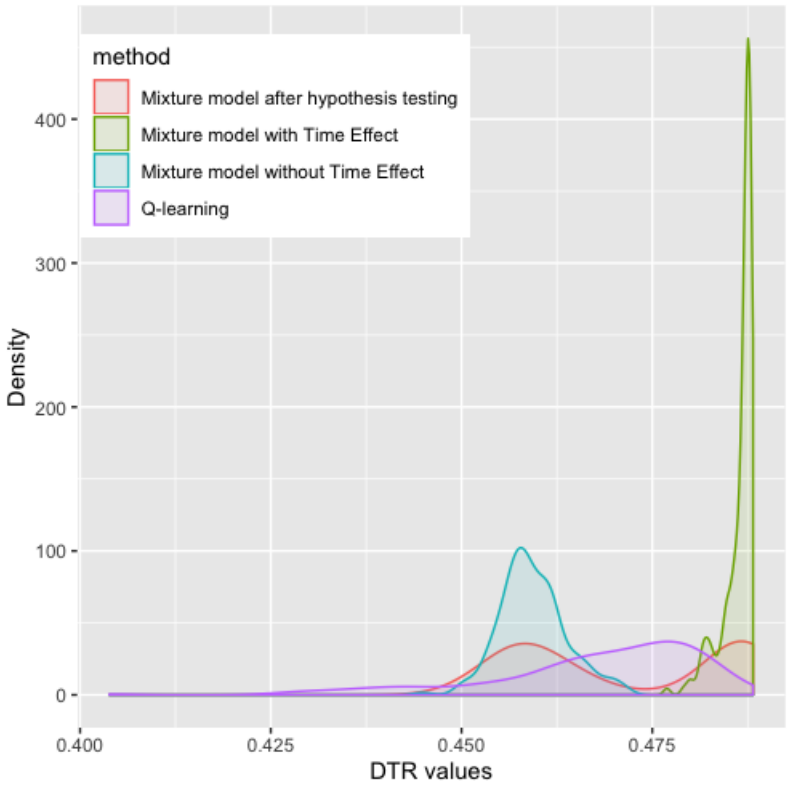

Figure 3.15. Smoothed histograms of estimated optimal DTR values

Table 3.19. Percentage of finding a better DTR than Q-learning when categorical time effects exist

Method	Percentage
Mixture model with time effects	99.5%
Mixture model without time effects	21.5%
Mixture model after hypothesis testing	58.5%

We further compare the values of the estimated optimal DTRs from each replication. Table 3.19 presents the percentage of finding a better DTR than Q-learning does. The percentage of finding a better DTR increases from 21.5% to 58.5% if we conduct hypothesis testing for the existence of time effects. Thus, it is important to identify the time effects and incorporate them into the model.

3.5 Application

The dataset from MD Anderson advanced prostate cancer trial includes baseline covariates, for example, age, prior definitive local therapy, strata, hemoglobin and alkaline phosphatase levels. In Section 2.4, we assume no prior knowledge about baseline covariates and apply our no-covariate model to fit the data. This time, we would incorporate some key covariates and assess the data again.

In previous literature, it has been found that age is a significant covariate[69]. Strata (low or high disease volume), defined in [7], is also often considered when analyzing the data[7], [68]. Therefore, we include these two covariates and view age as the covariate that impacts responses directly while strata as the covariate that might imply latent subgroups. Data are normalized before fitting. A binary score is used as the utility.

Table 3.20. Subgroup parameter estimates of MD Anderson prostate cancer trial

Parameters	$\gamma_{1,0}$	$\gamma_{1,1}$	$\gamma_{2,0}$	$\gamma_{2,1}$
Estimates	1.2799	4.1186	7.2291	8.6395
Parameters	$\gamma_{3,0}$	$\gamma_{3,1}$	$\gamma_{4,0}$	$\gamma_{4,1}$
Estimates	3.5000	4.9387	-8.2952	1.4961
Parameters	$\gamma_{12,0}$	$\gamma_{12,1}$	$\gamma_{13,0}$	$\gamma_{13,1}$
Estimates	-3.9761	-0.5351	-4.5929	-6.9271
Parameters	$\gamma_{14,0}$	$\gamma_{14,1}$	$\gamma_{23,0}$	$\gamma_{23,1}$
Estimates	8.2896	1.2236	0.5340	-6.1322
Parameters	$\gamma_{24,0}$	$\gamma_{24,1}$	$\gamma_{34,0}$	$\gamma_{34,1}$
Estimates	2.1038	-4.6506	5.7771	3.0087

There are four treatments/regimens (CVD, KA/VE, TEC, and TEE). We use our model with the homogeneous association assumption for subgroups to fit the dataset. Furthermore, we assume that age has the same effect on the responses. As a result, our model has 20 subgroup parameters and 6 response parameters. Table 3.20 presents subgroup parameter estimates.

We take a close look at how strata plays the role in subgroup probabilities. Figures 3.16 and 3.17 illustrate the underlying subgroup structure for low and high disease volume. For low volume disease, all marginal subgroup probabilities are above 95%, and 2-way interaction

subgroup probabilities are also very high, with 94.81% being the lowest. This indicates that there might be little heterogeneity among any of the four chemotherapy treatments. Patients with low volume disease, in general, can respond to any chemotherapy. On the other hand, for high volume disease, we observe huge differences among marginal subgroup probabilities. Only 23.62% can respond to CVD and 75.49% respond to TEE, while 99.68% and 94.08% can respond KA/VE and TEC respectively. This finding verifies our results in Section 2.4, where we apply our model without covariates and discover heterogeneity in CVD and TEE treatment responses. Furthermore, 2-way interaction subgroup probabilities are also consistent with our findings (Table 2.16) in Section 2.4. TEC and KA/VE have the highest overlap probability, followed by TEE and KA/VE, and TEC and TEE, which are similar. Others have similar low probabilities.

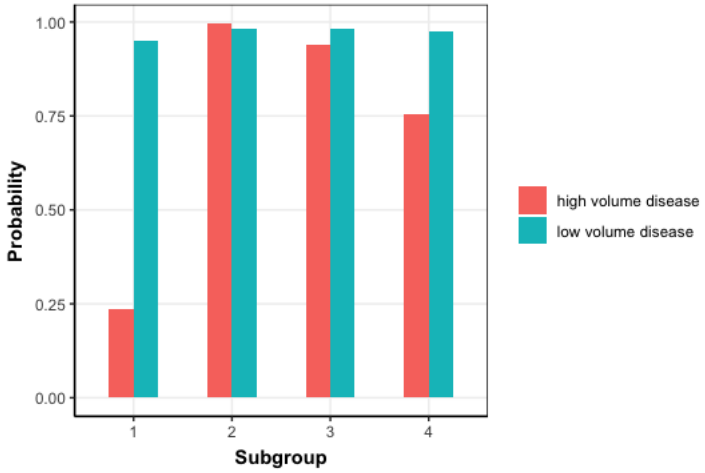

Figure 3.16. Bar plot of marginal subgroup probabilities of MD Anderson prostate cancer trial

We present response parameter estimates in Table 3.21. As we can see, patient's chances of responding favorably decreases as age increases. This is consistent with the findings [69]. KA/VE has the lowest subgroup-specific treatment effect, followed by CVD. TEC and TEE have similar treatment effects, while TEE is slightly higher.

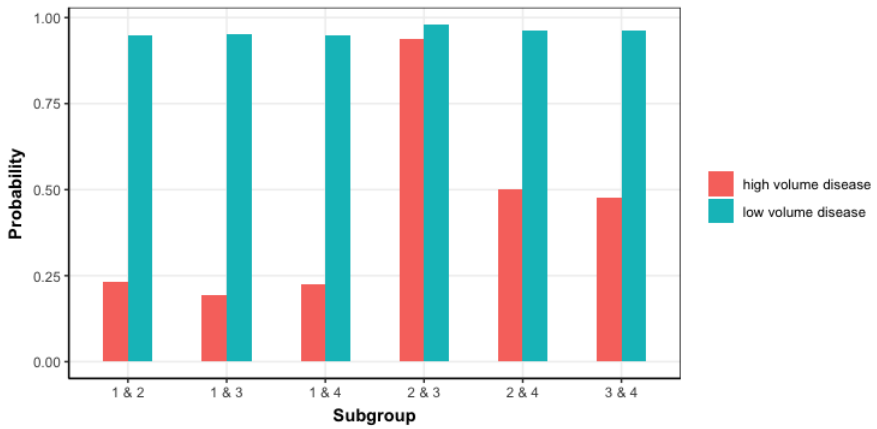

Figure 3.17. Bar plot of two-subgroup overlap probabilities of MD Anderson prostate cancer trial

Table 3.21. Response parameter estimates of MD Anderson prostate caner trial

Parameters	τ_1	τ_2	τ_3	τ_4	β_{age}	σ
Estimates	0.3624	0.2366	0.4154	0.4301	-0.2510	0.0088

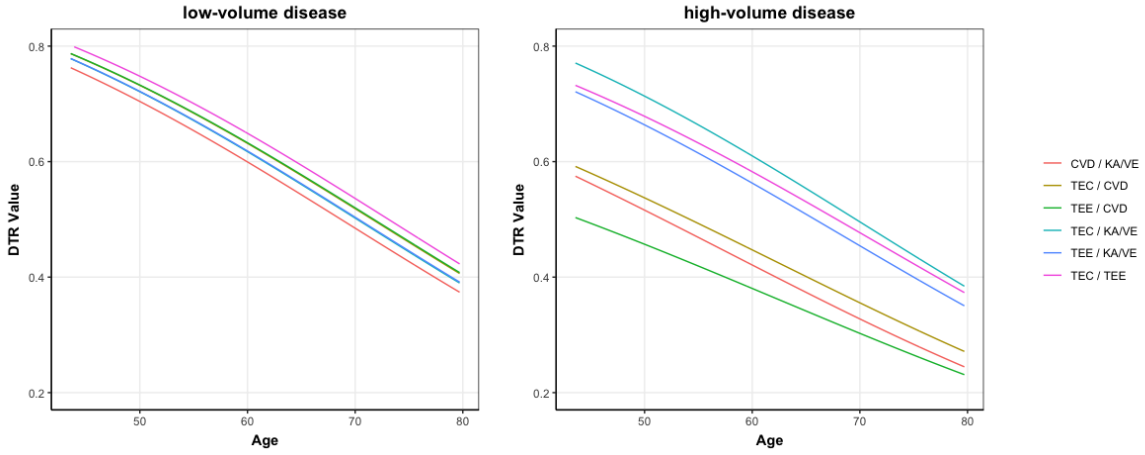

Figure 3.18. Line plots of DTR values of MD Anderson prostate cancer trial

Our goal is to discover the optimal individualized DTR for a certain disease volume and age. Based on the parameter estimates above, we calculate the values of DTRs given age for both low and high volume diseases. Figure 3.18 shows the line plot of age and DTR values.

As we can see, at both disease volumes, all DTR values go down when age increases. When disease volume is high, DTR TEC / KA/VE yields the highest value and should be assigned to patients as the optimal DTR. It is consistent with Thall's work[48], where TEC is the best first-line treatment while KA/VE is the best second-line treatment. The interesting part among patients with low volume disease. Our analysis shows that though these DTR values are relatively close to each other, TEC / TEE is the highest. Thus, the baseline covariate strata indeed determines the optimal individual DTR.

4. SUMMARY AND FUTURE WORK

4.1 Summary

Research in DTRs plays an important role in the management of long-term care for chronic diseases and complex behaviors. In this dissertation, we focused on multistage management programs with a binary response (i.e., the patient either responds or does not respond to treatment). Using data collected from a randomized trial like a SMART design, we described a model-based approach that can be used to compare the values of DTRs and find the optimal one. This modeling approach is unique in that it addresses discrete heterogeneity through the use of latent subgroups. We assume that for each treatment, patients can be divided into two subgroups with one subgroup representing patients who could benefit from the treatment and the other including those patients who would not. This particular type of heterogeneity has received little attention so far in DTR research.

Due to concerns over model misspecification, a large number of methods that have been proposed to estimate the optimal DTR are model-free, and therefore, cannot model the discrete heterogeneity. In taking a model-based approach, if the model is reasonable, we can not only find the optimal DTR but also provide prognostic information about the treatments. Furthermore, when the same treatments are assigned repeatedly across stages, outcome information can be combined across stages, thereby making our model-based approach more sample-efficient. As a result, we can better estimate the value of DTRs when our proposed model fits the data well. Currently, only very limited work has been attempted to characterize how patients react to treatments stage-by-stage through model-based methods [7], [49]. We also take this a step further by incorporating both the continuous heterogeneity that can be described by a smooth function of baseline covariates, and the discrete heterogeneity due to latent subgroups. Our primary contributions in this dissertation are to raise awareness of the discrete heterogeneity when determining optimal DTR and propose a model-based framework that accounts for it when the outcome is binary.

In Chapter 2, we introduced our model-based approach without considering baseline covariates. We described the probabilistic framework of treatment/response trajectories for a SMART design under this model and our approach to estimation. We focused on

the situation where the same K treatments are under consideration at each stage. In this setting, there are a total of 2^K subgroup proportions, K treatment effects, and the variance of a subject-specific random effect as unknown model parameters. The subject-specific random effect is also included in the logit model.

Because each of the 2^K subgroups has a unique response profile to the K treatments, we describe our model as a mixture of mixed logit models. We use an EM algorithm with Gaussian Hermite quadrature for the integration approximation, to estimate model parameters. Because mixture models often suffer from nonidentifiability, we include a discussion about this, distinguishing the identifiability of model parameters and the value of a DTR.

In Chapter 3, we extend our approach to include baseline covariates, thereby making the optimal DTR more individualized. These covariates are separated into two classes — those that are related to the probability of an effective response and those that are associated with the probability of subgroup identity. We include the former covariates in the logistic regression component of our model. As for the latter, we utilize the multivariate Bernoulli distribution to incorporate them into the determination of the latent subgroup. Both independent and homogeneous association structures are considered. In addition, we extend our model to incorporate two kinds of time effects into the logistic regression model for responses and illustrate how to test the existence of time effects. We adapted our EM algorithm to include covariates and time effects.

We conclude each of these chapters with simulation studies. These simulations demonstrate the accuracy of our approach, although larger than typical sample sizes are needed for high precision. We also compare the performance of our model-based approach to Q-Learning. When the model assumptions are correct, our approach outperforms Q-Learning. Our model's chance of finding the true optimal DTR is higher than Q-Learning. Even if the true optimal DTR is missed, we select DTRs with higher values as the estimated optimal DTR. This is because Q-Learning determines the optimal decision rule based solely on the information from a given stage, while our approach is capable of fully utilizing information across stages when treatments are repeatedly assigned across stages. Also, Q-Learning appears to struggle with a correlated subgroup structure and produces biased estimations of Q-functions. These advantages rely on valid model assumptions. In Section 3.4.3, we further

fit the data by a model which ignores the existing time effects. Compared with the model without time effects, Q-Learning can naturally adjust for time, and as a result, select the DTRs with higher values. However, hypothesis testing can be used to detect the time effects and find the adequate model most of the time.

4.2 Future Work

As with any novel statistics approach, several challenges arose that were not fully addressed in this dissertation. This section highlights some of the important ones.

4.2.1 The Existence of Subgroups

In all the simulations presented in the dissertation, we assumed that there were underlying latent subgroups. An obvious follow-up question is "how does the approach work if there are no subgroups?" In other words, it is worth exploring the performance when all the subjects do respond to each of the K treatments. Our expectations are that our model approach will be robust but further work is needed to validate this claim. We did conduct a preliminary study to assess this.

A total of 200 datasets with sample size $N = 1200$ were generated with X_R and X_Z following independent standard Normal distributions. Instead of using the multivariate Bernoulli model to generate the patient's subgroup, we assumed $\boldsymbol{Z} = (1, 1, 1, 1)$ for all patients in the data generation process. Treatment responses are generated as in Equation (3.17), which parameters are set as below:

τ_1	β_1	τ_2	β_2	τ_3	β_3	τ_4	β_4	σ^2
0.1	-0.5	0.5	0.6	-0.2	0.4	0.3	1	0.01

Each dataset is fitted by the proposed model with the independent subgroup structure that is described in Section 3.1.2. Figure 4.1 shows the boxplots of estimated subgroup probabilities given $X_Z = 1$ and -1. As we can see, the distributions of subgroup probabilities for each subgroup are similar. Also, the majority of the estimated probabilities are close to 1, regardless of the value of X_Z. These indicate that our model can, to some degree, detect that there are no latent subgroups.

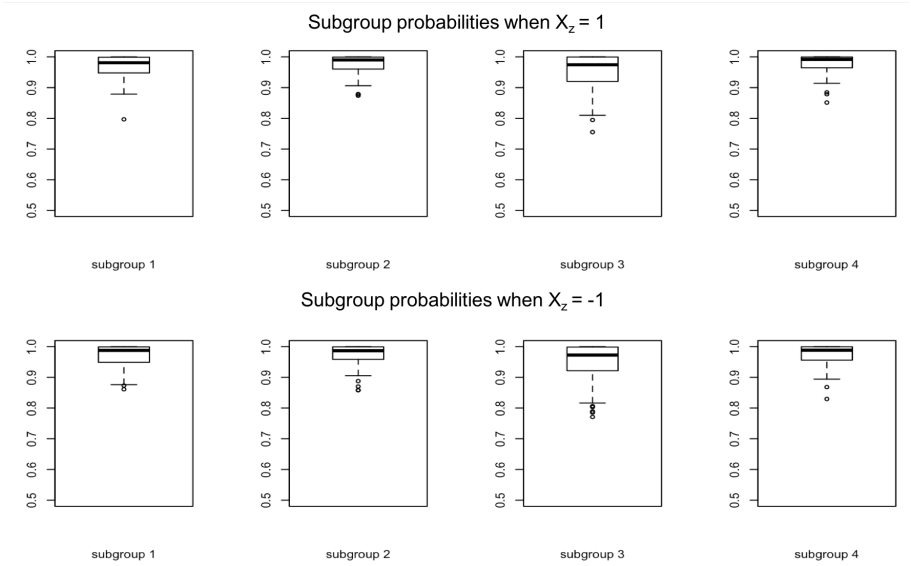

Figure 4.1. Boxplots of estimated subgroup probabilities given $X_Z = 1$ and -1. The true probability is 1.

Table 4.1. Means and standard deviations of estimated optimal DTR values when discrete heterogeneity doesn't exist. True optima DTR value = 0.6655

Model	Mean	Std Dev
Mixture Model	0.6649	0.0001
Q-Learning	0.6443	0.0179

Table 4.2. Means and standard deviations of probabilities of finding the true optimal DTR when discrete heterogeneity doesn't exist.

Model	Mean	Std Dev
Mixture Model	98.45%	2.36%
Q-Learning	73.48%	21.16%

We further fit the data by Q-Learning. Tables 4.1 and 4.2 present the results of the value of the estimated optimal DTR and the probabilities of finding the true optimal DTR for both methods. As we can see, our model has advantages over Q-Learning. This is because our model can adequately estimate the subgroup probabilities (close to 1) and also utilize information of treatment effects across stages, and therefore, is more efficient.

4.2.2 Model Diagnostics and Goodness of Fit of Linearity Assumptions

As we saw in Section 3.4.3, when time effects are ignored, our model has difficulty in providing accurate parameter estimates and finding the optimal DTR. Therefore, to achieve better performance, it is important to conduct model diagnosis and check the goodness of fit.

In both the multivariate Bernoulli model component describing subgroup probabilities and the mixed effect logistic model component for the response, we assume that the baseline covariates enter the model linearly. In this section, we consider some simple examples and show how to use randomized quantile residual plots for model diagnostics and the Hosmer-Lemeshow test for the goodness of fit. Further investigation and more comprehensive hypothesis testing approaches are needed to verify that these methods are helpful tools in general.

Randomized Quantile Residual Plots

Residual plots can play a crucial role in checking the linearity assumption of the covariates. They are commonly used in the diagnostics of the linear model. However, we cannot easily use traditional residual plots for logistic regression. Dunn (1996) proposed randomized quantile residuals [85], which was designed for generalized linear models. For each response, a uniform random variable is generated by inverting the cumulative distribution function. Then, the randomized quantile residual is defined as the value that finds the equivalent standard normal quantile.

Our model is much more complex than a logistic regression model. At each stage, there is a mixture of mixed logistic regression models, and together, we can have multiple stages in our full-fledged model. As a result, we cannot apply the randomized quantile residual plots to our model directly, but the core idea can be borrowed. In this section, we illustrate how to use randomized quantile residual plots to assess the linearity assumption in the response model component and the subgroup model component respectively. Our analysis is carried out based on generated datasets and assume there is one baseline covariate X_R for the response model and one X_Z for the subgroup model.

We first consider the response model component. Recall the protocol of MD Anderson' trial, subjects receive the same treatment in the next stage if they respond in the current stage. Since we assume the natural response rate is zero, the individual must be in the favorable treatment subgroup if the patient responds favorably. Therefore, if subjects receive the treatment for the first time and respond favorably at Stage t, they will be in the favorable subgroup of the treatment at Stage $t+1$. For example, in Stage 1 subjects are randomly assigned a treatment. If they respond favorably, then we know at Stage 2 they receive the treatment again and are in the favorable subgroup. As a result, given that subjects respond favorably to treatment ω_k for the first time at the previous stage, the logit function can be written as:

$$\log(\frac{p_t}{1-p_t}) = \tau_{\omega_k} + x_R\beta_{\omega_k} + s, \tag{4.1}$$

which doesn't depend on the latent subgroup indicator anymore.

We create a subset of data by keeping the current stage information only if subjects respond favorably to the treatments at the previous stage. We then fit the subset by mixed-effects logistic model directly and evaluate the linearity assumption in the response model component.

In Example 4.1, we generate a dataset of independence model with $N = 1200$. A quadratic term of X_R is included in the logistic regression:

$$\log(\frac{p_t}{1-p_t}) = \tau_{\omega_k} + x_R\beta_{\omega_k} + x_R^2\zeta_{\omega_k} + s,$$

Below is the parameter setting for the quadratic terms:

ζ_1	ζ_1	ζ_1	ζ_1
0.25	0.5	0	-0.8

The absolute values of coefficients of X_R^2 rank as treatment $4 > 2 > 1$ while no X_R^2 term for treatment 3, i.e., $\zeta_4 > \zeta_2 > \zeta_1 > \zeta_3 = 0$. The rest of parameters are the same as in Section 3.4.1.

After the subset of data is extracted, we fit the data with mixed effects logistic model as in Equation (4.1), pretending we don't know the existence of the quadratic term. We

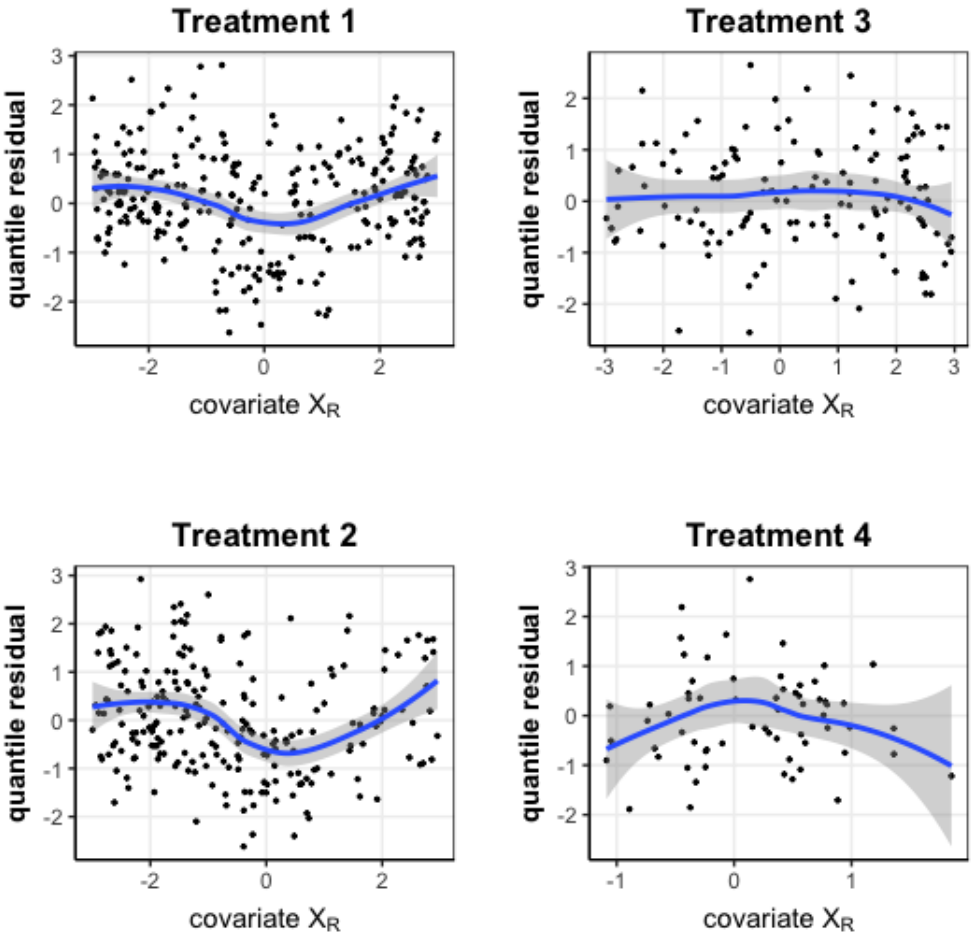

Figure 4.2. Residual plot of X_R in Example 4.1

integrate the subject effect and calculate the probability of $P(Y|X_R)$. Randomized quantile residuals are then computed and plotted versus X_R as in Figure 4.2. Clearly, we can see a bending curve in Treatment 4. Treatments 1 and 2 also show the sign of non-linearity. These indicate potential model misspecification and suggest that a nonlinear form should be considered in the response model component.

Since the subgroup identity is unobserved, it is not feasible to solely check the linearity assumption in the subgroup model component. Instead, we can calculate the randomized quantile residuals of the mixture model for the probability of responding favorably at each stage. Recall that in the later stages of the trial, this probability also depends on previous

treatment and response trajectory. For example, the probabilities of responding favorably to the same treatment at Stage 2 are different between the subjects who receive this treatment and respond favorably at Stage 1 and the subjects who receive another treatment and respond unfavorably at Stage 1. Therefore, additional calculations conditioning on previous treatment and response trajectories are required to compute the probability of responding favorably at later stages of the trial. For simplicity, we only focus on the analysis of the data collected at Stage 1.

The probability of Y_1 is $P(Y_1|X_Z, X_R) = \sum_{\boldsymbol{Z}} P(\boldsymbol{Z}|X_Z) \times P(Y_1|X_R, \boldsymbol{Z})$. Previously, we've shown how to perform model diagnostics for the response model component. If the response model $P(Y_1|X_R, \boldsymbol{Z})$ is appropriate, then by checking the randomized quantile residual plots of $P(Y_1|X_Z, X_R)$, we indirectly evaluate the model fit of the subgroup model $P(\boldsymbol{Z}|X_Z)$.

A dataset of independence model with $N = 1200$ is generated for Example 4.2. We include X_Z^2 in multivariate Bernoullli model, which now is:

$$\log \frac{P(Z_{\mathrm{i}} = 1, \text{and others are } 0|x_Z)}{P(\boldsymbol{Z} = (0,0,0,0)|x_Z))} = \gamma_{\mathrm{i},0} + x_Z\gamma_{\mathrm{i},1} + x_Z^2\gamma_{\mathrm{i},2} \tag{4.2}$$

where $\mathrm{i} = 1, 2, 3, 4$. Below is the parameter setting in the multivariate Bernoullli model.

$\gamma_{1,0}$	$\gamma_{1,1}$	$\gamma_{1,2}$	$\gamma_{2,0}$	$\gamma_{2,1}$	$\gamma_{2,2}$	$\gamma_{3,0}$	$\gamma_{3,1}$	$\gamma_{3,2}$	$\gamma_{4,0}$	$\gamma_{4,1}$	$\gamma_{4,2}$
0.3	1.5	0	1	-2	-1	-1	1	2	0.5	-1	1.5

The absolute values of coefficients of X_Z^2 in the subgroup model rank as treatment $3 > 4 > 2$, while no X_Z^2 term in the subgroup model for treatment 1. The parameter setting of the response model is the same as in Section 3.4.1.

We apply our mixture model with the independence subgroup structure to fit the data. Based on the parameter estimates, we integrate out the subject effect and calculate the probability of $P(Y_1|X_Z, X_R)$ at Stage 1. We further get the randomized quantile residuals. Figure 4.3 presents the randomized quantile residual plots of X_Z. The plots of Subgroups 3 and 4 suggest non-linearity while subgroup 2 shows a marginal sign. No significant trend is detected for subgroup 1. This is consistent with our parameter setting for X_Z^2.

Based on Example 4.1 and Example 4.2, non-linear relationships in the model can be detected by checking the randomized quantile residual plots of X_R and X_Z separately. In

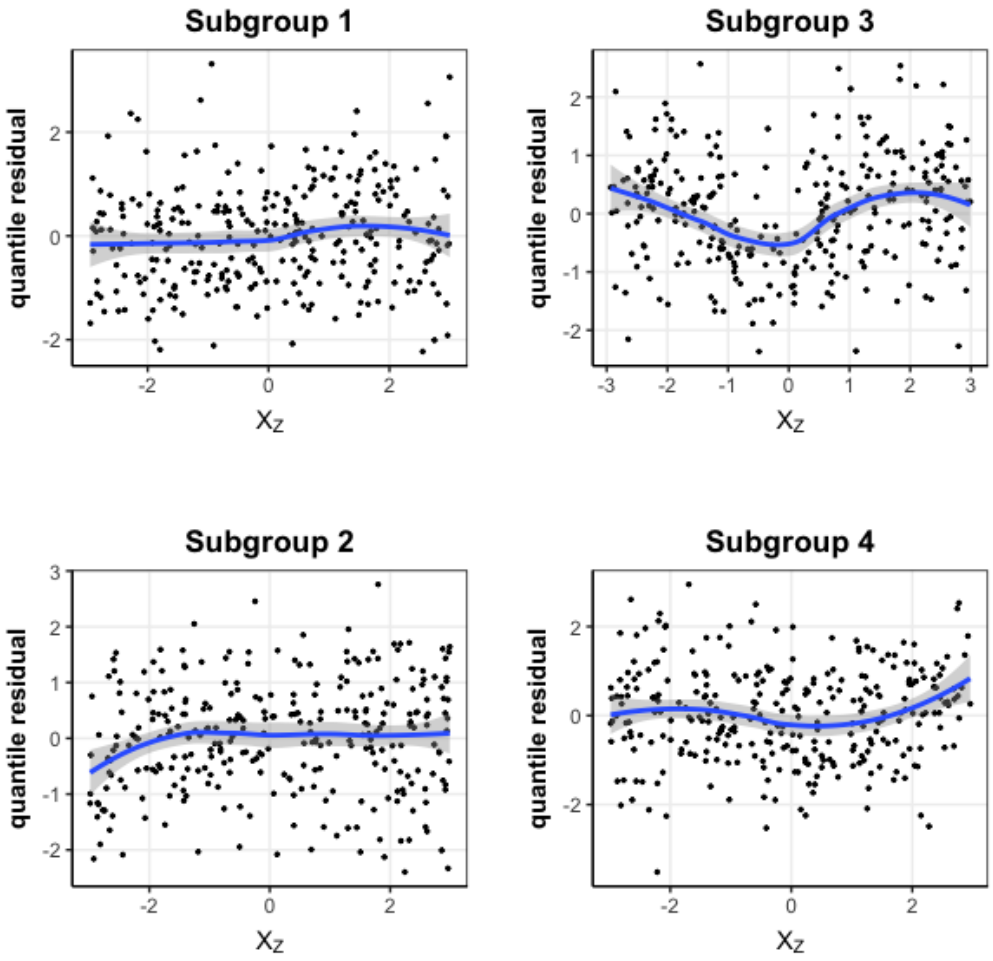

Figure 4.3. Residual plot of X_Z in Example 4.2

practice, given any dataset, we can first evaluate the linearity assumption in the response model component through randomized quantile residual plots of X_R. According to the pattern of the residual plots, additional terms may be considered. If the model looks appropriate, then we start to evaluate the linearity assumption in the subgroup model component.

Hosmer-Lemeshow test

Besides randomized quantile residual plots, we can check the linearity assumption by the goodness of fit test. Hosmer-Lemeshow test[86] is a statistical test for goodness of fit for the logistic regression model. This test divides the data into groups according to the

predicted probabilities first, and then evaluates whether or not the observation matches the expectation, which is similar to Pearson's chi-square test. In this section, we illustrate how to conduct the Hosmer-Lemeshow test to check the goodness of fit of our proposed model.

Due to the reason mentioned before that the probability of responding favorably depends on the previous treatment and response trajectory, we conduct our analysis for Stage 1 data only. To get the Hosmer-Lemeshow test statistic, we first fit the data by our proposed model and obtain the predicted probability of $P(Y_1|X_Z, X_R)$ based on the parameter estimates. Next, we sort $P(Y_1|X_Z, X_R)$ from the smallest to the largest and equally divide them into M groups in order. Then, in group m $(m = 1, ..., M)$, we count the number of observed $Y = 0$ and $Y = 1$, which are denoted as $O_{0,m}$ and $O_{1,m}$ respectively. Let $E_{1,m}$ be the expected number of $Y = 1$. It is computed as the sum of the probabilities of responding favorably for all subjects in the group m. Similarly, we get $E_{0,m}$, the expected number of $Y = 0$, as the sum of the probabilities of an unfavorable response for all subjects in the group m. Eventually, the Hosmer-Lemeshow test statistic can be calculated as below:

$$H = \sum_{m=1}^{M} \left(\frac{(O_{1,m} - E_{1,m})^2}{E_{1,m}} + \frac{(O_{0,m} - E_{0,m})^2}{E_{0,m}} \right)$$

In Example 4.1, following the steps above and dividing the data into 10 groups, we can get the Hosmer–Lemeshow test statistics and p-values shown in Table 4.3. With $\alpha = 0.05$, we can reject the null hypothesis of the linearity model for treatments 4 and 2, which have the top 2 large coefficients of X_R^2.

Table 4.3. Hosmer-Lemeshow test statistic for quadratic form in response model in Example 4.1

Treatment	**1**	**2**	**3**	**4**
Hosmer-Lemeshow statistic	9.010408	16.56187	6.407669	14.89042
p-value	0.3414187	0.03500873	0.6016663	0.06131151

We do the same for the subgroup model in Example 4.2. Table 4.4 shows the Hosmer-Lemeshow test statistic. It shows linearity assumption fails for subgroups 2 and 3 while 4 is on the edge.

Table 4.4. Hosmer-Lemeshow test statistic for quadratic form in subgroup model in Example 4.2

Treatment	**1**	**2**	**3**	**4**
Hosmer-Lemeshow statistic	6.048619	17.444	25.76688	14.97124
p-value	0.6417857	0.02580385	0.001151223	0.059707

As we can see from both tables, the Hosmer-Lemeshow test can detect the non-linearity in the model when the magnitude of the quadratic term is large enough.

4.2.3 Computational Challenge

Our current estimation approach involves the EM algorithm, which can be slow to converge. In our setting, this is further complicated by the heavy computation needed at the M step when baseline covariates are taken into consideration. For example, in Chapter 3 where we add a covariate in both the response and subgroup components of the model, we presented that the conditional expectation of complete data log-likelihood can be maximized by maximizing two parts separately. One is for parameters for the subgroup identity, and the other is for parameters in the logistic regression for the response. The latter

$$
\begin{aligned}
&Q_2(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}) \\
&= \sum_{l=1}^{2^K} P(C_l^{\mathrm{i}} = 1|\boldsymbol{Y}^{\mathrm{i}}, \boldsymbol{A}^{\mathrm{i}}, \boldsymbol{X}^{\mathrm{i}}, \boldsymbol{\theta^{(t)}}) \{\sum_{\mathrm{i}=1}^{N} \log \int \{\prod_{t=1}^{T_{\mathrm{i}}} \exp(y_t^{\mathrm{i}} \eta_t^{\mathrm{i}} - \log(1+\eta_t^{\mathrm{i}}))\} \phi(s) ds\},
\end{aligned}
$$

includes the random subject effect that we have to integrate out. Since there is no closed-form solution, Gauss-Hermite quadrature with 25 points is implemented to approximate the integration numerically. Then $Q_2(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})$ is approximated by

$$
\sum_{l=1}^{2^K} P(C_l^{\mathrm{i}} = 1|\boldsymbol{Y}^{\mathrm{i}}, \boldsymbol{A}^{\mathrm{i}}, \boldsymbol{X}^{\mathrm{i}}, \boldsymbol{\theta^{(t)}}) \Big\{ \sum_{\mathrm{i}=1}^{N} \log \Big\{ \frac{1}{\sqrt{\pi}} \sum_{\mathrm{j}=1}^{J} (\prod_{t=1}^{T_{\mathrm{i}}} \exp(y_t^{\mathrm{i}} \tilde{\eta}_{t\mathrm{j}}^{\mathrm{i}} - \log(1+\tilde{\eta}_{t\mathrm{j}}^{\mathrm{i}}))) g_{\mathrm{j}} \Big\} \Big\}.
$$

Due to the complex form of this approximation, i.e., taking the logarithm of summations, we use the numerical gradient instead of the analytic one. As a result, this part takes a significantly longer time for maximization. For example, with sample size $N = 2000$, at

each iteration in the EM algorithm, maximization over subgroup parameters takes from a few seconds to one and half minutes. In contrast, maximizing $Q_2(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})$ can take from 100 seconds to more than 25 minutes. In general, we might wait for 4 to 6 hours when $N = 600$, and 15 to 30 hours when N increases to 2000.

One immediate opportunity is to reduce the computational time in the approximation of $Q_2(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})$. The expected value of $Q_2(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})$ can be viewed as the sum of log-likelihood of the trajectory weighted by conditional subgroup probabilities given observed data. As mentioned in Lee (2016) [87], according to this inherent structure of the EM algorithm, most of the expressions that need to be evaluated on the M-steps can be performed independently for each subgroup. Therefore, we could consider parallelizing the calculation of $Q_2(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})$ by subgroups. The conditional expectation of log-likelihood of trajectory given subgroup identity is computed on an individual core. Then, the results from each core are combined to obtain the global result as if it requires the whole data to be analyzed at once. Given that K treatments of interest mean 2^K subgroups, by paralleling the data across subgroups, the evaluation time can be reduced to about $1/2^K$ of the original time. As the number of treatments increases, the computational time decreases dramatically compared with computing everything on a single core. Note that the partition takes place on subgroups rather than the dataset. This is because the sample size is often relatively small in the research of DTR. Dividing the dataset into many very small chunks may result in unfulfilled utilization of each core's computation power and more time spent on bringing results back.

So far, only two baseline covariates are considered in our simulation in Chapter 3. When more baseline covariates are available, we anticipate the computation time of the current EM algorithm to dramatically increase. One may have to find better methods for maximization. For example, Laplace approximation, Adaptive Gaussian Quadrature approximation and Penalized quasi-likelihood can be considered for the approximation of the integration. These methods are widely used for generalized linear mixed models, and SAS and R packages have been developed. However, not much work has been done to extend them for latent subgroups. In addition, Monte Carlo EM can also be implemented to avoid numerical approximation. It is worth comparing the performance of these methods and figuring out the best one for our model.

4.2.4 Model Extensions

In order to accurately find the optimal DTR, it is of great importance to avoid model misspecification and to identify all possible factors that affect individuals' responses. Our model consists of two parts. One is to model subgroup memberships, and the other is to model treatment response given subgroup memberships. Given that subgroup membership does not change over time, it is enough to utilize baseline covariates to predict subgroup memberships. On the other hand, there are still many other covariates that can be taken into account for modeling treatment responses. In Section 3.3.4, we explore the model performance under time effects. We foresee potential opportunities of incorporating more factors into the logistic regression model for treatment response.

Residual effects of previous treatments, for example, are worth checking. Sometimes when there is no washout period between stages, the treatment can have a direct effect on the individual when assigned at the stage of its application, as well as an i^{th}-order residual effect at the i^{th} stage after the treatment's application is discontinued [88]. In general, the second and higher-order residual effects are negligible, but the first-order residual effect can be incorporated into the logistic regression for the response model. For example, when $t > 2$, the logit function in our mixture of mixed Logit models (see Equation (2.2)) can be written as:

$$\log(\frac{p_t}{1-p_t}) = \mu + \sum_{k=1}^{K} \mathbb{1}\{d_t(a_1, ..., y_{t-1}) = \omega_k\}\tau_{\omega_k} z_{\omega_k} + \sum_{k=1}^{K} \mathbb{1}\{a_{t-1} = \omega_k\}\delta_{\omega_k} + s$$

where $\sum_{k=1}^{K} \mathbb{1}\{a_{t-1} = \omega_k\}\delta_{\omega_k}$ is the first order residual effect of previous treatment.

Besides, like Thall (2007) [7], one can consider including previous stage response Y_{t-1} as well as entire response history $\{Y_1, ..., Y_{t-1}\}$ into the model at t^{th} stage. It measures how previous responses, especially unfavorable responses, affect the individual's response at later stages.

Currently, in our work, we assume the individual random effect stays the same across stages. This assumption can be relaxed. Let s_t be the subject effect at t^{th} stage and $\boldsymbol{s} = (s_1, ..., s_T)$ is the vector of s_t. We can rewrite Equation (2.2) as:

$$\log(\frac{p_t}{1-p_t}) = \mu + \sum_{k=1}^{K} \mathbb{1}\{d_t(a_1..., y_{t-1}) = \omega_k\}\tau_{\omega_k} z_{\omega_k} + s_t.$$

where $\boldsymbol{s}$ follows multivariate Normal distribution $\boldsymbol{N}(\boldsymbol{0}, \Sigma_T)$. Σ_T is the covariance matrix of within-subject errors. The same changes can be made for Equation (3.2). One can make assumptions on the covariance matrix, like variance components, first-order autoregressive AR(1), or even unstructured covariance, to allow for a more complex correlation structure of subject effects. As for model estimation, multivariate Gauss-Hermite quadrature [89] can be implemented to evaluate the integration numerically.

4.2.5 Continuous Response

In this research, we focus on multi-stage trials with binary responses, like respond or not respond. Sometimes, continuous responses are observed at each stage. We may observe a certain subgroup of patients have a higher treatment effect than the rest of the patients do. In these cases, with consideration of latent subgroups, we can assume that the response is a mixture of Normal distributions. As for model estimation, we could either use the EM algorithm or adopt the Newton-Raphson family method. If the EM algorithm is adopted, random effects are viewed as missing values. Compared with a binary response, the complete-data likelihood is a joint Normal distribution, and therefore a much nicer form can be obtained after taking the logarithm on the likelihood. Furthermore, if the responses follow a Normal distribution, the optimization of the log-likelihood can be transferred to a non-linear least-square problem. Cécile Proust [90], [91] has pointed out that the Marquardt algorithm [92] has advantages over the EM algorithm due to better convergence rate and computational speed.

4.2.6 Identifiability

When baseline covariates are not considered in Chapter 2, we list the sufficient condition of local identifiability. We also demonstrate our model-based approach is still able to estimate trajectory probability and the value of DTRs, even though in the situation when parameters are not identifiable in section 2.3. However, identifiability remains an unsolved issue if we plan to extend our model to incorporate more baseline covariates and other factors like residual effects and history responses, or to relax subject effect covariance constraints. Future research needs to be done to figure out these identifiability conditions, in order to guarantee the unique parameter estimates and better understand the heterogeneity in treatment response.

REFERENCES

[1] I. B. C. S. G. (IBCSG), "Endocrine responsiveness and tailoring adjuvant therapy for postmenopausal lymph node-negative breast cancer: A randomized trial," *CancerSpectrum Knowledge Environment*, vol. 94, no. 14, pp. 1054–1065, Jul. 2002. DOI: 10.1093/jnci/94.14.1054. [Online]. Available: https://doi.org/10.1093/jnci/94.14.1054.

[2] A. Ramamoorthy, M. Pacanowski, J. Bull, and L. Zhang, "Racial/ethnic differences in drug disposition and response: Review of recently approved drugs," *Clinical Pharmacology & Therapeutics*, vol. 97, no. 3, pp. 263–273, Jan. 2015. DOI: 10.1002/cpt.61. [Online]. Available: https://doi.org/10.1002/cpt.61.

[3] F. R. Vogenberg, C. Isaacson Barash, and M. Pursel, "Personalized medicine: part 1: evolution and development into theranostics," *P T*, vol. 35, no. 10, pp. 560–576, Oct. 2010.

[4] P. W. Lavori and R. Dawson, "Dynamic treatment regimes: Practical design considerations," *Clinical Trials*, vol. 1, no. 1, pp. 9–20, 2004, PMID: 16281458. DOI: 10.1191/1740774504cn002oa. eprint: https://doi.org/10.1191/1740774504cn002oa. [Online]. Available: https://doi.org/10.1191/1740774504cn002oa.

[5] L. M. Collins, S. A. Murphy, and K. L. Bierman, "A conceptual framework for adaptive preventive interventions," *Prevention Science*, vol. 5, no. 3, pp. 185–196, Sep. 2004. DOI: 10.1023/b:prev.0000037641.26017.00. [Online]. Available: https://doi.org/10.1023/b:prev.0000037641.26017.00.

[6] B. Chakraborty and S. A. Murphy, "Dynamic treatment regimes," *Annual Review of Statistics and Its Application*, vol. 1, no. 1, pp. 447–464, 2014. DOI: 10.1146/annurev-statistics-022513-115553. eprint: https://doi.org/10.1146/annurev-statistics-022513-115553. [Online]. Available: https://doi.org/10.1146/annurev-statistics-022513-115553.

[7] P. F. Thall, C. Logothetis, L. C. Pagliaro, S. Wen, M. A. Brown, D. Williams, and R. E. Millikan, "Adaptive therapy for androgen-independent prostate cancer: A randomized selection trial of four regimens," *JNCI Journal of the National Cancer Institute*, vol. 99, no. 21, pp. 1613–1622, Oct. 2007. DOI: 10.1093/jnci/djm189. [Online]. Available: https://doi.org/10.1093/jnci/djm189.

[8] B. Chakraborty and E. E. M. Moodie, "Statistical methods for dynamic treatment regimes: Reinforcement learning, causal inference, and personalized medicine," in. New York, NY: Springer New York, 2013, ISBN: 978-1-4614-7428-9. DOI: 10.1007/978-1-4614-7428-9_2. [Online]. Available: https://doi.org/10.1007/978-1-4614-7428-9_2.

[9] J. Splawa-Neyman, D. M. Dabrowska, and T. P. Speed, "On the application of probability theory to agricultural experiments. essay on principles. section 9," *Statistical Science*, vol. 5, no. 4, pp. 465–472, Nov. 1990. DOI: 10.1214/ss/1177012031. [Online]. Available: https://doi.org/10.1214/ss/1177012031.

[10] D. B. Rubin, "Estimating causal effects of treatments in randomized and nonrandomized studies.," *Journal of Educational Psychology*, vol. 66, no. 5, pp. 688–701, 1974. DOI: 10.1037/h0037350. [Online]. Available: https://doi.org/10.1037/h0037350.

[11] D. B. Rubin, "Bayesian inference for causal effects: The role of randomization," *The Annals of Statistics*, vol. 6, no. 1, pp. 34–58, Jan. 1978. DOI: 10.1214/aos/1176344064. [Online]. Available: https://doi.org/10.1214/aos/1176344064.

[12] J. Robins, "A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect," *Mathematical Modelling*, vol. 7, no. 9-12, pp. 1393–1512, 1986. DOI: 10.1016/0270-0255(86)90088-6. [Online]. Available: https://doi.org/10.1016/0270-0255(86)90088-6.

[13] J. M. Robins, "Causal inference from complex longitudinal data," in *Latent Variable Modeling and Applications to Causality*, M. Berkane, Ed., New York, NY: Springer New York, 1997, pp. 69–117, ISBN: 978-1-4612-1842-5.

[14] D. B. Rubin, "Comment," *Journal of the American Statistical Association*, vol. 75, no. 371, pp. 591–593, Sep. 1980. DOI: 10.1080/01621459.1980.10477517. [Online]. Available: https://doi.org/10.1080/01621459.1980.10477517.

[15] P. W. Lavori and R. Dawson, "A design for testing clinical strategies: Biased adaptive within-subject randomization," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, vol. 163, no. 1, pp. 29–38, 2000. DOI: 10.1111/1467-985X.00154. eprint: https://rss.onlinelibrary.wiley.com/doi/pdf/10.1111/1467-985X.00154. [Online]. Available: https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/1467-985X.00154.

[16] Y. Liu, Y. Wang, M. Kosorok, Y. Zhao, and D. Zeng, "Robust hybrid learning for estimating personalized dynamic treatment regimens," Nov. 2016, arXiv:1611.02314v1.

[17] S. A. Murphy, "An experimental design for the development of adaptive treatment strategies," *Statistics in Medicine*, vol. 24, no. 10, pp. 1455–1481, 2005. DOI: 10.1002/sim.2022. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/sim.2022. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.2022.

[18] R. Sutton and A. Barto, "Reinforcement learning: An introduction," *IEEE Transactions on Neural Networks*, vol. 9, no. 5, pp. 1054–1054, Sep. 1998. DOI: 10.1109/tnn.1998.712192. [Online]. Available: https://doi.org/10.1109/tnn.1998.712192.

[19] C. J. C. H. WATKINS, "Learning from delayed rewards," *PhD thesis, Cambridge University*, 1989. [Online]. Available: https://ci.nii.ac.jp/naid/10000025057/en/.

[20] C. J. C. H. Watkins and P. Dayan, "Q-learning," *Machine Learning*, vol. 8, no. 3, pp. 279–292, 1992, ISSN: 1573-0565. DOI: 10.1007/BF00992698. [Online]. Available: https://doi.org/10.1007/BF00992698.

[21] J. Pineau, M. G. Bellemare, A. J. Rush, A. Ghizaru, and S. A. Murphy, "Constructing evidence-based treatment strategies using methods from computer science," *Drug and Alcohol Dependence*, vol. 88, S52–S60, May 2007. DOI: 10.1016/j.drugalcdep.2007.01.005. [Online]. Available: https://doi.org/10.1016/j.drugalcdep.2007.01.005.

[22] Y. Zhao, M. R. Kosorok, and D. Zeng, "Reinforcement learning design for cancer clinical trials," *Statistics in Medicine*, vol. 28, no. 26, pp. 3294–3315, Nov. 2009. DOI: 10.1002/sim.3720. [Online]. Available: https://doi.org/10.1002/sim.3720.

[23] I. Nahum-Shani, M. Qian, D. Almirall, W. E. Pelham, B. Gnagy, G. A. Fabiano, J. G. Waxmonsky, J. Yu, and S. A. Murphy, "Q-learning: A data analysis method for constructing adaptive interventions.," *Psychological Methods*, vol. 17, no. 4, pp. 478–494, 2012. DOI: 10.1037/a0029373. [Online]. Available: https://doi.org/10.1037/a0029373.

[24] S. A. Murphy, "Optimal dynamic treatment regimes," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 65, no. 2, pp. 331–355, 2003. DOI: 10.1111/1467-9868.00389. eprint: https://rss.onlinelibrary.wiley.com/doi/pdf/10.1111/1467-9868.00389. [Online]. Available: https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/1467-9868.00389.

[25] D. Blatt, S. A. Murphy, and J. Zhu, "A-learning for approximate planning," The Methodology Center, Pennsylvania State University., 2004, pp. 04–63.

[26] J. M. Robins, "Optimal structural nested models for optimal sequential decisions," in *In Proceedings of the Second Seattle Symposium on Biostatistics*, Springer, 2004.

[27] R. Bellman, "Dynamic programming," *Science*, vol. 153, no. 3731, pp. 34–37, 1966, ISSN: 0036-8075. DOI: 10.1126/science.153.3731.34. eprint: https://science.sciencemag.org/content/153/3731/34.full.pdf. [Online]. Available: https://science.sciencemag.org/content/153/3731/34.

[28] S. A. Murphy, D. W. Oslin, A. J. Rush, and J. Zhu, "Methodological challenges in constructing effective treatment sequences for chronic psychiatric disorders," *Neuropsychopharmacology*, vol. 32, no. 2, pp. 257–262, Nov. 2006. DOI: 10.1038/sj.npp.1301241. [Online]. Available: https://doi.org/10.1038/sj.npp.1301241.

[29] J. M. Robins, "The analysis of randomized and nonrandomized aids treatment trials using a new approach to causal inference in longitudinal studies," in *Health Service Research Methodology: A Focus on AIDS*, L. Sechrest, H. Freeman, and A. Mulley, Eds., New York, NY: US Public Health Service, 1993, pp. 113–159.

[30] J. ROBINS, "Estimation of the time-dependent accelerated failure time model in the presence of confounding factors," *Biometrika*, vol. 79, no. 2, pp. 321–334, 1992. DOI: 10.1093/biomet/79.2.321. [Online]. Available: https://doi.org/10.1093/biomet/79.2.321.

[31] D. B. RUBIN, "Inference and missing data," *Biometrika*, vol. 63, no. 3, pp. 581–592, 1976. DOI: 10.1093/biomet/63.3.581. [Online]. Available: https://doi.org/10.1093/biomet/63.3.581.

[32] J. M. Robins, A. Rotnitzky, and L. P. Zhao, "Estimation of regression coefficients when some regressors are not always observed," *Journal of the American Statistical Association*, vol. 89, no. 427, pp. 846–866, Sep. 1994. DOI: 10.1080/01621459.1994.10476818. [Online]. Available: https://doi.org/10.1080/01621459.1994.10476818.

[33] J. M. Robins, M. Á. Hernán, and B. Brumback, "Marginal structural models and causal inference in epidemiology," *Epidemiology*, vol. 11, no. 5, pp. 550–560, Sep. 2000. DOI: 10.1097/00001648-200009000-00011. [Online]. Available: https://doi.org/10.1097/00001648-200009000-00011.

[34] J. M. Robins, "Marginal structural models versus structural nested models as tools for causal inference," in *Statistical Models in Epidemiology, the Environment, and Clinical Trials*, Springer New York, 2000, pp. 95–133. DOI: 10.1007/978-1-4612-1284-3_2. [Online]. Available: https://doi.org/10.1007/978-1-4612-1284-3_2.

[35] M. Á. Hernán, B. Brumback, and J. M. Robins, "Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men," *Epidemiology*, vol. 11, no. 5, pp. 561–570, Sep. 2000. DOI: 10.1097/00001648-200009000-00012. [Online]. Available: https://doi.org/10.1097/00001648-200009000-00012.

[36] S. A. Murphy, M. J. van der Laan, and J. M. R. and, "Marginal mean models for dynamic regimes," *Journal of the American Statistical Association*, vol. 96, no. 456, pp. 1410–1423, Dec. 2001. DOI: 10.1198/016214501753382327. [Online]. Available: https://doi.org/10.1198/016214501753382327.

[37] M. J. van der Laan and D. Rubin, "Targeted maximum likelihood learning," *The International Journal of Biostatistics*, vol. 2, no. 1, Jan. 2006. DOI: 10.2202/1557-4679.1043. [Online]. Available: https://doi.org/10.2202/1557-4679.1043.

[38] M. A. Hernan, E. Lanoy, D. Costagliola, and J. M. Robins, "Comparison of dynamic treatment regimes via inverse probability weighting," *Basic Clinical Pharmacology Toxicology*, vol. 98, no. 3, pp. 237–242, Mar. 2006. DOI: 10.1111/j.1742-7843.2006.pto_329.x. [Online]. Available: https://doi.org/10.1111/j.1742-7843.2006.pto_329.x.

[39] M. J. van der Laan and M. L. Petersen, "Causal effect models for realistic individualized treatment and intention to treat rules," *The International Journal of Biostatistics*, vol. 3, no. 1, Jan. 2007. DOI: 10.2202/1557-4679.1022. [Online]. Available: https://doi.org/10.2202/1557-4679.1022.

[40] M. L. Petersen, S. G. Deeks, and M. J. van der Laan, "Individualized treatment rules: Generating candidate clinical trials," *Statistics in Medicine*, vol. 26, no. 25, pp. 4578–4601, 2007. DOI: 10.1002/sim.2888. [Online]. Available: https://doi.org/10.1002/sim.2888.

[41] C. A. Cotton and P. J. Heagerty, "A data augmentation method for estimating the causal effect of adherence to treatment regimens targeting control of an intermediate measure," *Statistics in Biosciences*, vol. 3, no. 1, pp. 28–44, Jul. 2011. DOI: 10.1007/s12561-011-9038-1. [Online]. Available: https://doi.org/10.1007/s12561-011-9038-1.

[42] L. Orellana, A. Rotnitzky, and J. M. Robins, "Dynamic regime marginal structural mean models for estimation of optimal dynamic treatment regimes, part i: Main content," *The International Journal of Biostatistics*, vol. 6, no. 2, Jan. 2010. DOI: 10.2202/1557-4679.1200. [Online]. Available: https://doi.org/10.2202/1557-4679.1200.

[43] M. Qian and S. A. Murphy, "Performance guarantees for individualized treatment rules," *Ann. Statist.*, vol. 39, no. 2, pp. 1180–1210, Apr. 2011. DOI: 10.1214/10-AOS864. [Online]. Available: https://doi.org/10.1214/10-AOS864.

[44] Y. Zhao, D. Zeng, A. J. Rush, and M. R. Kosorok, "Estimating individualized treatment rules using outcome weighted learning," *Journal of the American Statistical Association*, vol. 107, no. 499, pp. 1106–1118, 2012, ISSN: 01621459. [Online]. Available: http://www.jstor.org/stable/23427417.

[45] C. Cortes and V. Vapnik, *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995. DOI: 10.1023/a:1022627411411. [Online]. Available: https://doi.org/10.1023/a:1022627411411.

[46] B. Zhang, A. A. Tsiatis, M. Davidian, M. Zhang, and E. Laber, "Estimating optimal treatment regimes from a classification perspective," *Stat*, vol. 1, no. 1, pp. 103–114, Oct. 2012. DOI: 10.1002/sta.411. [Online]. Available: https://doi.org/10.1002/sta.411.

[47] Y.-Q. Zhao, D. Zeng, E. B. Laber, and M. R. Kosorok, "New statistical learning methods for estimating optimal dynamic treatment regimes," *Journal of the American Statistical Association*, vol. 110, no. 510, pp. 583–598, 2015, PMID: 26236062. DOI: 10.1080/01621459.2014.937488. eprint: https://doi.org/10.1080/01621459.2014.937488. [Online]. Available: https://doi.org/10.1080/01621459.2014.937488.

[48] P. F. Thall, R. E. Millikan, and H.-G. Sung, "Evaluating multiple treatment courses in clinical trials," *Statistics in Medicine*, vol. 19, no. 8, pp. 1011–1028, 2000. DOI: 10.1002/(SICI)1097-0258(20000430)19:8<1011::AID-SIM414>3.0.CO;2-M. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/%28SICI%291097-0258%2820000430%2919%3A8%3C1011%3A%3AAID-SIM414%3E3.0.CO%3B2-M. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/%28SICI%291097-0258%2820000430%2919%3A8%3C1011%3A%3AAID-SIM414%3E3.0.CO%3B2-M.

[49] P. F. Thall, H.-G. Sung, and E. H. Estey, "Selecting therapeutic strategies based on efficacy and death in multicourse clinical trials," *Journal of the American Statistical Association*, vol. 97, no. 457, pp. 29–39, Mar. 2002. DOI: 10.1198/016214502753479202. [Online]. Available: https://doi.org/10.1198/016214502753479202.

[50] R. Wang, S. W. Lagakos, J. H. Ware, D. J. Hunter, and J. M. Drazen, "Statistics in medicine — reporting of subgroup analyses in clinical trials," *New England Journal of Medicine*, vol. 357, no. 21, pp. 2189–2194, 2007, PMID: 18032770. DOI: 10.1056/NEJMsr077003. eprint: https://doi.org/10.1056/NEJMsr077003. [Online]. Available: https://doi.org/10.1056/NEJMsr077003.

[51] Z. Shahn and D. Madigan, "Latent class mixture models of treatment effect heterogeneity," *Bayesian Anal.*, vol. 12, no. 3, pp. 831–854, Sep. 2017. DOI: 10.1214/16-BA1022. [Online]. Available: https://doi.org/10.1214/16-BA1022.

[52] D. P. Byar, "Assessing apparent treatment—covariate interactions in randomized clinical trials," *Statistics in Medicine*, vol. 4, no. 3, pp. 255–263, Jul. 1985. DOI: 10.1002/sim.4780040304. [Online]. Available: https://doi.org/10.1002/sim.4780040304.

[53] L. Gunter, J. Zhu, and S. Murphy, "Variable selection for optimal decision making," in *Artificial Intelligence in Medicine*, R. Bellazzi, A. Abu-Hanna, and J. Hunter, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 149–154, ISBN: 978-3-540-73599-1.

[54] T. Cai, L. Tian, P. H. Wong, and L. J. Wei, "Analysis of randomized comparative clinical trial data for personalized treatment selections," *Biostatistics*, vol. 12, no. 2, pp. 270–282, Sep. 2010, ISSN: 1465-4644. DOI: 10.1093/biostatistics/kxq060. eprint: https://academic.oup.com/biostatistics/article-pdf/12/2/270/18608731/kxq060.pdf. [Online]. Available: https://doi.org/10.1093/biostatistics/kxq060.

[55] W. Lu, H. H. Zhang, and D. Zeng, "Variable selection for optimal treatment decision," *Statistical Methods in Medical Research*, vol. 22, no. 5, pp. 493–504, Nov. 2011. DOI: 10.1177/0962280211428383. [Online]. Available: https://doi.org/10.1177/0962280211428383.

[56] X. Su, C.-L. Tsai, H. Wang, D. M. Nickerson, and B. Li, "Subgroup analysis via recursive partitioning," *SSRN Electronic Journal*, 2009. DOI: 10.2139/ssrn.1341380. [Online]. Available: https://doi.org/10.2139/ssrn.1341380.

[57] J. C. Foster, J. M. Taylor, and S. J. Ruberg, "Subgroup identification from randomized clinical trial data," *Statistics in Medicine*, vol. 30, no. 24, pp. 2867–2880, Aug. 2011. DOI: 10.1002/sim.4322. [Online]. Available: https://doi.org/10.1002/sim.4322.

[58] J. Kang, X. Su, B. Hitsman, K. Liu, and D. Lloyd-Jones, "Tree-structured analysis of treatment effects with large observational data," *Journal of Applied Statistics*, vol. 39, no. 3, pp. 513–529, Mar. 2012. DOI: 10.1080/02664763.2011.602056. [Online]. Available: https://doi.org/10.1080/02664763.2011.602056.

[59] C. Kang, H. Janes, and Y. Huang, "Combining biomarkers to optimize patient treatment recommendations," *Biometrics*, vol. 70, no. 3, pp. 695–707, May 2014. DOI: 10.1111/biom.12191. [Online]. Available: https://doi.org/10.1111/biom.12191.

[60] T. Stiehl, C. Lutz, and A. Marciniak-Czochra, "Emergence of heterogeneity in acute leukemias," *Biology Direct*, vol. 11, no. 1, p. 51, 2016, ISSN: 1745-6150. DOI: 10.1186/s13062-016-0154-1. [Online]. Available: https://doi.org/10.1186/s13062-016-0154-1.

[61] P. Wang, M. L. Puterman, I. Cockburn, and N. Le, "Mixed poisson regression models with covariate dependent rates," *Biometrics*, vol. 52, no. 2, p. 381, Jun. 1996. DOI: 10.2307/2532881. [Online]. Available: https://doi.org/10.2307/2532881.

[62] P. Wang and M. L. Puterman, "Mixed logistic regression models," *Journal of Agricultural, Biological, and Environmental Statistics*, vol. 3, no. 2, p. 175, Jun. 1998. DOI: 10.2307/1400650. [Online]. Available: https://doi.org/10.2307/1400650.

[63] C. S. Wong, "On a logistic mixture autoregressive model," *Biometrika*, vol. 88, no. 3, pp. 833–846, Oct. 2001. DOI: 10.1093/biomet/88.3.833. [Online]. Available: https://doi.org/10.1093/biomet/88.3.833.

[64] M. E. Sobel and B. Muthén, "Compliance mixture modelling with a zero-effect complier class and missing data," *Biometrics*, vol. 68, no. 4, pp. 1037–1045, 2012. DOI: 10.1111/j.1541-0420.2012.01791.x. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1541-0420.2012.01791.x. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1541-0420.2012.01791.x.

[65] J. Shen and X. He, "Inference for subgroup analysis with a structured logistic-normal mixture model," *Journal of the American Statistical Association*, vol. 110, no. 509, pp. 303–312, 2015. DOI: 10.1080/01621459.2014.894763. eprint: https://doi.org/10.1080/01621459.2014.894763. [Online]. Available: https://doi.org/10.1080/01621459.2014.894763.

[66] J. Shen and A. Qu, "Subgroup analysis based on structured mixed-effects models for longitudinal data," *Journal of Biopharmaceutical Statistics*, vol. 30, no. 4, pp. 607–622, Mar. 2020. DOI: 10.1080/10543406.2020.1730867. [Online]. Available: https://doi.org/10.1080/10543406.2020.1730867.

[67] N. M. Laird and J. H. Ware, "Random-effects models for longitudinal data," *Biometrics*, vol. 38, no. 4, p. 963, Dec. 1982. DOI: 10.2307/2529876. [Online]. Available: https://doi.org/10.2307/2529876.

[68] L. Wang, A. Rotnitzky, X. Lin, R. E. Millikan, and P. F. Thall, "Evaluation of viable dynamic treatment regimes in a sequentially randomized trial of advanced prostate cancer," *Journal of the American Statistical Association*, vol. 107, no. 498, pp. 493–508, 2012, PMID: 22956855. DOI: 10.1080/01621459.2011.641416. eprint: https://doi.org/10.1080/01621459.2011.641416. [Online]. Available: https://doi.org/10.1080/01621459.2011.641416.

[69] X. Huang, S. Choi, L. Wang, and P. F. Thall, "Optimization of multi-stage dynamic treatment regimes utilizing accumulated data," *Statistics in Medicine*, vol. 34, no. 26, pp. 3424–3443, Jun. 2015. DOI: 10.1002/sim.6558. [Online]. Available: https://doi.org/10.1002/sim.6558.

[70] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via theEMAlgorithm," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 39, no. 1, pp. 1–22, Sep. 1977. DOI: 10.1111/j.2517-6161.1977.tb01600.x. [Online]. Available: https://doi.org/10.1111/j.2517-6161.1977.tb01600.x.

[71] J. Pan and R. Thompson, "Gauss-hermite quadrature approximation for estimation in generalised linear mixed models," *Computational Statistics*, vol. 18, no. 1, pp. 57–78, Mar. 2003. DOI: 10.1007/s001800300132. [Online]. Available: https://doi.org/10.1007/s001800300132.

[72] D. C. Liu and J. Nocedal, "On the limited memory BFGS method for large scale optimization," *Mathematical Programming*, vol. 45, no. 1-3, pp. 503–528, Aug. 1989. DOI: 10.1007/bf01589116. [Online]. Available: https://doi.org/10.1007/bf01589116.

[73] B. Fornberg and D. M. Sloan, "A review of pseudospectral methods for solving partial differential equations," *Acta Numerica*, vol. 3, pp. 203–267, Jan. 1994. DOI: 10.1017/s0962492900002440. [Online]. Available: https://doi.org/10.1017/s0962492900002440.

[74] T. J. Rothenberg, "Identification in parametric models," *Econometrica*, vol. 39, no. 3, pp. 577–591, 1971, ISSN: 00129682, 14680262. [Online]. Available: http://www.jstor.org/stable/1913267.

[75] "Finite mixture modeling," in *Finite Mixture and Markov Switching Models*. New York, NY: Springer New York, 2006, ISBN: 978-0-387-35768-3. DOI: 10.1007/978-0-387-35768-3_1. [Online]. Available: https://doi.org/10.1007/978-0-387-35768-3_1.

[76] R. B. McHugh, "Efficient estimation and local identification in latent class analysis," *Psychometrika*, vol. 21, no. 4, pp. 331–347, 1956, ISSN: 1860-0980. DOI: 10.1007/BF02296300. [Online]. Available: https://doi.org/10.1007/BF02296300.

[77] L. A. Goodman, "Exploratory latent structure analysis using both identifiable and unidentifiable models," *Biometrika*, vol. 61, no. 2, pp. 215–231, 1974, ISSN: 00063444. [Online]. Available: http://www.jstor.org/stable/2334349.

[78] J. L. Teugels, "Some representations of the multivariate bernoulli and binomial distributions," *Journal of Multivariate Analysis*, vol. 32, no. 2, pp. 256–268, Feb. 1990. DOI: 10.1016/0047-259x(90)90084-u. [Online]. Available: https://doi.org/10.1016/0047-259x(90)90084-u.

[79] B. Dai, S. Ding, and G. Wahba, "Multivariate bernoulli distribution," *Bernoulli*, vol. 19, no. 4, pp. 1465–1483, Sep. 2013. DOI: 10.3150/12-bejsp10. [Online]. Available: https://doi.org/10.3150/12-bejsp10.

[80] C. G. BROYDEN, "The convergence of a class of double-rank minimization algorithms 1. general considerations," *IMA Journal of Applied Mathematics*, vol. 6, no. 1, pp. 76–90, 1970. DOI: 10.1093/imamat/6.1.76. [Online]. Available: https://doi.org/10.1093/imamat/6.1.76.

[81] D. Goldfarb, "A family of variable-metric methods derived by variational means," *Mathematics of Computation*, vol. 24, no. 109, pp. 23–23, Jan. 1970. DOI: 10.1090/s0025-5718-1970-0258249-6. [Online]. Available: https://doi.org/10.1090/s0025-5718-1970-0258249-6.

[82] R. Fletcher, "A new approach to variable metric algorithms," *The Computer Journal*, vol. 13, no. 3, pp. 317–322, Mar. 1970. DOI: 10.1093/comjnl/13.3.317. [Online]. Available: https://doi.org/10.1093/comjnl/13.3.317.

[83] D. F. Shanno, "Conditioning of quasi-newton methods for function minimization," *Mathematics of Computation*, vol. 24, no. 111, pp. 647–647, Sep. 1970. DOI: 10.1090/s0025-5718-1970-0274029-x. [Online]. Available: https://doi.org/10.1090/s0025-5718-1970-0274029-x.

[84] K. E. Train, *Discrete Choice Methods with Simulation*. Cambridge University Press, Jan. 2001. DOI: 10.1017/cbo9780511805271. [Online]. Available: https://doi.org/10.1017/cbo9780511805271.

[85] P. K. Dunn and G. K. Smyth, "Randomized quantile residuals," *Journal of Computational and Graphical Statistics*, vol. 5, no. 3, p. 236, Sep. 1996. DOI: 10.2307/1390802. [Online]. Available: https://doi.org/10.2307/1390802.

[86] D. W. Hosmer, S. Lemeshow, and R. X. Sturdivant, *Applied Logistic Regression*. John Wiley & Sons, Inc., Mar. 2013. DOI: 10.1002/9781118548387. [Online]. Available: https://doi.org/10.1002/9781118548387.

[87] S. X. Lee, K. L. Leemaqz, and G. J. McLachlan, "A simple parallel EM algorithm for statistical learning via mixture models," in *2016 International Conference on Digital Image Computing: Techniques and Applications (DICTA)*, IEEE, Nov. 2016. DOI: 10.1109/dicta.2016.7796997. [Online]. Available: https://doi.org/10.1109/dicta.2016.7796997.

[88] D. Raghavarao and L. Padgett, *Repeated Measurements and Cross-Over Designs*. John Wiley & Sons, Inc, Apr. 2014. DOI: 10.1002/9781118709153. [Online]. Available: https://doi.org/10.1002/9781118709153.

[89] A. Genz and B. Keister, "Fully symmetric interpolatory rules for multiple integrals over infinite regions with gaussian weight," *Journal of Computational and Applied Mathematics*, vol. 71, no. 2, pp. 299–309, Jul. 1996. DOI: 10.1016/0377-0427(95)00232-4. [Online]. Available: https://doi.org/10.1016/0377-0427(95)00232-4.

[90] C. Proust and H. Jacqmin-Gadda, "Estimation of linear mixed models with a mixture of distribution for the random effects," *Computer Methods and Programs in Biomedicine*, vol. 78, no. 2, pp. 165–173, May 2005. DOI: 10.1016/j.cmpb.2004.12.004. [Online]. Available: https://doi.org/10.1016/j.cmpb.2004.12.004.

[91] C. Proust-Lima, V. Philipps, and B. Liquet, "Estimation of extended mixed models using latent classes and latent processes: The r package lcmm," *Journal of Statistical Software*, vol. 78, no. 2, 2017. DOI: 10.18637/jss.v078.i02. [Online]. Available: https://doi.org/10.18637/jss.v078.i02.

[92] D. W. Marquardt, "An algorithm for least-squares estimation of nonlinear parameters," *Journal of the Society for Industrial and Applied Mathematics*, vol. 11, no. 2, pp. 431–441, Jun. 1963. DOI: 10.1137/0111030. [Online]. Available: https://doi.org/10.1137/0111030.

A. FIGURES

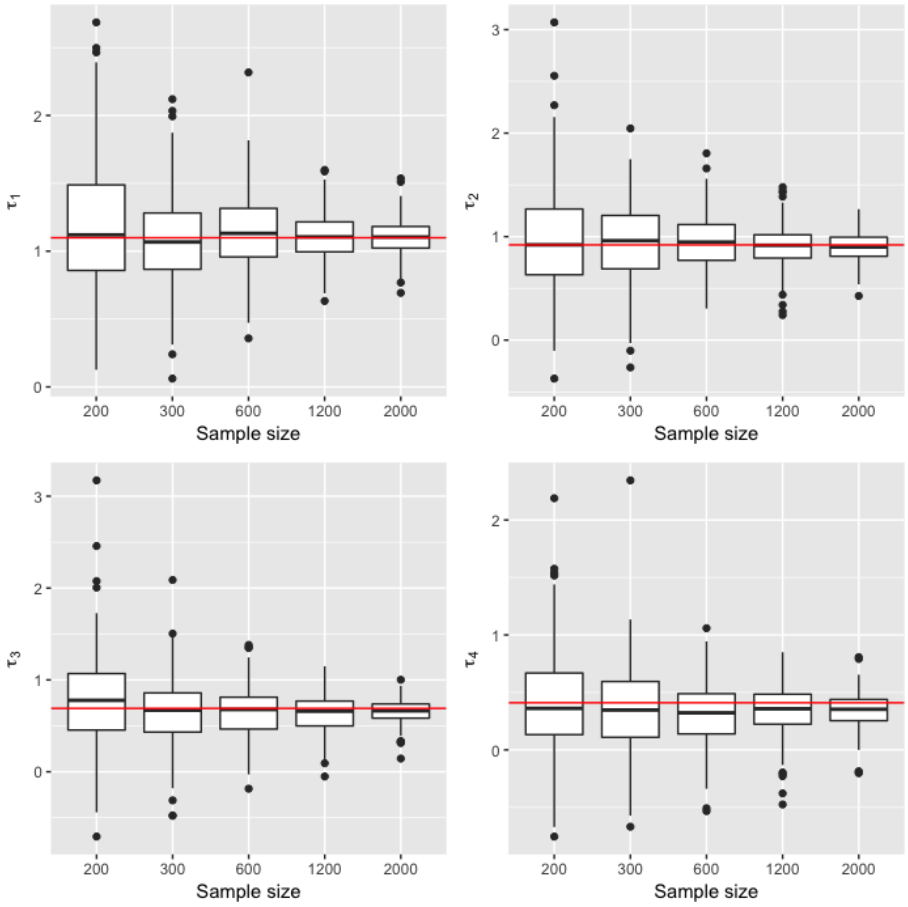

Figure A.1. Boxplots of estimated treatment effects of setting 1. The red line represents the true value.

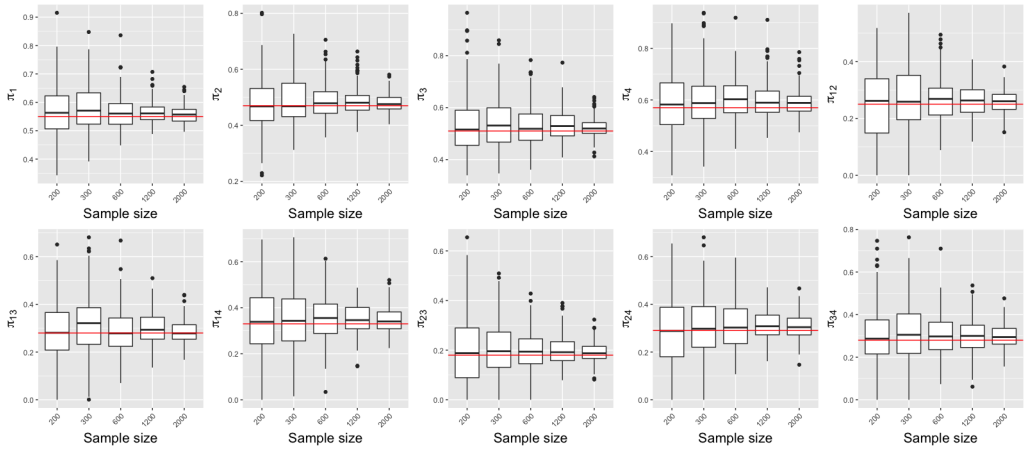

Figure A.2. Boxplots of estimated subgroup proportions of setting 1. The red line represents the true value.

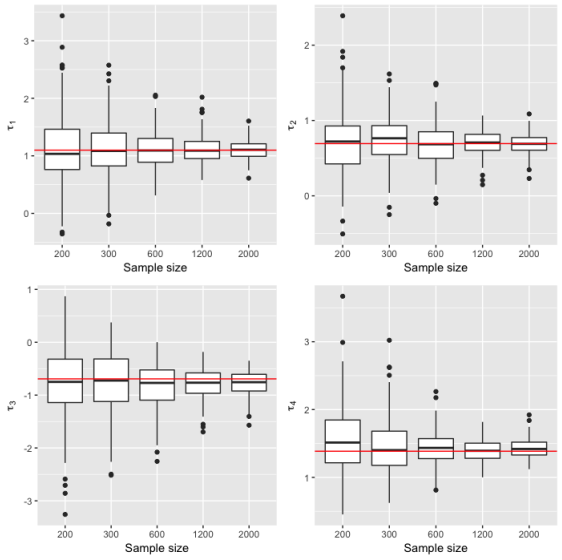

Figure A.3. Boxplots of estimated treatment effects of setting 2. The red line represents the true value.

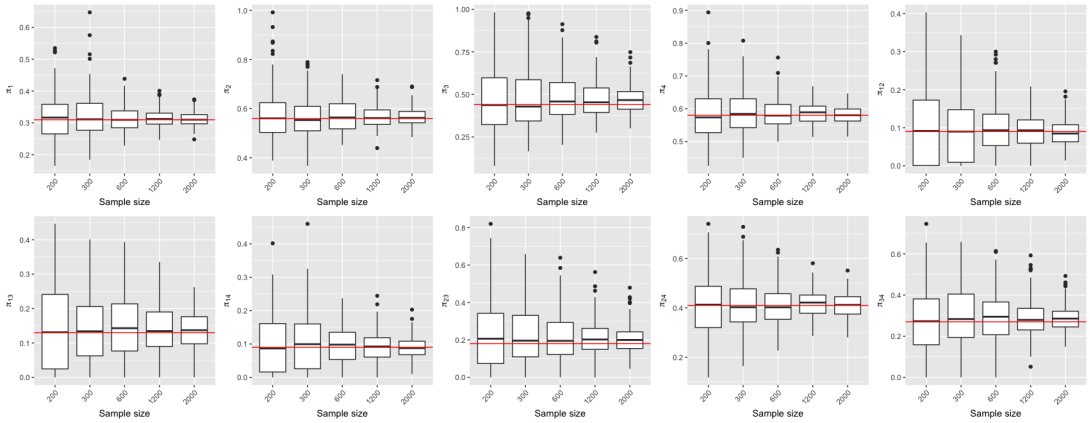

Figure A.4. Boxplots of estimated subgroup proportions of setting 2. The red line represents the true value.

B. TABLES

Table B.1. Means and standard deviations of parameter estimates for 4-treatment independence model when N = 200, 600, and 1200

True value	$N = 200$		$N = 600$		$N = 1200$	
	Mean	Std.Dev.	Mean	Std.Dev.	Mean	Std.Dev.
$\gamma_{1,0}$= 0.3000	0.7721	1.4998	0.4312	0.3790	0.3535	0.2170
$\gamma_{1,1}$= 1.5000	2.1696	1.6928	1.6167	0.3883	1.5772	0.2673
$\gamma_{2,0}$= 1.0000	2.0438	2.0994	1.3162	0.7815	1.0658	0.4582
$\gamma_{2,1}$= -3.0000	-4.9298	3.4035	-3.4768	1.1635	-3.1306	0.6473
$\gamma_{3,0}$= -1.0000	-1.0178	0.4423	-0.9539	0.2350	-0.9805	0.1637
$\gamma_{3,1}$= 1.0000	1.1812	0.5451	1.0088	0.2359	0.9982	0.1758
$\gamma_{4,0}$= 0.5000	0.8590	1.2793	0.5792	0.3746	0.5619	0.2121
$\gamma_{4,1}$= -1.0000	-1.4278	1.3171	-1.0777	0.3602	-1.0469	0.1975
τ_1 = 0.8000	0.8045	0.4349	0.8019	0.2269	0.7805	0.1481
β_1= -0.5000	-0.5694	0.3781	-0.5255	0.1765	-0.5254	0.1334
τ_2 = 0.3000	0.2697	0.3372	0.2849	0.1935	0.2979	0.1223
β_2= 0.6000	0.6495	0.2931	0.6440	0.1695	0.6160	0.1162
τ_3= 1.1000	1.8046	6.8807	1.1395	0.3540	1.1077	0.2426
β_3= 0.4000	0.4560	2.7781	0.4493	0.3036	0.4381	0.2023
τ_4 = 0.6000	0.6462	0.4639	0.6000	0.2277	0.5724	0.1523
β_4= 1.0000	1.1634	0.5538	1.0806	0.2317	1.0095	0.1533
σ= -2.3026	0.1949	0.3156	0.2569	0.3192	0.2020	0.2731