Evaluation and Comparison of Dynamic Treatment Regimes: Methods and Challenges

by

Xi Lu

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Statistics)
in the University of Michigan
2015

Doctoral Committee:

Professor Susan A. Murphy, Chair
Research Assistant Professor Daniel Almirall
Assistant Professor Lu Wang
Assistant Professor Shuheng Zhou

2015

Dedication

To my parents

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Evaluation and Comparison of Dynamic Treatment Regimes: Methods and Challenges

by

Xi Lu

Chair: Professor Susan A. Murphy

Dynamic treatment regimes (DTRs) are sequences of decision rules that link the patient history with treatment recommendations. Clinical scientists have become increasingly interested in the development of DTRs in various fields including substance abuse, mental health and cancer. The Sequential Multiple Assignment Randomized Trial (SMART) is a multi-stage trial design that explicitly targets the development of high-quality DTRs. In this dissertation, we develop statistical methodologies, which can be applied to SMART data, that either address novel research questions regarding the construction of a high-quality DTR, or exhibit better performance than existing statistical methods.

CHAPTER 1

Introduction

In many areas of health, treatment response is heterogeneous in which case clinicians will need to consider providing a sequence of treatments in order to obtain sufficient treatment response. Furthermore patients with chronic illnesses often require changes in treatment, that is, sequences of treatments, so as to maintain a good response. As a result clinical scientists have become increasingly interested in, and active in, the development of interventions that are composed of treatment sequences [25] in various fields including alcoholism [48], substance abuse [18, 35], leukemia [75] and autism spectrum disorder [19]. The treatment sequences are adapted to the dynamics of the evolving illness. The idea is that the adaptation should accommodate treatment response heterogeneity so as to result in more efficacious and less burdensome/costly treatment. Treatment policies [33, 82, 83] – also called dynamic treatment regimes (DTRs) [51, 56, 57, 58, 64, 41], adaptive treatment strategies [25, 27, 26, 40, 74, 75] or adaptive interventions [43, 44, 1] – operationalize the dynamic adaption via a sequence of decision rules, one for each stage in the treatment process; the decision rule inputs measurements of patients' time-varying covariates and outputs recommended treatments.

The Sequential Multiple Assignment Randomized Trial (SMART; [25, 40, 9]), a multi-stage trial design, was developed explicitly for the development of high-quality DTRs. Specifically, data from SMART design is useful in addressing key research questions that inform the construction of a high-quality DTR. Each stage in a SMART corresponds to one of the critical decisions involved in the DTR. Each participant moves through the multiple stages and at each stage the participant is (re)randomized to one of several intervention options. A variety of these trials have been conducted, with some of the earliest taking place in cancer research, for the purpose of developing medication algorithms for leukemia [75], or to develop adaptive treatments of prostate cancer [74]. A selection of SMART studies may be found at `http://methodology.psu.edu/ra/adap-inter/projects`. Common research questions regarding DTRs that can be addressed by analyzing SMART data include: (a) the comparison of different intervention options at each of multiple stages of

the intervention; (b) the comparison among a pre-determined set of DTRs (usually "embedded" in the design of a SMART, which we later explain in Chapter 3) in terms of an end-of-study primary outcome.

In the following chapters in this dissertation, we will develop statistical methodologies that either address novel research questions regarding the construction of a high-quality DTR, or exhibit better performance than existing methods/statistical procedures. The topics that are discussed in this dissertation will cover a variety of aspects in the analysis of data arising from SMARTs, or more generally, randomized clinical trials. It is worthwhile to note that there are also extensive works concerning the design of (multi-stage) randomized clinical trials, for the purpose of optimizing various objectives. Those topics are not in the scope of this dissertation.

The dissertation is organized as follows. In the remainder of Chapter 1, we review the literature of methodological works related to evaluating and optimizing DTRs. In Chapter 2, we develop an "assisted estimator" that can be used to compare the mean outcomes of a pair of competing DTRs. The term "assisted" refers to the fact that estimators from the Structural Nested Mean Model (SNMM), a parametric model for the causal effect of treatment at each time point, are used in the process of estimating the mean outcome. This novel estimator significantly improves efficiency compared to the existing inverse-probability-of-treatment-weighted type of methods, by imposing parametric modeling assumptions on the components of the data distribution that are easily interpretable. Additionally, based on Robins' G-estimators for the SNMM, we present an easy-to-implement least-squares estimator for the parameters in the SNMM.

In Chapter 3, we focus on the comparison of a pre-determined set of DTRs, in terms of a repeated-measures outcome that spans across multiple treatment stages. Modeling the marginal mean trajectories of a repeated-measures outcome arising from a SMART presents challenges, because traditional longitudinal models used for randomized clinical trials do not take into account the unique design features of SMART. In this chapter, we fill in this gap by discussing modeling considerations for various forms of SMART designs, emphasizing the importance of considering the timing of the repeated measures in relation to the treatment stages in a SMART. For illustration, we present three case studies with increasing level of complexity, in autism, child attention deficit hyperactivity disorder (ADHD), and adult alcoholism. The weighted-and-replicated estimators, which were originally proposed for comparing DTRs in terms of an end-of-study outcome, are generalized to estimate the parameters in our repeated-measures model.

In Chapter 4, we concentrate on one particular aspect of the weighted-and-replicated (WR) estimators, namely the performance of the WR estimators on data sets with small

sample sizes. More specifically, in some numerical studies we found that the sandwich estimator for the variance of WR estimators derived from the standard Taylor series arguments does not provide confidence intervals that have good coverage, when the sample size is sufficiently small. The same phenomenon has been discovered in the GEE literature; intuitively, this happens because using the estimated parameters as a surrogate for the true unknown values of the parameters in general tend to "underestimate the variance of the true errors". Based on [34] in the GEE literature, we propose a small-sample adjusted estimator for the variance of WR estimators. The adjustment is developed for the WR estimators with both known weights (due to the SMART design) and estimated weights.

In Chapter 5, we consider a novel research question regarding the search for the optimal decision rule. Primarily the goal is to identify the optimal policy, i.e., the one that yields the highest mean of an outcome variable, within a pre-specified class of parametrized policies. On top of this goal, we are interested in understanding the usefulness of a particular variable in decision making, i.e., whether using this variable in addition to all the other variables in the specified policy form to construct a policy would remarkably increase the optimal achievable policy value. It turns out that estimating the optimal policy by simply searching for the policy associated to the highest (non-parametrically) estimated policy value does not answer the second part of our research question, due to some interesting ill-posedness issues. In this chapter some preliminary endeavor is made towards this research question. We propose a regularized estimator for the optimal policy, with two components of regularization motivated by two issues of the original unregularized estimator.

1.1 Review of Existing Work on Dynamic Treatment Regime Methodologies

Here we give a brief review of the literature on DTRs, mostly using data arising from an experiment study such as SMART.

A vast literature is available concerning the estimation of the optimal DTR based on data collected from both SMARTs and observational studies. A DTR is considered to be optimal if it yields the highest value of mean outcome when the entire population receive treatment sequences that are specified by this DTR. Q-learning [72, 42, 69, 37] has been the most well studied methodology in this direction. A backward induction procedure mimicking dynamic programming is implemented, estimating the Q-function at each time point, which is the conditional mean of the primary outcome given certain values of current history of covariates and treatments, assuming that the optimal treatment is always

assigned in each of subsequent stages. The optimal DTR can then be estimated to consist of decision rules that recommend the treatment that maximizes the estimated Q-function at each stage. However, unbiased estimates and valid inference about the estimated optimal DTR can be difficult, because the "max" operator used in the Q-learning procedure can cause non-regularity. Some variations of Q-learning have been proposed, including the use of thresholding [8, 38] and the combination with LASSO [71]. There are also works that aim to directly draw valid inference from Q-learning based on bootstrap or m-out-of-n bootstrap techniques [23, 7].

On the other hand, the optimality, or good performance, of the estimated optimal DTR relies severely on the correct model specification of the Q-functions. More specifically, both the main effects of the covariates and the treatment interaction effects in each of the Q-functions have to be correctly specified to guarantee the optimality of the derived optimal DTR. This is a rather strong modeling assumption; mis-specification of the Q-functions may potentially lead to low mean value of the estimated optimal DTR. Advantage learning (A-learning; [39, 64]) is an alternative approach to estimating the optimal DTR. Unlike Q-learning, in A-learning only the part of outcome regression model that represents the contrasts among the treatments is parametrically modeled; this makes A-learning in general more robust than Q-learning to model mis-specification.

Another line of research that targets the estimation of the optimal DTR is the statistical learning based methods developed by [89] and [90]; the former focuses on one time point and the latter extends the methodology for single time point to sequential treatments. The approaches proposed in these works cast the optimization of the value under the DTRs as weighted classification problems, where weights depend on the outcomes; as a consequence, existing machine learning algorithms can be directly applied to achieve the search for the optimal DTR. In particular, in these works the authors adopt the support vector machine (SVM) algorithm to relax the weighted classification problem; therefore, the theoretical properties of the proposed methods naturally follow from the established theory of SVM.

Marginal mean model [41] is a model for the marginal mean of a primary outcome under a DTR, conditional on some baseline covariates. This methodology is essentially non-parametric in that it does not make parametric assumptions on the relationship between time-varying covariates and the outcome; consistency only relies on correct specification of the treatment assignment probability in the observed data, which helps to connect the mean in the hypothetical population where all individuals follow the specified DTR, to a weighted mean in the observed population. Such a model is estimated by a doubly robust inverse probability weighted estimator that contains some working models for a series of

conditional means to provide additional guarantee of robustness and potential efficiency improvement. Since the treatment assignment probability is known in SMARTs, the estimator based on the marginal mean model can easily be consistent. [87] provides another perspective of the estimator for marginal mean model that arises from coarsening of the sequential data, and has some insights about how the users might obtain reasonably good working estimates for the nuisance functions.

G-computation estimators [51] are another class of estimators that can be used to estimate the marginal mean of the outcome under a DTR. This class of estimators is based on the representation of the marginal mean with conditional mean of the outcome given time-varying covariates, and the conditional distribution of the time-varying covariates. In the G-computation estimator, these conditional means and conditional distributions are replaced by their estimates, respectively. This approach is conceptually intuitive; however, it requires correct model specification of many components in the data, which is particularly difficult when the covariates are of high dimension.

Marginal Structural Models (MSMs; [55], [63]) are a class of methods that were originally proposed to model the causal effect of time-varying treatment as a function of baseline prognostic factors. MSMs can be readily applied to handle various types of primary outcomes. Later the MSM methodology was extended to investigate the causal effect of DTRs conditional on baseline prognostic factors [47, 78, 50, 3, 46]. This is achieved by modeling the mean of potential outcomes associated with each of the DTRs in the class of DTRs of interest. By nature of the MSM methodology, the model adopted in MSM needs to be chosen according to the class of DTRs under study. Doubly robust inverse probability weighted estimating equation can be used to estimate such models. Then the optimal DTR among the specified class of DTRs can be readily estimated by identifying the optimizer of the estimated values of the DTRs.

Targeted maximum likelihood estimation (TMLE; [77, 6]) is another estimation procedure that can be taken to estimate a pre-specified parameter of the distribution of the observed data, such as the mean of an end-of-study outcome. The TMLE procedure targets a pre-specified estimand; more specifically, the procedure estimates the likelihood functions in a way that matches the efficient influence curve of the targeted parameter. The estimated likelihood functions are later used to construct the estimator for the targeted parameter via the G-computation formula. A TMLE is a substitution estimator, i.e., an estimator that can be conceptualized by replacing the unknown true underlying distribution with a particularly estimated distribution, in the defining formula of the estimand. Therefore it enjoys advantages that are specific to substitution estimators (e.g., the values of the estimator always lie in the reasonable range).

Methodological work that targets other types of outcomes is also available. Survival outcome (i.e., time to event outcome) is a particular type of outcome that usually requires special methodologies, and there have been a series of works about the comparison of DTRs regarding a survival outcome. [10] and [29] present weighted log-rank test statistic to compare a pair of DTRs that do not share observations (i.e., no participant can have treatment sequence that is consistent with these two DTRs at the same time). [21] develop weighted log-rank test statistic that can compare any pair of competing DTRs.

[85] proposes Bayesian inference methodology for the estimation and inference about DTRs. Under the Bayesian framework, potential outcomes under all possible treatment sequences are conceptualized as unknown parameters, and therefore posterior predictive distribution can be formed for the potential outcomes to facilitate estimation and inference about static and dynamic regimes. Moreover, the Bayesian approach naturally offers the potential to pool information across treatment paths and individuals in the same group/cluster.

CHAPTER 2

Comparing Treatment Policies with Assistance from the Structural Nested Mean Model

2.1 Introduction

In many health domains, a treatment sequence that is adapted to patients' evolving characteristics and past treatment history is needed. This is because the response to the same treatment/intervention can vary among patients with different baseline characteristics and time-varying health status. Moreover, a treatment that is associated with short-term success may not be preferable for controlling the disorder in the long term. One way to operationalize the adaptation of the sequence of treatment to patients' evolving status over time is via the treatment policies, which compose of a sequence of decision rules, one for each critical decision time point in the treatment process. At each critical decision time point (i.e., at each treatment stage), the decision rule takes the measurements of patients' time-varying covariates as input, and determines the recommended treatments/interventions.

Often scientists construct treatment policies that represent competing approaches to managing an illness. For example in the treatment of attention deficit hyperactivity disorder (ADHD), the American Psychological Association recommends starting with behavioral treatment and moving to a medication only if the behavioral treatment is not effective [4], whereas the American Academy of Child and Adolescent Psychiatry recommends starting with medication [49]. Or one treatment policy might represent a least intensive or least costly version, whereas another treatment policy may represent a most intensive, most costly version. For example, the Extending Treatment Effectiveness of Naltrexone (ExTENd) trial of alcohol dependence treatments (PI: Oslin; [48]) involves multiple treatment policies, of which one is the most intensive and another is the least intensive.

In this chapter, we develop and discuss statistical methodologies for the evaluation of a treatment policy and the comparison between two competing treatment policies, regarding the mean of a pre-specified primary outcome variable, that is either measured at the end

of the study, or an outcome variable calculated from the variables measured during the study. Some other endpoints for the comparison/evaluation of treatment policies might be possible. In particular, in the next chapter, we discuss the comparison among treatment policies regarding the mean trajectories of a repeated-measures outcome, the measurement of which spans through multiple treatment stages in a study.

A common approach to comparing the mean outcomes of two competing treatment policies, is to use a non-parametric estimation procedure that involves inverse-probability-weights (IPW), such as those described in [41] and [87]. These estimators are non-parametric in the sense that they do not require nor take advantage of models that relate baseline or time-varying covariates with the outcome. Robins and colleagues [54, 46] generalized the [41] methods to consider multiple treatment policies.

In this chapter, we develop an alternative approach for contrasting two treatment policies. This approach combines the non-parametric IPW estimators with a model-based approach based on Robins' Structural Nested Mean Model [52]. In the Structural Nested Mean Model, intermediate treatment effect functions, also called "treatment blips," are parametrically modeled. The intermediate treatment effects isolate the causal effect of treatment at each time point, conditional on baseline and time-varying covariate history up to that time point. We call the resulting estimator, an "assisted" estimator to convey that the model-based approach is intended to assist the non-parametric estimator in estimating the mean outcomes of competing treatment policies.

In this chapter we first focus on the comparison of two-stage treatment policies. Most sequentially randomized trials, also known as Sequential Multiple Assignment Randomized Trials (SMART) [24, 40], concern two stages of treatment. In particular, ExTENd is a two-stage SMART. Towards the end of this chapter we will briefly discuss the extension of the proposed methodology to the scenario of more than two treatment stages. In Section 2.2, we formulate the estimand in a precise manner. In this section we provide a class of assisted estimators for the mean outcome based on data from a SMART; theoretical properties of the estimators are also provided. In Section 2.3, we briefly introduce how these estimators can be used to compare a pair of treatment policies and make inference. Simulation studies, in Section 2.4, are used to investigate different aspects of the methodology, including the performance of the proposed estimator under various levels of mis-specifying treatment effects. In Section 2.5, the methodology is illustrated by an analysis of the ExTENd data. In Section 2.6, we briefly introduce some ideas about how the proposed estimator can be extended to apply to three-stage problems. Finally, a discussion of the paper, including ideas for future work, is presented in Section 2.7. Proofs of the theorems and lemmas are relegated to the appendix.

2.2 Assisted Estimator for Policy Value

A two-stage treatment policy consists of two decision rules, $d = (d_1, d_2)$. Each decision rule inputs available patient information at the current stage and outputs a treatment recommendation. Denote the outcome by Y (Y may be observed after the study or may be a function of the data collected during the study). The value of a policy is the expectation of Y that would result if the treatments were selected using the treatment policy d. A useful way to define the value of a policy is via the potential outcome framework [45, 68]. For each variable and each treatment sequence, we conceptualize a "potential outcome" that would have been observed under that treatment sequence. Using X_j to denote observations available prior to the j-th decision and using X_3 to denote observations available after the second-stage treatment, the potential outcomes are $\{X_1, X_2(a_1), X_3(a_1, a_2);$ for all possible sequences of treatments $(a_1, a_2)\}$. Here the outcome $Y(a_1, a_2)$ is a known function of $\{X_1, X_2(a_1), X_3(a_1, a_2)\}$. The value of the policy, d, is given by $V_d = E\left[Y(a_1, a_2)|_{a_2=d_2(H_2(a_1)),a_1=d_1(H_1)}\right]$ where $H_1 = X_1$ and $H_2(a_1) = (X_1, a_1, X_2(a_1))$ are the potential outcome history vectors prior to the treatments at stage one and stage two.

The value of a treatment policy d, can also be written as a function of the intermediate treatment effects or "treatment blip functions," from Robins' Structural Nested Mean Model [52]. We deviate briefly to define these intermediate treatment effects. Corresponding to the two stages of treatment, there are two intermediate treatment effects given by $\mu_2(h_2, a_2) = E[Y(a_1, a_2)|H_2(a_1) = h_2] - E[Y(a_1, 0)|H_2(a_1) = h_2]$ and $\mu_1(h_1, a_1) = E[Y(a_1, 0)|H_1 = h_1] - E[Y(0, 0)|H_1 = h_1]$, where $a_t = 0$ is the coding for a reference treatment (e.g., control treatment). The intermediate treatment effect, μ_2, quantifies the effect of treatment a_2 relative to the reference treatment at stage two on the mean of Y, among individuals with history h_2. The intermediate treatment effect, μ_1, quantifies the effect of treatment a_1 relative to the stage one reference treatment, if always followed by the reference treatment at stage two, on the mean of Y, among individuals with history h_1 at stage one. In addition to this type of treatment blip functions, there are other types of blips, such as the optimal-blip-to-zero functions for A-learning [39] and regime-specific SNMMs [64].

Consider randomized treatments, denoted by capitalized letters, A_1, A_2, where the conditional distribution of A_1 given $H_1 = h_1$ is denoted by $p_1(\cdot|h_1)$ and the conditional distribution of A_2 given $H_2(A_1) = h_2$ is denoted by $p_2(\cdot|h_2)$. Throughout this chapter we implicitly make all required measurability assumptions as well as existence of regular conditional densities. We have the following lemma.

Lemma 2.2.1. *Assume that (i)* $\max\{E|Y(a_1,a_2)|, E|\mu_1(H_1,a_1)|, E|\mu_2(H_2(a_1),a_2)|\} < \infty$ *for any treatment sequence* (a_1,a_2) *and (ii) for some* $\delta > 0$, $p_1(a_1|h_1) \geq \delta$ *a.s. for* (h_1,a_1)*, then*

$$\begin{aligned} V_d = \; & E\Big[Y(A_1,A_2) - \mu_2(H_2(A_1),A_2) - \mu_1(H_1,A_1) + \mu_1(H_1,d_1(H_1)) \\ & \qquad + \mu_2(H_2(a_1),d_2(H_2(a_1)))|_{a_1=d_1(H_1)}\Big] \\ = \; & E\Big[Y(A_1,A_2) - \mu_2(H_2(A_1),A_2) - \mu_1(H_1,A_1) + \mu_1(H_1,d_1(H_1)) \\ & \qquad + \frac{I\{A_1 = d_1(H_1)\}}{p_1(A_1|H_1)}\mu_2(H_2(A_1),d_2(H_2(A_1)))\Big]. \end{aligned} \tag{2.1}$$

This representation of the value, V_d, will form the basis for our method. The intuition behind this representation is that the potential outcome of Y under treatment policy d can be constructed or recovered from the potential outcome associated with the treatment sequence (A_1,A_2), by subtracting the intermediate treatment effects due to the sequence (A_1,A_2) and then adding in the intermediate treatment effects due to the policy d. The fraction involving the randomization probability in the last term (2.1) is used to account for the fact that the intermediate treatment effect of the second stage treatment under policy d depends on $H_2(a_1)|_{a_1=d_1(H_1)}$ (the covariate history that would occur if the first stage treatment were assigned according to policy d); that is, this fraction adjusts for the fact that $H_2(A_1)$ is not always equal to $H_2(d_1(H_1))$.

2.2.1 The Data and the Estimation Method

The observed data on each participant in a two-stage SMART is $\{X_1,A_1,X_2,A_2,X_3\}$ where X_t denotes covariates observed prior to the t-th stage and A_t denotes the t-th stage randomized treatment. Let $H_2 = (X_1,A_1,X_2)$ and $H_1 = X_1$. The randomization probability for an individual's treatment may be a function of the individual's observed data (say $P[A_t = a|H_t] = p_t(a|H_t)$). For example, in ExTENd (see Figure 2.1), participants were initially randomized uniformly to one of two criteria for early non-response to Naltrexone: the stringent definition (two or more heavy drinking days) or the lenient definition (five or more heavy drinking days). A heavy drinking day is defined as a day with more than five standard drinks for males or more than four standard drinks for females. Participants were assessed weekly for non-response; as soon as a participant met the non-response criterion, he/she was re-randomized to either switch to combined behavioral interventions (CBI) or to a combination of CBI and Naltrexone. If the participant did not meet his/her assigned

non-response criterion by the end of two months, then the participant was re-randomized to one of two relapse prevention options: usual care (UC) or telephone disease management (TDM). Thus non-responding participants had probability 0 of being assigned a relapse prevention option whereas responding participants had probability 0 of being assigned CBI or the combination of CBI and Naltrexone.

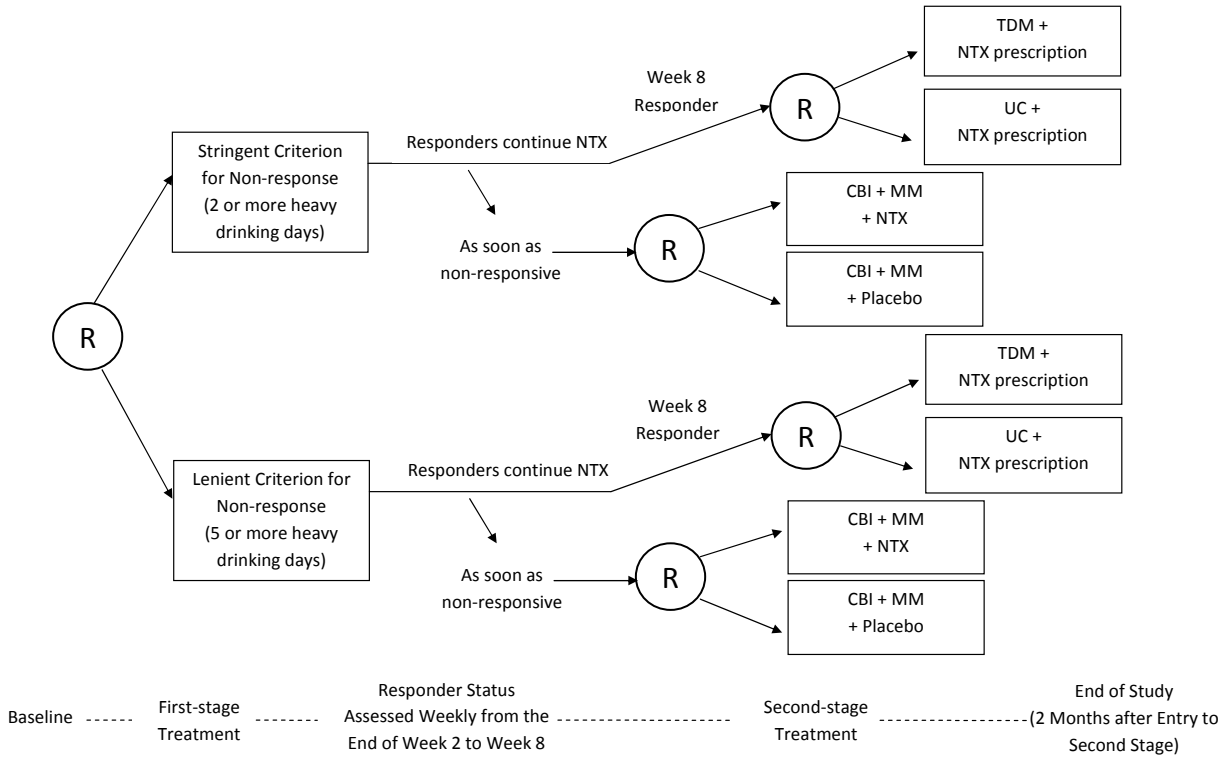

Figure 2.1: ExTENd SMART design for the treatment of alcohol dependence. "R" stands for (re-)randomization. TDM = Telephone Disease Management, UC = Usual Care, NTX = Naltrexone, CBI = Combined Behavioral Intervention, MM = Medical Management

Denote the primary outcome by Y (we assume a higher value is more favorable; in ExTENd Y might be percent days abstinent or a mental health score). To express the intermediate effects and the value (2.1) in terms of the observed data, we relate the observed data to the potential outcomes. We assume [66, 53, 50], (A1) Consistency: $X_2 = X_2(A_1), X_3 = X_3(A_1, A_2), Y = Y(A_1, A_2)$ and (A2) Sequential Randomization: A_1 is independent of all potential outcomes given observed X_1; A_2 is independent of all potential outcomes given observed (X_1, A_1, X_2). The consistency assumption states that the observed covariates are

identical to the potential outcomes of the covariates evaluated at the observed treatment sequence. In particular this assumption implies that each subject's outcomes are uninfluenced by other subjects' assigned treatments. This assumption may be violated if for example, treatment is provided in a group setting (group counseling). The sequential randomization assumption is valid in the setting of SMART trials because the treatment is randomized.

The intermediate treatment effects and the value, V_d, can be expressed in terms of the observed data as follows.

Lemma 2.2.2. *Assume A1 and A2 and (i)* $\max\{E|Y|, E|\mu_1(H_1, a_1)|, E|\mu_2(H_2, a_2)|\} < \infty$ *for any treatment sequence* (a_1, a_2) *and (ii) for some* $\delta > 0$, $p_1(a_1|h_1) \geq \delta$ *a.s. for* (h_1, a_1), *then*

(a) $\mu_2(h_2, a_2) = E[Y|H_2 = h_2, A_2 = a_2] - E[Y|H_2 = h_2, A_2 = 0]$,

(b) $\mu_1(h_1, a_1) = E[E[Y|H_2, A_2 = 0]|H_1 = h_1, A_1 = a_1] - E[E[Y|H_2, A_2 = 0]|H_1 = h_1, A_1 = 0]$ *and*

(c) $V_d = E\Big[Y - \mu_2(H_2, A_2) - \mu_1(H_1, A_1) + \mu_1(H_1, d_1(H_1)) + \frac{I\{A_1 = d_1(H_1)\}}{p_1(A_1|H_1)}\mu_2(H_2, d_2(H_2))\Big]$.

Suppose the intermediate treatment effects are known up to a finite-dimensional parameter: $\mu_1(h_1, a_1) = \mu_1(h_1, a_1; \beta_1), \mu_2(h_2, a_2) = \mu_2(h_2, a_2; \beta_2)$. [52] provides a class of "g-estimators" for the parameters, $\beta = (\beta_1, \beta_2)$. Each member in the class corresponds to a different choice of model for each of several nuisance functions; consistency of the g-estimators does not require correct models for the nuisance functions (see [52] for a detailed discussion). Furthermore this class of estimators does not require knowledge of the treatment policy, d. Thus β can be estimated and then used to form the estimators of the values of a variety of treatment policies. In the next section, we review the class of g-estimators. Each estimator in this class is consistent for the true value $\beta_0 = (\beta_{10}, \beta_{20})$ of β, and is asymptotically normally distributed (assuming a correctly specified SNMM and some finite moment conditions). Throughout the chapter we implicitly assume consistency and asymptotic normality of $\hat{\beta}$.

Then, given the results of Lemma 2.2.2 and estimators, $\hat{\beta}$, a natural assisted estimator of the value of the policy d, V_d is:

$$\hat{V}_0(d; \hat{\beta}) = \mathbb{P}_n\Big[Y - \mu_2(H_2, A_2; \hat{\beta}_2) - \mu_1(H_1, A_1; \hat{\beta}_1) + \mu_1(H_1, d_1(H_1); \hat{\beta}_1) + \frac{I\{A_1 = d_1(H_1)\}}{p_1(A_1|H_1)}\mu_2(H_2, d_2(H_2); \hat{\beta}_2)\Big], \tag{2.2}$$

where $\mathbb{P}_n f(X_1, A_1, X_2, A_2, X_3)$ denotes a sample average. This estimator belongs to a class of assisted estimators, given by

$$\hat{V}_m(d;\hat{\beta}) = \mathbb{P}_n\Big[Y - \mu_2(H_2, A_2; \hat{\beta}_2) - \mu_1(H_1, A_1; \hat{\beta}_1) + \mu_1(H_1, d_1(H_1); \hat{\beta}_1) \\ + \frac{I\{A_1 = d_1(H_1)\}}{p_1(A_1|H_1)} \Big\{\mu_2(H_2, d_2(H_2); \hat{\beta}_2) - m(H_1, A_1)\Big\} + m(H_1, d_1(H_1))\Big], \tag{2.3}$$

indexed by the function $m(h_1, a_1)$. Note the former assisted estimator, $\hat{V}_0(d; \hat{\beta})$, corresponds to setting $m(h_1, a_1) \equiv 0$. We have the following lemma:

Lemma 2.2.3. *Assume that the assumptions for Lemma 2.2.2 hold, then*

(a) The estimating function in (2.3) *is unbiased for any choice of* m *that satisfies* $E|m(H_1, a_1)| < \infty$ *for any* a_1.

(b) Assume (i) $E|Y|^2 < \infty$*; (ii)* $\dot{\mu}_1(h_1, a_1; \beta_1) := \frac{\partial}{\partial \beta_1}\mu_1(h_1, a_1; \beta_1)$ *exists for all* β_1*, a.s., and* $\dot{\mu}_2(h_2, a_2; \beta_2) := \frac{\partial}{\partial \beta_2}\mu_2(h_2, a_2; \beta_2)$ *exists for all* β_2*, a.s.; and (iii) there exists some* $\delta > 0$ *such that* $\sum_{a_1} E \sup_{\|\beta_1 - \beta_{10}\| \leq \delta} |\mu_1(H_1, a_1; \beta_1)|^2 + |\dot{\mu}_1(H_1, a_1; \beta_1)|^2 < \infty$*, and* $\sum_{a_2} E \sup_{\|\beta_2 - \beta_{20}\| \leq \delta} |\mu_2(H_2, a_2; \beta_2)|^2 + |\dot{\mu}_2(H_2, a_2; \beta_2)|^2 < \infty$*. Then if* $\hat{\beta}$ *belongs to a subclass* $\mathscr{B}$ *of g-estimators, the choice of* m *resulting in the lowest variance for* $\hat{V}_m(d; \hat{\beta})$ *satisfies* $m(h_1, d_1(h_1)) = E[\mu_2(H_2, d_2(H_2))|H_1 = h_1, A_1 = d_1(h_1)]$.

The subclass $\mathscr{B}$ corresponds to g-estimators for which a particular nuisance function is correctly modeled. This subclass is defined in Section 2.2.2 after a general review of g-estimators; in particular, in the simulation section we will first use an estimator $\hat{\beta}$ based on a correctly specified model for the nuisance function, thus $\hat{\beta} \in \mathscr{B}$. We will also provide additional simulation results when using a $\hat{\beta}$ that does not belong to $\mathscr{B}$.

The lemma above provides a guide for the choice of m; in practice $m(h_1, a_1)$ in (2.3) can be replaced by a working estimator $\hat{m}(h_1, a_1) := m(h_1, a_1; \hat{\alpha}_m)$ of $E[\mu_2(H_2, d_2(H_2))|H_1 = h_1, A_1 = a_1]$, resulting in $\hat{V}_{\hat{m}}(d; \hat{\beta})$. Next we provide consistency and asymptotic normality results for the estimators of the value. We assume A1 and A2; in addition, we assume that $\mu_1(h_1, a_1; \beta_1)$ and $\mu_2(h_2, a_2; \beta_2)$ are functions that correctly specify the SNMM, with true parameter value $\beta_0 = (\beta_{10}, \beta_{20})$. In particular, Theorem 2.2.4 below implies that the assisted estimator is consistent regardless of the choice of function m (indeed one can set $m \equiv 0$).

Theorem 2.2.4. *Assume that the assumptions for Lemma 2.2.3 hold; moreover, assume: (1)* $\hat{\alpha}_m$ *converges in probability to some limit* α_m^+*; (2) there exists some* $\delta > 0$ *such that* $\sum_{a_1} E \sup_{\|\alpha_m - \alpha_m^+\| \leq \delta} |m(H_1, a_1; \alpha_m)| < \infty$*; and (3)* $\dot{m}(h_1, a_1; \alpha_m) := \frac{\partial}{\partial \alpha_m} m(h_1, a_1; \alpha_m)$ *exists for all* α_m*, a.s. Then* $\hat{V}_{\hat{m}}(d; \hat{\beta})$ *is a consistent estimator for the policy value of* d*,* V_d.

Theorem 2.2.5. *Assume that the assumptions for Theorem 2.2.4 hold; moreover, assume: (1) there exists some* $\delta > 0$ *such that* $\sum_{a_1} E \sup_{\|\alpha_m - \alpha_m^+\| \leq \delta} |m(H_1, a_1; \alpha_m)|^2 + |\dot{m}(H_1, a_1; \alpha_m)|^2 < \infty$ *and (2)* $\sqrt{n}(\hat{\alpha}_m - \alpha_m^+) = O_p(1)$. *Then* $\sqrt{n}\left(\hat{V}_{\hat{m}}(d; \hat{\beta}) - V_d\right)$ *is asymptotically normal.*

The asymptotic variance of the limiting normal distribution in Theorem 2.2.5 is provided in the appendix. Recall that if $m(h_1, a_1; \alpha_m)$ is a correct model for $E[\mu_2(H_2, d_2(H_2))|H_1 = h_1, A_1 = a_1]$, then this asymptotic variance achieves the lowest value among all choices of m, provided that $\hat{\beta}$ belongs to the subclass $\mathscr{B}$ of g-estimators.

2.2.2 Estimators for SNMM

2.2.2.1 Review: Robins' G-Estimators for SNMM

Here we give a brief review of Robins' class of g-estimating equations [52] and the semiparametric locally efficient g-estimator. Assume that the SNMM is correctly specified. A class of estimating equations which can be used to solve for consistent estimators for β is:

$$\mathbb{P}_n\Big\{ r_1(H_1, A_1)\left(Y - \mu_2(H_2, A_2; \beta_2) - \mu_1(H_1, A_1; \beta_1) - q_1(H_1)\right) + r_2(H_2, A_2)\left(Y - \mu_2(H_2, A_2; \beta_2) - q_2(H_2)\right) \Big\} = 0,$$

where r_1, r_2 are arbitrary functions, both of the same dimension as the length of (β_1^T, β_2^T), that satisfy $E[r_1(H_1, A_1)|H_1] \equiv 0, E[r_2(H_2, A_2)|H_2] \equiv 0$; q_1, q_2 are arbitrary functions.

Assume that $Var(Y - \mu_2(H_2, A_2) - \mu_1(H_1, A_1)|H_1, A_1) \equiv Var(Y - \mu_2(H_2, A_2) - \mu_1(H_1, A_1)|H_1)$, which we will denote as $\sigma_1^2(H_1)$, and that $Var(Y - \mu_2(H_2, A_2)|H_2, A_2) \equiv Var(Y - \mu_2(H_2, A_2)|H_2)$, which we will denote as $\sigma_2^2(H_2)$. Robins provides r_1, r_2, q_1, q_2 functions that make the estimating equation semiparametric locally efficient; in particular the semiparametric locally efficient estimating equation is obtained by setting

$$q_1^*(h_1) = E[Y - \mu_2(H_2, A_2; \beta_{20}) - \mu_1(H_1, A_1; \beta_{10})|H_1 = h_1],$$

$$q_2^*(h_2) = E[Y - \mu_2(H_2, A_2; \beta_{20})|H_2 = h_2],$$

$$r_1^*(h_1, a_1) = \sigma_1^{-2}(h_1) \begin{pmatrix} \dot{\mu}_1(h_1, a_1; \beta_{10}) - E[\dot{\mu}_1(H_1, A_1; \beta_{10})|H_1 = h_1] \\ E[\dot{\mu}_2(H_2, A_2; \beta_{20})|H_1 = h_1, A_1 = a_1] - E[\dot{\mu}_2(H_2, A_2; \beta_{20})|H_1 = h_1] \end{pmatrix}$$

and

$$r_2^*(h_2, a_2) = \sigma_2^{-2}(h_2) \begin{pmatrix} 0 \\ \dot{\mu}_2(h_2, a_2; \beta_{20}) - E[\dot{\mu}_2(H_2, A_2; \beta_{20})|H_2 = h_2] \end{pmatrix}.$$

Consider models for $r_1(\cdot), r_2(\cdot), q_1(\cdot), q_2(\cdot)$, namely $r_1(\cdot;\eta), r_2(\cdot;\eta), q_1(\cdot;\xi), q_2(\cdot;\xi)$. If the parametric models specified for r_1, r_2, q_1, q_2 contain the truth (i.e., $r_1^*, r_2^*, q_1^*, q_2^*$), the estimator for β is then semiparametric efficient.

Definition of $\mathscr{B}$: The subclass $\mathscr{B}$ of g-estimators is defined as the collection of g-estimators in which $q_1(h_1;\xi)$ is a correctly specified model for $q_1^*(h_1)$. In Lemma 2.2.3, we show that the optimal m function in the assisted estimator can be identified if $\hat{\beta}$ belongs to this subclass. Note that the semiparametric efficient estimator belongs to this subclass.

2.2.2.2 Regression-Type Implementation of the G-Estimator

It turns out that for particular models of the nuisance functions (i.e., r_1, r_2, q_1, q_2) in the g-estimating equation, one can estimate both the nuisance functions and the β's simultaneously via least-squares. We use this approach to estimate the β parameters in the intermediate treatment effects in our simulations. We assume that the treatment effect functions are linear in the unknown parameters: $\mu_1(h_1, a_1; \beta_1) = \phi_1(h_1, a_1)^T\beta_1$ and $\mu_2(h_2, a_2; \beta_2) = \phi_2(h_2, a_2)^T\beta_2$, where ϕ_t is some feature of (h_t, a_t). The estimation is as follows:

1. First solve a linear regression of Y on $(\phi_2(H_2, A_2) - E[\phi_2(H_2, A_2)|H_2], M_2)$, in which M_2 is a summary of the history H_2. Note that in the setting of a randomized trial, the distribution of A_2 is known; thus $E[\phi_2(H_2, A_2)|H_2]$ can be calculated. Put $\hat{\beta}_2$ equal to the vector of the estimated coefficients for $\phi_2(H_2, A_2) - E[\phi_2(H_2, A_2)|H_2]$.

2. Second solve a linear regression of $Y - \phi_2(H_2, A_2)^T\hat{\beta}_2$ on $(\phi_1(H_1, A_1) - E[\phi_1(H_1, A_1)|H_1], M_1)$, in which M_1 is a summary of the history H_1. Again since the distribution of A_1 is known, $E[\phi_1(H_1, A_1)|H_1]$ can be calculated. Put $\hat{\beta}_1$ equal to the vector of the estimated coefficients for $\phi_1(H_1, A_1) - E[\phi_1(H_1, A_1)|H_1]$.

$\hat{\beta}$ obtained from this least-squares implementation is equivalent to a g-estimator with the following choice of nuisance functions: $r_1(H_1, A_1) = \tilde{\phi}_1(H_1, A_1)$, $r_2(H_2, A_2) = \tilde{\phi}_2(H_2, A_2)$, $q_1(H_1) = M_1^T\kappa_1^+ - E[\phi_1(H_1, A_1)|H_1]^T\beta_{10}$, $q_2(H_2) = M_2^T\kappa_2^+ - E[\phi_2(H_2, A_2)|H_2]^T\beta_{20}$, where $\tilde{\phi}_1 \equiv \tilde{\phi}_1(H_1, A_1) = \phi_1(H_1, A_1) - E[\phi_1(H_1, A_1)|H_1]$ and $\tilde{\phi}_2 \equiv \tilde{\phi}_2(H_2, A_2) = \phi_2(H_2, A_2) - E[\phi_2(H_2, A_2)|H_2]$; κ_1^+ and κ_2^+ denote the probabilistic limits of the estimated coefficients of M_1 and M_2 in the least-squares procedure. In particular, this regression-type estimator is consistent with correctly specified SNMM. Note that $\hat{\beta}$ obtained from this least-squares implementation belongs to the subclass $\mathscr{B}$ defined previously, provided that $M_1^T\kappa_1$ is a correct model for $q_1^*(H_1) + E[\phi_1(H_1, A_1)|H_1]^T\beta_{10}$.

Each member of the class of g-estimators is consistent and asymptotically normal. In particular, the asymptotic distribution of $\sqrt{n}(\hat{\beta} - \beta_0)$ is a multivariate normal with mean zero and var-covariance matrix $B^{-1}\Sigma B^{-1,T}$ where

$$B = \begin{pmatrix} E[\tilde{\phi}_1\tilde{\phi}_1^T] & E[\tilde{\phi}_1\phi_2^T] \\ 0 & E[\tilde{\phi}_2\tilde{\phi}_2^T] \end{pmatrix}$$

and $\Sigma = E\left(\left((Y - \phi_2^T\beta_{20} - \tilde{\phi}_1^T\beta_{10} - M_1^T\kappa_1^+)\tilde{\phi}_1^T, (Y - \tilde{\phi}_2^T\beta_{20} - M_2^T\kappa_2^+)\tilde{\phi}_2^T\right)^T\right)^{\otimes 2}$, where $V^{\otimes 2} = VV^T$. Plug-in estimates $\hat{B}$ and $\hat{\Sigma}$ can be obtained by replacing population expectation in B and Σ with sample mean, and replacing β, κ by the estimates from the series of least squares.

Prior to this least-squares type estimator for the SNMM, [2] proposed a parametric two-stage estimator that can be implemented by linear regression. Consistency of the estimator therein requires correct model specification for both the intermediate treatment effects (i.e., μ_1, μ_2) and the nuisance functions associated to the time-varying error terms.

2.2.3 Existing Work Regarding the Evaluation of A Treatment Policy

Here we review the methodologies for the evaluation and comparison of treatment policies proposed by [41] and [87]. We present those methods in the two-stage scenario.

[41] introduces the marginal mean models for the estimation of a mean response to a treatment policy (called DTR there). For simplicity, we ignore the discussion there about the mean value of a treatment policy over interesting subpopulations (denoted by Z in [41]), and only consider the estimation of the marginal mean value of a policy in the entire population. The estimator is based on the equality

$$E_d[Y] = E_{obs}\left[W_d(\bar{A}_2, \bar{X}_2)Y\right],$$

where $W_d(\bar{a}_2, \bar{x}_2) = \omega_{d,1}(a_1, x_1)\omega_{d,2}(\bar{a}_2, \bar{x}_2)$ and $\omega_{d,1}(a_1, x_1) = \frac{I\{A_1 = d_1(X_1)\}}{p_1(A_1|H_1)}, \omega_{d,2}(\bar{a}_2, \bar{x}_2) = \frac{I\{A_2 = d_2(\bar{X}_2, A_1)\}}{p_2(A_2|H_2)}$; E_d is the expectation in the population where all individuals follow the treatment policy d and E_{obs} is the expectation in the observed population. Thus the basic IPW estimator based on the marginal mean model is $\hat{V} = \mathbb{P}_n\left[W_d(\bar{A}_2, \bar{X}_2)Y\right]$.

There is a potential to improve the efficiency of this estimator by augmenting it (motivated by projecting the original estimating equation off the score functions for the treatment assignment probabilities, which are nuisance parameters for the estimation of policy value),

namely

$$\begin{aligned}\hat{V}(\hat{\alpha}_g) =& \mathbb{P}_n\big(\omega_{d,1}\omega_{d,2}Y + (g_1(X_1, d_1(X_1); \hat{\alpha}_g) - \omega_{d,1}g_1(X_1, A_1; \hat{\alpha}_g)) \\ &+ \omega_{d,1}(g_2(\bar{X}_2, A_1, d_2(\bar{X}_2, A_1); \hat{\alpha}_g) - \omega_{d,2}g_2(\bar{X}_2, \bar{A}_2; \hat{\alpha}_g))\big),\end{aligned}$$

in which $g_2(\bar{x}_2, \bar{a}_2; \alpha_g)$ is a model for $g_2(\bar{x}_2, \bar{a}_2) := E_{obs}[Y|\bar{X}_2 = \bar{x}_2, \bar{A}_2 = \bar{a}_2]$ and $g_1(x_1, a_1; \alpha_g)$ is a model for $g_1(x_1, a_1) := E_{obs}[g_2(\bar{X}_2, A_1, d_2(\bar{X}_2, A_1))|X_1 = x_1, A_1 = a_1]$.

These estimators are in essence non-parametric estimators that are obtained by properly weighting the observations in the data that happen to have the entire treatment sequences consistent with the policy, d, of interest. Data from those who have treatments consistent with the policy d only until the first stage, or data from those with treatments inconsistent with the policy d from the entry to study, are utilized to varying degrees in the augmented estimators, to potentially improve the efficiency.

[87] presents a robust augmented inverse probability weighted estimator for the values of a restricted class of treatment policies. In their paper the problem of policy value estimation is cast as one of monotone coarsening; however, with some calculation one can show that the general class of estimators proposed in this paper is equivalent to the estimators arising from the marginal mean model in Murphy et al. (2001). Here we briefly present the equivalence in the case of a two-stage problem.

For each two-stage policy $d = (d_1, d_2)$, conceptualize the complete data to be the potential outcomes associated with d: $(X_1, X_2(d_1), Y(d_1, d_2))$. Then a coarsening variable C_d can be defined for the complete data as below: If $A_1 \neq d_1(H_1)$, then $C_d = 1$. If $A_1 = d_1(H_1)$ and $A_2 \neq d_2(H_2)$, then $C_d = 2$. If $A_1 = d_1(H_1)$ and $A_2 = d_2(H_2)$, then $C_d = \infty$. Then define the hazard functions for this coarsening variable C_d as follows (coarsening at random is assumed, and in the scenario of sequential randomized trials this assumption is naturally satisfied): $\lambda_{d,1}(X_1) = Pr(C_d = 1|X_1)$, and $\lambda_{d,2}(X_1, X_2) = Pr(C_d = 2|C_d \geq 2, X_1, X_2)$. Then the class of estimators (indexed by the functions $L_1(x_1)$ and $L_2(x_1, a_1, x_2)$) proposed in Zhang et al. (2013) can be written as:

$$\begin{aligned}\mathbb{P}_n\Big\{\frac{I\{C_d = \infty\}}{(1-\lambda_{d,1})(1-\lambda_{d,2})}Y + \frac{I\{C_d = 1\} - \lambda_{d,1}}{1-\lambda_{d,1}}L_1(X_1) + \\ \frac{I\{C_d = 2\} - \lambda_{d,2}I\{C_d \geq 2\}}{(1-\lambda_{d,1})(1-\lambda_{d,2})}L_2(X_1, A_1, X_2)\Big\},\end{aligned}$$

in which $\lambda_{d,1} = \lambda_{d,1}(X_1), \lambda_{d,2} = \lambda_{d,2}(X_1, X_2)$. The consistency of any estimator in this class is guaranteed, regardless of the choices of L_1, L_2.

Equivalency of this class of estimators and the estimators in [41] can be established by

setting $L_1(X_1) = g_1(X_1, d_1(X_1))$ and $L_2(X_1, A_1, X_2) = g_2(\bar{X}_2, A_1, d_2(\bar{X}_2, A_1))$.

In the simulation section, we will compare the assisted estimators with the estimators arising from the marginal mean models. Note that the model specification for $g_t(\cdot)$ does not have an impact on the consistency of the estimator $\hat{V}(\hat{\alpha}_g)$. Suggested by [41], to guarantee that the models for g_t are consistent with each other under the null, in the simulation experiments we model g_t as linear in $\bar{x}_t$ and independent of $\bar{a}_t$. In particular, we estimate $g_2(X_1, A_1, X_2, A_1; \hat{\alpha}_g)$ by regressing Y on intercept and $\bar{X}_2$, then regress the fitted values on intercept and X_1 to obtain $g_1(X_1, A_1; \hat{\alpha}_g)$.

2.3 Comparison between Treatment Policies

Suppose we are interested in comparing treatment policies $d = (d_1, d_2)$ and $\tilde{d} = (\tilde{d}_1, \tilde{d}_2)$. Then, given an estimator $\hat{\beta}$ for the intermediate treatment effects, we obtain the following consistent estimator for the contrast between d and $\tilde{d}$, i.e., $V_{\tilde{d}} - V_d$:

$$
\begin{aligned}
(\hat{V}_{m_{\tilde{d}}}(\tilde{d}; \hat{\beta}) - \hat{V}_{m_d}(d; \hat{\beta})) = & \mathbb{P}_n \Big[\mu_1(H_1, \tilde{d}_1(H_1); \hat{\beta}_1) - \mu_1(H_1, d_1(H_1); \hat{\beta}_1) \\
& + \frac{I\{A_1 = \tilde{d}_1(H_1)\}}{p_1(A_1|H_1)} \Big\{ \mu_2(H_2, \tilde{d}_2(H_2); \hat{\beta}_2) - m_{\tilde{d}}(H_1, A_1) \Big\} \\
& - \frac{I\{A_1 = d_1(H_1)\}}{p_1(A_1|H_1)} \Big\{ \mu_2(H_2, d_2(H_2); \hat{\beta}_2) - m_d(H_1, A_1) \Big\} \\
& + m_{\tilde{d}}(H_1, \tilde{d}_1(H_1)) - m_d(H_1, d_1(H_1)) \Big],
\end{aligned} \tag{2.4}
$$

where the function $m(h_1, a_1)$ is now subscripted by the policy d, to reflect that a good choice of function m varies with d (see the following lemma). For ease of notation, define $\Delta_d(h_1, a_1) = m_d(h_1, a_1) - E[\mu_2(H_2, d_2(H_2))|H_1 = h_1, A_1 = a_1]$.

Lemma 2.3.1. *Assume that the conditions for Lemma 2.2.3 are satisfied; in particular, assume that $\hat{\beta}$ belongs to the subclass $\mathscr{B}$ of g-estimators. Then the choice of m_d and $m_{\tilde{d}}$ resulting in the lowest asymptotic variance for $\sqrt{n}(\hat{V}_{m_{\tilde{d}}}(\tilde{d}; \hat{\beta}) - \hat{V}_{m_d}(d; \hat{\beta}))$, among the class of estimators in* (2.4) *with m_d and $m_{\tilde{d}}$ being arbitrary functions of (h_1, a_1), satisfy: (1) for h_1 such that $d_1(h_1) \neq \tilde{d}_1(h_1)$, $\Delta_{\tilde{d}}(h_1, \tilde{d}_1(h_1)) = \Delta_d(h_1, d_1(h_1)) = 0$; (2) for h_1 such that $d_1(h_1) = \tilde{d}_1(h_1)$, $\Delta_{\tilde{d}}(h_1, \tilde{d}_1(h_1)) = \Delta_d(h_1, d_1(h_1))$.*

Lemma 2.3.1 implies that, for the purpose of estimating the policy contrast, it is reasonable to replace $m_d(h_1, a_1)$ with a working estimate $m_d(h_1, a_1; \hat{\alpha}_m)$ of $E[\mu_2(H_2, d_2(H_2))|H_1 = h_1, A_1 = a_1]$. Then we have the following lemma concerning the estimator of the contrast in (2.4) with $m_d(h_1, a_1)$ replaced by $m_d(h_1, a_1; \hat{\alpha}_m)$. We will also refer to this estimator as

an "assisted estimator". This lemma assumes that $m_d(h_1, a_1; \hat{\alpha}_m)$ is modeled via a linear model $D_m^T \alpha_m$ where D_m is a function of (H_1, A_1) and α_m is estimated via least squares.

Lemma 2.3.2. *Assume that the conditions for Theorem 2.2.4 and 2.2.5 are satisfied; then* $\sqrt{n}\big((\hat{V}_{\hat{m}_{\tilde{d}}}(\tilde{d}; \hat{\beta}) - \hat{V}_{\hat{m}_d}(d; \hat{\beta})) - (V_{\tilde{d}} - V_d)\big)$ *converges in distribution to a normal distribution with mean zero and var-covariance matrix,* Σ_Δ*. The plug-in estimator* $\hat{\Sigma}_\Delta$ *is a consistent estimator of* Σ_Δ*.*

The formulae for Σ_Δ and $\hat{\Sigma}_\Delta$ are provided in the appendix.

2.4 Simulation

All simulation experiments are based on generative models mimicking the ExTENd study. More specifically, the structure of the simulated data is: $(X_1, A_1, X_2, R, A_2, Y)$. X_1 is a 3-dimension baseline covariate simulating the distribution of {baseline percent days heavy drinking, baseline craving score, baseline mental composite score}, A_1 is the binary indicator of the randomized non-response criterion, X_2 is a 2-dimension covariate simulating the distribution of {phase 1 duration, phase 1 percent days drinking}, R is the binary indicator of early response, A_2 is the re-randomized binary treatment at the second stage. Y is a primary outcome simulating the distribution of the end-of-study craving score (lower values are better). We will study various simulation scenarios that are all based on the following Y:

$$Y = \eta_0(X_1) + A_1(1, X_1^T)\beta_1 + \eta_1(X_1, A_1, X_2) + A_2(1, X_2^T, A_1, R, RX_2^T, RA_1)\beta_2 + \epsilon. \quad (2.5)$$

in which the terms involving β's are the intermediate treatment effects and $\eta_0(\cdot), \eta_1(\cdot)$ and ϵ are other components in the distribution of Y that correspond to the main effect of X_1, the effect of X_2 conditional on (X_1, A_1) and the error term, respectively. We use estimates of $\eta_0(\cdot)$ and $\eta_1(\cdot)$ that are by-products of estimating an SNMM with the ExTENd data; the by-products of the estimation of SNMM also include an estimate of the variance of the error term, and we use that variance estimate to generate ϵ in our simulations. More details are provided in the appendix.

We create nine simulation scenarios by varying β_1, β_2 in the generating model for Y. This procedure alters the magnitude of the main effects of the treatments at both stages and also the extent to which there are treatment by time-varying covariate interactions. In particular, the first coordinates in β_1 and β_2 reflect the main effects of A_1 and A_2, and the remaining coordinates reflect the interactions of A_1 and A_2 with time-varying covariates. We

adopt the following definition of standardized effect size of a coordinate in β_j by slightly modifying Cohen's d measure to: $SES(\beta_{jk}) = \beta_{jk}/\sqrt{Var(\eta_0(X_1)) + Var(\eta_1(X_1, A_1, X_2)) + Var(\epsilon)}$. We adopt this definition of standardized effect size because $\eta_0(X_1), \eta_1(X_1, A_1, X_2)$ and ϵ are uncorrelated components in the generative model of primary outcome Y, and the sum of their variances contributes to the majority of the variance in Y. Note that to ensure that this definition of standardized effect size is meaningful, we will use standardized covariates (each covariate in X_1, X_2 is standardized to come from a population with mean 0 and standard deviation equal to 1). The nine simulation scenarios correspond to combinations of no treatment effect, low treatment effect and medium treatment effect at both stages. We define no A_j treatment effect ($j = 1, 2$) as $\beta_j = 0$, define low A_j treatment effect as setting all coordinates in β_j to have SES equal to 0.2, and define medium A_j treatment effect as setting the first two coordinates in β_j to have SES equal to 0.5 (i.e., main effect and interaction effect with X_{j1}), and the other coordinates in β_j to have SES equal to 0.2. The rationale for only one medium level interaction in medium A_j treatment effect case is that it is unlikely (in real data) for the treatment to interact with many covariates at medium level. The sign of each coordinate in β_j is determined by a preliminary fit to the ExTENd data. In each simulation scenario, we generate 1000 simulated data sets.

Throughout $\hat{\beta}$ in the assisted estimator is one of Robins' g-estimators that belongs to $\mathscr{B}$ ($\hat{\beta}$ is the solution to a series of least squares problems; indeed if, as discussed above a particular nuisance function is correctly modeled, then this least squares solution will belong to $\mathscr{B}$). In the appendix we provide results when $\hat{\beta}$ does not belong to $\mathscr{B}$; the simulation results are similar. Also throughout $\hat{m}_d$ is estimated via least squares with $(1, X_1, A_1)$ as predictors.

Let the triple (c_1, c_2, c_3) denote a policy in which c_1 is the assigned non-response criterion, c_2 is the assigned binary treatment for early responders at the second stage, and c_3 is the assigned binary treatment for early non-responders at the second stage. To investigate different aspects of the proposed methodology, we perform two sets of simulation experiments: The first set studies the bias and MSE of the assisted estimators of the difference in values of the most intensive policy, (1,1,1) and the least intensive policy, (0,0,0). The second set illustrates the efficiency gain of using the assisted estimator, compared with a non-parametric policy value estimator that is based on the marginal mean model.

Simulation 1: Here we compare bias and MSE for three types of assisted estimators for difference in value. We use the assisted estimator, $\hat{V}_{\hat{m}_d}(d; \hat{\beta})$ with $\hat{m}_d$, an estimator of $E[\mu_2(H_2, d_2(H_2))|H_1, A_1]$, and $\hat{V}_0(d; \hat{\beta})$, to estimate the contrast between embedded policies $(1, 1, 1)$ and $(0, 0, 0)$. We also consider $\hat{V}_{m_d}(d; \hat{\beta})$ in which m_d is the unknown $E[\mu_2(H_2, d_2(H_2))|H_1, A_1]$; we call this an "oracle" assisted estimator, because in prac-

tice the optimal m_d will be unknown. The coverage of confidence intervals based on the asymptotic standard errors of each of the two non-oracle estimators is also provided.

Table 2.1: Simulation 1: Statistical properties of the assisted estimators of the contrast between values of policies (1,1,1) and (0,0,0). Oracle = contrast estimator based on $\hat{V}_{m_d}(d;\hat{\beta})$ with the true optimal m_d. Assist = contrast estimator based on $\hat{V}_{\hat{m}_d}(d;\hat{\beta})$ with a working estimate of the optimal m_d. Assist ($m_d = 0$) = contrast estimator based on $\hat{V}_0(d;\hat{\beta})$. The displayed numbers for confidence interval coverage are the coverage proportion $\times$ 100. An Asterisk indicates that the MSE of Oracle or Assist ($m_d = 0$) is significantly different from MSE of Assist (at 0.05 level).

$N = 100$

Scenario	True Value	Bias / SD			MSE			ASE Coverage	
		Oracle	Assist	Assist ($m_d = 0$)	Oracle	Assist	Assist ($m_d = 0$)	Assist	Assist ($m_d = 0$)
(none,none)	0	0.04	0.04	0.04	3.51*	3.46	3.51*	95.7	95.4
(none,low)	-2.4	0.01	0.01	0.01	4.26	4.26	4.31	95.1	95.6
(none,med)	-5.2	0.03	0.03	0.01	3.94	3.93	4.3*	95.2	95.4
(low,none)	-1.4	-0.01	-0.01	-0.01	3.31	3.3	3.31	95.5	96.3
(low,low)	-3.8	0	0	0	4.08	4.14	4.12	95.5	95.9
(low,med)	-6.6	0.04	0.04	0.04	4.09	4.1	4.25*	95.6	96.3
(med,none)	-3.6	0.03	0.03	0.03	3.96	3.93	3.96	95.9	95.4
(med,low)	-6.0	-0.01	-0.01	-0.01	4.33	4.36	4.38	95.2	95.5
(med,med)	-8.8	0.01	0.01	0	4.02	4.04	4.24*	95	95.7

$N = 250$

Scenario	True Value	Bias / SD			MSE			ASE Coverage	
		Oracle	Assist	Assist ($m_d = 0$)	Oracle	Assist	Assist ($m_d = 0$)	Assist	Assist ($m_d = 0$)
(none,none)	0	0	0	0	1.3	1.31	1.3	95	95
(none,low)	-2.4	0.03	0.03	0.03	1.44	1.45	1.47*	95.1	95.1
(none,med)	-5.2	0.03	0.02	0.02	1.4	1.4	1.48*	94.6	95.6
(low,none)	-1.4	-0.01	-0.01	-0.01	1.31	1.32	1.31	95.7	95.6
(low,low)	-3.8	-0.02	-0.02	-0.02	1.69	1.71	1.71	93.2	93.6
(low,med)	-6.6	0	-0.01	0	1.42	1.42	1.54*	95.3	95.2
(med,none)	-3.6	-0.03	-0.03	-0.03	1.38	1.38	1.38	95.2	95.1
(med,low)	-6.0	0	0	0.01	1.64	1.64	1.67	94.8	95
(med,med)	-8.8	-0.02	-0.02	-0.01	1.55	1.55	1.63*	94.8	94.7

The simulation results with $N = 100$ and $N = 250$ are shown in Table 2.1. Based on the ratio of bias and standard deviation, we conclude that, as expected, the assisted estimators provide an unbiased estimate of the contrast between policies. The MSEs of all the three estimators are similar; $\hat{V}_{\hat{m}_d}(d;\hat{\beta})$ tends to be slightly more efficient than $\hat{V}_0(d;\hat{\beta})$. The coverage of the confidence intervals based on the asymptotic standard errors is close to 95% in all cases.

In the appendix we provide additional simulations; these simulations illustrate that $\hat{V}_{\hat{m}_d}(d;\hat{\beta})$ will provide a noticeable efficiency improvement over $\hat{V}_0(d;\hat{\beta})$ in some extreme settings. However, in most practical scenarios, a sophisticated chosen m_d does not substantially improve the efficiency over $m_d \equiv 0$; therefore for simplicity we recommend using the assisted estimator with $m_d \equiv 0$.

Simulation 2: Here we assess the robustness via the bias, MSE and confidence interval coverage provided by the assisted estimators to misspecification of the SNMM. As a comparison we consider estimators from the marginal mean model [41] as these estimators do not require the SNMM. The marginal mean models are estimated via a non-parametric inverse-weighted estimator. Note that when the goal is to evaluate the difference between two policies, the estimators in [46] under particular choices of nuisance functions reduce to the marginal mean model estimators.

$\hat{V}_{\hat{m}_d}(d;\hat{\beta})$ is estimated with two differently mis-specified SNMMs in addition to the correctly specified SNMM. The true SNMM is implied by the generative model in (2.5), i.e., $\mu_1(H_1, A_1) = A_1(1, X_1^T)\beta_1, \mu_2(H_2, A_2) = A_2(1, X_2^T, A_1, R, RX_2^T, RA_2)\beta_2$. The first mis-specification of the SNMM excludes X_{11} from the model for $\mu_1(H_1, A_1)$ and excludes X_{21}, RX_{21} from the model for $\mu_2(H_2, A_2)$ (denoted as Assist2 in Table 2.2). The second mis-specification models $\mu_1(H_1, A_1)$ as $A_1(1, X_1^{*T})\beta_1$ and models $\mu_2(H_2, A_2)$ as $A_2(1, X_2^{*T})\beta_2$, where X_1^* and X_2^* are 3-dimensional and 7-dimensional covariates (denoted as Assist3 in Table 2.2). X_1^* and X_2^* generated independently of all the other covariates; the dimensions of X_1^* and X_2^* are chosen so that the model complexity is the same as in the correctly specified SNMM.

We focus on the estimation of two contrasts: the first is the contrast between the policies (1,1,1) and (0,0,0), and the second is the contrast between a "tailored" treatment policy and the policy $(0, 0, 0)$. This tailored treatment policy assigns $a_1 = 1$ if $X_{13} > 0$; $a_2 = 1$ to all early responders and $a_2 = 1$ to early non-responders if $X_{21} < 0$. In each of the nine simulation scenarios we compare the marginal-mean-model-based estimator with the assisted estimators for three differently specified SNMMs.

Table 2.2: Simulation 2: Comparison between the marginal-mean-model-based estimators and the assisted estimators, with respect to the performance in estimating the policy contrasts, with $N = 100$. MM = Marginal-mean-model-based estimator. Assist1 = Assisted estimator with correctly specified SNMM. Assist2 = Assisted estimator with mis-specified SNMM that excludes X_{11}, X_{21}, RX_{21}. Assist3 = Assisted estimator with mis-specified SNMM that excludes all the covariates interacting with treatments. Bias significantly different from 0, and coverage proportion significantly different from 95%, are marked with an asterisk. Relative MSE is calculated as the ratio of MSE with that of MM.

$N = 100$											
Estimation of the first contrast,(1, 1, 1) vs (0,0,0)											
Scenario	Bias x 100				Coverage of 95% CI x 100				Relative MSE		
	MM	Assist1	Assist2	Assist3	MM	Assist1	Assist2	Assist3	Assist1	Assist2	Assist3
(none,none)	2.4	4.9	5.2	4.9	95.2	96.2	96	96.1	0.94	0.93	0.99
(none,low)	5.8	4.6	4.7	6	94.5	96	95.4	95.2	0.95	0.94	1.04
(none,med)	12	-6.8	-6.8	-2.6	93.6*	93.9	93.6*	94.6	0.95	0.95	1.01
(low,none)	-1.9	2.5	1.7	4.8	95.6	94.6	94	95	1.01	1.01	1.09
(low,low)	-12.5	-10.8	-11	-10.3	94.3	94.5	93.5*	94.6	0.92	0.92	0.97
(low,med)	11	-9.9	-10.4	-5.8	93.9	94.8	94.7	95.5	0.84	0.84	0.93
(med,none)	8.9	4.2	5.4	3.4	95.5	95.9	95.3	96.2	0.89	0.87	0.89
(med,low)	9.7	-1.9	-2.7	-7.1	94.3	94.8	94.1	94.9	0.85	0.85	0.93
(med,med)	28.9*	4.2	5.4	4.7	93.7	94.9	95.2	94.9	0.8	0.79	0.85
Estimation of the second contrast, the tailored policy vs (0,0,0)											
Scenario	Bias x 100				Coverage of 95% CI x 100				Relative MSE		
	MM	Assist1	Assist2	Assist3	MM	Assist1	Assist2	Assist3	Assist1	Assist2	Assist3
(none,none)	6	1	2.4	2.3	96.2	97*	96.6*	96.1	0.78	0.76	0.57
(none,low)	6.4	4.8	-2.8	16.7*	95.6	96	95.7	94.7	0.79	0.77	0.59
(none,med)	11.5	-2.8	-22.1*	-43.9*	94.9	95.8	95.1	94.4	0.78	0.77	0.67
(low,none)	5.3	11.2*	9.7	42.9*	95.5	95.3	94.8	93.8	0.81	0.8	0.69
(low,low)	-7.3	-6.3	-15.1*	46.3*	93.9	95.3	93.9	95	0.77	0.74	0.59
(low,med)	6.7	-1.8	-23.6*	-2.8	94	96.3	94.9	95.7	0.7	0.69	0.5
(med,none)	9.3	8	9.1	50*	95.9	96.5*	95.8	95.4	0.76	0.74	0.57
(med,low)	13.7*	9.4	-0.3	53.3*	93.2*	95	95.2	94.1	0.7	0.67	0.57
(med,med)	24.7*	5.2	-15.1*	9.9*	93.1*	95.5	95.3	95.6	0.66	0.64	0.49

Table 2.3: Simulation 2: Comparison between the marginal-mean-model-based estimators and the assisted estimators, with respect to the performance in estimating the policy contrasts, with $N = 250$. MM = Marginal-mean-model-based estimator. Assist1 = Assisted estimator with correctly specified SNMM. Assist2 = Assisted estimator with mis-specified SNMM that excludes X_{11}, X_{21}, RX_{21}. Assist3 = Assisted estimator with mis-specified SNMM that excludes all the covariates interacting with treatments. Bias significantly different from 0, and coverage proportion significantly different from 95%, are marked with an asterisk. Relative MSE is calculated as the ratio of MSE with that of MM.

$N = 250$											
Estimation of the first contrast,(1, 1, 1) vs (0,0,0)											
Scenario	Bias x 100				Coverage of 95% CI x 100				Relative MSE		
	MM	Assist1	Assist2	Assist3	MM	Assist1	Assist2	Assist3	Assist1	Assist2	Assist3
(none,none)	2.6	4.7	4.4	4.8	93.5*	94.6	94.4	94.4	0.87	0.86	0.89
(none,low)	2	1.5	1.6	2.5	93.8	94.7	94.5	94.6	0.82	0.81	0.87
(none,med)	7.2	-1.2	-1.4	0.3	94.6	94.7	94.9	94.9	0.82	0.83	0.85
(low,none)	-4.6	-2.4	-3.3	-3.8	95.2	95.1	95	95.3	0.83	0.83	0.86
(low,low)	-5.1	-6	-5.8	-6.4	94.5	93.9	93.6*	93.8	0.87	0.87	0.89
(low,med)	6	-0.3	-0.5	1.3	96	95.4	95.4	95.8	0.79	0.8	0.84
(med,none)	-2.3	-1.3	-1.8	-1.1	94.5	94.3	94.3	95.9	0.75	0.76	0.78
(med,low)	9.3*	7.6	7.9	8.1	94.4	94.5	94.3	94.1	0.79	0.79	0.8
(med,med)	20.6*	11.2*	10.6*	14.3*	94.5	94.5	93.7	94.2	0.73	0.74	0.78
Estimation of the second contrast, the tailored policy vs (0,0,0)											
Scenario	Bias x 100				Coverage of 95% CI x 100				Relative MSE		
	MM	Assist1	Assist2	Assist3	MM	Assist1	Assist2	Assist3	Assist1	Assist2	Assist3
(none,none)	-0.8	0.2	-0.2	2.3	95.1	93.6*	93.3*	95.4	0.69	0.67	0.48
(none,low)	0.2	0.5	-7.3*	13.2*	93.6*	94.7	94.3	94.9	0.65	0.64	0.46
(none,med)	6.7	0.9	-20*	-39.5*	93.8	94.6	93.7	92.8*	0.69	0.71	0.58
(low,none)	-1.6	0.1	-1	38.5*	95.1	95.1	95.1	92.4*	0.67	0.66	0.6
(low,low)	-3.8	-1.5	-9.2*	49.1*	95.5	94.8	94.9	91.5*	0.68	0.68	0.62
(low,med)	0.7	-0.5	-21.9*	-0.7	95.1	95.2	94.7	96	0.68	0.7	0.45
(med,none)	2	3.6	2.4	46.9*	95.3	94.9	95.4	91.1*	0.6	0.59	0.55
(med,low)	10.4*	7.3*	-0.5	63.7*	94.4	94.6	94.8	88.8*	0.62	0.6	0.65
(med,med)	8.7*	6.7	-15*	15.3*	94.7	95.4	94.9	94.1	0.64	0.62	0.46

The experiment results when $N = 100$ are shown in Table 2.2; results for $N = 250$ are shown in Table 2.3. Instead of the MSE of the estimators, we present the relative MSE of the assisted estimators, with the MSE of the marginal-mean-model-based estimator (MM) as the reference. From the simulation results with $N = 100$, we found that, for the comparison between the two embedded policies, the assisted estimators with correctly specified SNMM outperform MM in terms of the MSE in most cases; mis-specifying the SNMM does not seem to introduce bias, but severe mis-specification (Assist3 in the Table) can lead to lower efficiency, and sometimes can even cause the assisted estimators to have a larger MSE than MM. For the comparison between the tailored policy and the reference policy, the assisted estimators with correctly specified SNMM outperform MM in terms of the MSE, and the advantage is greater than that of the first contrast. Mis-specifying the SNMM introduces bias; in particular, severe mis-specification (Assist3) leads to considerable bias. However, this bias does not seem to greatly impact the performance of the confidence interval. Interestingly, for the estimation of this contrast, mis-specifying the SNMM may even result in a smaller MSE despite of the bias, due to a smaller standard deviation in the estimate.

With a larger sample size ($N = 250$ as compared to $N = 100$), the advantage of the assisted estimators in terms of having a lower MSE than the marginal-mean-model-based estimators is more evident. Similar to the $N = 100$ experiments, mis-specifying the SNMM introduces bias in some scenarios, but even in those scenarios the performance of the assisted estimators in terms of the MSE does not worsen, because reduction in the variance dominates the bias-variance tradeoff. We notice that under the most severe mis-specification of SNMM (Assist3), the confidence interval of the contrast between the tailored policy and the policy $(0, 0, 0)$ has noticeable under-coverage. However, we expect that in practice, such severe mis-specification, which fails to use any variable correlated with the variables in the true SNMM, might be unlikely to happen.

2.5 Illustration with the ExTENd Data

The ExTENd study (see Figure 2.1) includes 302 participants, with 49 participants dropping out prior to experiencing two heavy drinking days. These participants are removed from our analysis as they did not experience the first randomization and both they and the clinicians were blind to this randomization. Only three participants dropped out during the first treatment stage after experiencing two heavy drinking days. The data from these participants is also removed for simplicity. Thus the data we analyze has a sample size of 250.

We use both the marginal-mean-model-based estimator and the assisted estimator to compare the most intensive versus the least intensive policies. Treatment policy (1,1,1) represents the most intensive policy in the SMART, in which the early non-response is deemed to occur if and when there are 5 or more heavy drinking days in the first 8 weeks, in which early responders are provided TDM and in which early non-responders are provided NTX+CBI. Treatment policy (0,0,0) represents the least intensive policy, in which early non-response is deemed to occur if and when there are 2 or more heavy drinking days in the first 8 weeks, in which early responders are provided UC and in which early non-responders are provided CBI only.

Besides the two treatment policies above, we will also compare a more "deeply tailored" policy versus the policy (0, 0, 0). At stage one, this tailored policy assigns the 5 or more heavy drinking days definition of non-response to participants for whom the standardized pre-treatment mental score is above zero and the 2 or more heavy drinking days definition of non-response to participants with a pre-treatment mental score below zero. Among early responders this policy assigns TDM if they have at least one heavy drinking day during stage one and assigns UC otherwise. Among early non-responders this policy assigns NTX+CBI if their stage one duration is shorter than 49 days and otherwise assigns CBI only. The justification of this treatment policy comes from the belief that participants who were in worse mental health condition (indicated by a lower mental composite score) at baseline should proceed to stage two earlier to receive more intensive treatments. Moreover, it is considered that responders and non-responders who performed worse in stage one (i.e., responders who experienced at least one heavy drinking day and non-responders who transitioned to stage two sooner) should receive more intensive intervention in stage two.

We compare the treatment policies in terms of the Penn Alcohol Craving Scale (PACS). Here we reverse code this scale such that higher values imply less craving thus are more favorable. PACS is collected every two months during stage two. The outcome Y is the average of the measurement at two months and four months after entry into stage two. Among the 250 participants in our data set, 46 participants are missing Y. We deal with this missingness in the outcome, Y, by adopting a slightly adjusted assisted estimator that handles missingness via inverse-probability-weights (see [62] for example). The adjustment requires an estimator of the conditional probability of missing the outcome. This adjustment is briefly presented in the appendix. In particular, we make the assumption that the missing Y's are missing at random [65]. The marginal-mean-model-based estimator is also adjusted similarly to accommodate for missingness.

In the analysis model, we choose to include the following covariates: X_1 is a 10-

Table 2.4: Illustrative data analysis results with the ExTENd data. Evaluate the policy contrasts of both the policy (1, 1, 1) and the proposed tailored policy, in relation to the policy (0, 0, 0), with respect to PACS. MM = Marginal-mean-model-based estimator. Assist1 = Assisted estimator with a parsimonious SNMM. Assist2 = Assisted estimator with a complex SNMM.

		(1,1,1) vs (0,0,0)			Tailored vs (0,0,0)		
		Est (s.e.)	Lower Bound	Upper Bound	Est (s.e.)	Lower Bound	Upper Bound
PACS	MM	2.98 (1.30)	0.44	5.52	0.21 (1.05)	-1.85	2.27
	Assist1	2.83 (1.44)	0.00	5.66	0.91 (0.99)	-1.02	2.85
	Assist2	2.95 (1.48)	0.04	5.85	1.25 (1.05)	-0.80	3.31

dimensional baseline covariate including mean-centered versions of {gender, age, years of alcohol use, indicator of drug abuse, pre-treatment percent days heavy drinking, indicator of being married, years of alcohol intoxication, pre-treatment alcohol intoxication days within 30 days, pre-treatment percent days drinking, pre-treatment mental composite score}; X_2 is 5-dimensional covariate measured prior to re-randomization, including {duration of the first stage, number of heavy drinking days during the first stage, percent days drinking during the first stage, percent days heavy drinking during the first stage, average number of pills taken per day during the first stage}. Moreover, A_1 indicates whether $(A_1 = 1)$ or not $(A_1 = 0)$ a patient is randomized to the lenient definition (i.e., five or more heavy drinking days) of non-response as opposed to the stringent definition (i.e., two or more heavy drinking days); R is the indicator of being an early responder; A_2 indicates whether $(A_2 = 1)$ or not $(A_2 = 0)$ a responder is re-randomized to TDM as opposed to UC, or whether or not a non-responder is re-randomized to NTX+CBI as opposed to Placebo+CBI.

We run two sets of analysis with the assisted estimators, under two different SNMMs: in the first analysis we adopt a parsimonious model for SNMM by assuming $\mu_1(H_1, A_1) = A_1(1, \tilde{X}_1^T)\beta_1$ and $\mu_2(H_2, A_2) = A_2(1, \tilde{X}_2^T, A_1, R, R\tilde{X}_2^T, RA_1)\beta_2$, where $\tilde{X}_1$ is the first five dimensions in X_1 and $\tilde{X}_2$ is the first three dimensions in X_2; in the second analysis we adopt a more complex model for SNMM by assuming $\mu_1(H_1, A_1) = A_1(1, X_1^T)\beta_1$ and $\mu_2(H_2, A_2) = A_2(1, X_2^T, A_1, R, RX_2^T, RA_1)\beta_2$. Asymptotic standard errors of the policy contrast estimates are calculated and used to construct the 95% confidence intervals for the policy contrasts. Table 2.4 presents the analysis results.

The three estimators (including two assisted estimators with different SNMMs) produce similar estimates, considering the relatively large standard errors. The analyses suggest that the most intensive, (1,1,1) policy is estimated to approximately lower PACS by 3 on

average compared to the least intensive, (0,0,0) policy, and this difference is significant at 0.05 level, across all three estimators. The proposed more tailored policy, on the other hand, does not significantly differ from the (0,0,0) policy. Note that the marginal-mean-model based estimator has standard error no greater than that of the assisted estimators; this might be due to either small treatment effects in the ExTENd data, or the variance due to the considerable amount of missingness in the data.

2.6 Extension to More than 2 Stages

In this chapter we focused on the comparison of two-stage treatment policies. Most of the SMART studies that have been completed or are on-going are two-stage trials, that is, in these studies each participant was at most randomized twice. None-the-less SMART studies with more than two stages are likely to be proposed in the future. Here we discuss how the proposed assisted estimator might be extended to a three-stage problem; similar ideas can be used to extend to more stages.

The observed data on each participant is $\{X_1, A_1, X_2, A_2, X_3, A_3, Y\}$, and a three-stage policy would be $d = (d_1, d_2, d_3)$; the history before each treatment decision time point is $H_1 = X_1, H_2 = (X_1, A_1, X_2), H_3 = (X_1, A_1, X_2, A_2, X_3)$ respectively. Potential outcomes can be conceptualized similarly. Intermediate treatment effects are given by $\mu_3(h_3, a_3) = E[Y(a_1, a_2, a_3)|H_3(a_1, a_2) = h_3] - E[Y(a_1, a_2, 0)|H_3(a_1, a_2) = h_3]$, $\mu_2(h_2, a_2) = E[Y(a_1, a_2, 0)|H_2(a_1) = h_2] - E[Y(a_1, 0, 0)|H_2(a_1) = h_2]$ and $\mu_1(h_1, a_1) = E[Y(a_1, 0, 0)|H_1 = h_1] - E[Y(0, 0, 0)|H_1 = h_1]$.

Suppose we have models for μ_1, μ_2, μ_3 with estimators of the parameters in these models. Then a straightforward estimator for the value of d, V_d, is

$$\begin{aligned}\mathbb{P}_n\big\{&Y - \mu_1(H_1, A_1; \hat{\beta}_1) - \mu_2(H_2, A_2; \hat{\beta}_2) - \mu_3(H_3, A_3; \hat{\beta}_3) \\ &+ \mu_1(H_1, d_1(H_1); \hat{\beta}_1) \\ &+ \frac{I\{A_1 = d_1(H_1)\}}{p_1(A_1|H_1)}\mu_2(H_2, d_2(H_2); \hat{\beta}_2) \\ &+ \frac{I\{A_1 = d_1(H_1)\}}{p_1(A_1|H_1)} \cdot \frac{I\{A_2 = d_2(H_2)\}}{p_2(A_2|H_2)}\mu_3(H_3, d_3(H_3); \hat{\beta}_3)\big\},\end{aligned} \tag{2.6}$$

in which the third line aims to estimate the effect of the treatment specified by d_2 if the first stage treatment were assigned according to d_1, and the fourth line aims to estimate the effect of the treatment specified by d_3 if the first and second stage treatments were assigned according to (d_1, d_2). This estimator is assisted by the intermediate treatment effect functions (in terms of the primary outcome Y), borrowing the inverse-probability-

weight idea to estimate the effect of the policy at subsequent stages.

Similar to the two-stage assisted estimator, one can construct assisted estmimators to replace the fourth line in (2.6), by viewing $\mu_3(H_3, d_3(H_3))$ as the end-of-study primary outcome. More specifically, define $y(\bar{a}_2) = \mu_3(H_3(\bar{a}_2), d_3(H_3(\bar{a}_2))$; this is the intermediate effect of the treatment specified by d_3 on Y, conditional on the history $H_3(\bar{a}_2)$. Define $\nu_2(h_2, a_2) = E[y(a_1, a_2)|H_2(a_1) = h_2] - E[y(a_1, 0)|H_2(a_1) = h_2]$; this is the effect of treatment a_2 relative to the reference treatment at stage two, on the mean of the treatment effect of d_3, among individuals with history h_2. Define $\nu_1(h_1, a_1) = E[y(a_1, 0)|H_1 = h_1] - E[y(0, 0)|H_1 = h_1]$; this is the effect of treatment a_1 relative to the stage one reference treatment, if always followed by the reference treatment at stage two, on the mean of the treatment effect of d_3. Suppose that ν_1, ν_2 can be modeled as $\nu_1(h_1, a_1; \tau_1)$ and $\nu_2(h_2, a_2; \tau_2)$ and that τ_1, τ_2 can be consistently estimated. Estimators of ν_1 and ν_2 can be used to form an assisted estimator of (the fourth line in (2.6) is altered) V_d:

$$\begin{aligned}
\mathbb{P}_n\big\{ & Y - \mu_1(H_1, A_1; \hat{\beta}_1) - \mu_2(H_2, A_2; \hat{\beta}_2) - \mu_3(H_3, A_3; \hat{\beta}_3) \\
& + \mu_1(H_1, d_1(H_1); \hat{\beta}_1) \\
& + \frac{I\{A_1 = d_1(H_1)\}}{p_1(A_1|H_1)} \mu_2(H_2, d_2(H_2); \hat{\beta}_2) \\
& + \mu_3(H_3, d_3(H_3); \hat{\beta}_3) - \nu_1(H_1, A_1; \hat{\tau}_1) - \nu_2(H_2, A_2; \hat{\tau}_2) \\
& + \nu_1(H_1, d_1(H_1); \hat{\tau}_1) + \frac{I\{A_1 = d_1(H_1)\}}{p_1(A_1|H_1)} \nu_2(H_2, d_2(H_2); \hat{\tau}_2)
\end{aligned} \tag{2.7}$$

We have discussed two possible estimators of the value of a three-stage treatment policy; in fact, there is bias-variance-tradeoff between the two estimators, similar to the estimators for the value of a two-stage treatment policy.

2.7 Discussion

Our simulations indicate that the MSE performance of the assisted estimators is not sensitive to misspecification of the model for the intermediate treatment effects. None-the-less to reduce bias, efforts should be made to ensure good model fit in estimating the intermediate treatment effects. Data analysts should make efforts to collect all the time-varying covariates that may moderate the effect of treatment at each stage on the primary outcome and include them in the treatment effects models. Specific subject knowledge, and possibly results from past studies, may provide valuable information for choosing the models.

In this chapter we did not derive the semi-parametrically efficient estimator for policy

value and/or policy contrast. To obtain the most efficient estimator of the policy contrast, one needs to subtract from the influence function of the assisted estimator its projection on all tangent spaces that are orthogonal to the tangent space associated to the policy contrast; this appears difficult because the policy contrast is a functional of a collection of finite or infinite dimensional parameters in the data distribution and the functional is dependent on the specific policies being studied. We plan to investigate this efficiency problem in future research.

The methodology proposed in this chapter is only applicable when a few candidate treatment policies have been pre-specified. When there are more than a few candidate treatment policies, usually one of the candidate treatment policies can be considered as a reference policy, and comparison can be made between any of the remaining policies and this reference policy. In future work, we will also consider a multiple comparison procedures for many treatment policies.

The assisted estimators are based upon the structural nested mean models for continuous primary outcomes. Multiplicative structural mean models [53] and generalized structural mean models [81] have been proposed to deal with non-continuous primary outcomes and non-linear treatment effects. We expect that the assisted estimators can also be extended to deal with more complicated primary outcomes and more complicated underlying interaction between treatments and covariates, with the assistance of these more recent variations of SNMMs.

2.8 Appendix

Proof of Lemma 2.2.1

We first write a telescoping sum of the conditional mean of $Y(A_1, A_2)$. Since $A_1|H_1 = h_1$ has a conditional distribution given by $p_1(\cdot|h_1)$ (*) and $A_2|H_2(A_1) = h_2$ has a conditional distribution given by $p_2(\cdot|h_2)$ (**), we have $E[Y(A_1, 0)|H_2(A_1), A_2] = E[Y(A_1, 0)|H_2(A_1)]$ and $E[Y(0, 0)|H_1, A_1] = E[Y(0, 0)|H_1]$. Thus we have:

$$\begin{aligned} E[Y(A_1, A_2)|H_2(A_1), A_2] = {} & E[Y(A_1, A_2)|H_2(A_1), A_2] - E[Y(A_1, 0)|H_2(A_1), A_2] \\ & + E[Y(A_1, 0)|H_2(A_1)] - E[E[Y(A_1, 0)|H_2(A_1)]|H_1, A_1] \\ & + E[Y(A_1, 0)|H_1, A_1] - E[Y(0, 0)|H_1, A_1] \\ & + E[Y(0, 0)|H_1] \end{aligned}$$

Note that the first line on the right hand side is equal to $\mu_2(H_2(A_1), A_2)$ due to (**) and the third line is equal to $\mu_1(H_1, A_1)$ due to (*); the second line has a conditional mean zero, conditional on (H_1, A_1). Thus we conclude that $E[Y(A_1, A_2) - \mu_2(H_2(A_1), A_2) - \mu_1(H_1, A_1)] = E[Y(0, 0)]$.

For a fixed policy d, the associated potential outcomes are $\{X_1, X_2(d_1), Y(d_1, d_2)\}$. Now let us focus on the telescoping sum of the conditional mean of $Y(d_1, d_2)$. Due to (*), we have $E[Y(A_1, a_2)|X_1, A_1, X_2(A_1)]1_{A_1=a_1} = E[Y(a_1, a_2)|X_1, X_2(a_1)]1_{A_1=a_1}$; this implies $E[Y(A_1, a_2)|X_1, A_1, X_2(A_1)]1_{A_1=d_1(H_1)} = E[Y(d_1, a_2)|X_1, X_2(d_1)]1_{A_1=d_1(H_1)}$ because $d_1(H_1)$ is known given X_1. Moreover, since $d_2(H_2(A_1)) = d_2(H_2(d_1))$ on event $\{A_1 = d_1(H_1)\}$, we have
$E[Y(A_1, d_2)|X_1, A_1, X_2(A_1)]1_{A_1=d_1(H_1)} = E[Y(d_1, d_2)|X_1, X_2(d_1)]1_{A_1=d_1(H_1)}$. Now let $p_1(\cdot|h_1)$ be a degenerate distribution, that concentrates on $d_1(h_1)$, we then conclude that $E[Y(d_1, d_2)|X_1, X_2(d_1)] - E[Y(d_1, 0)|X_1, X_2(d_1)] = \mu_2(H_2(a_1), a_2)|_{a_2=d_2(H_2(a_1)),a_1=d_1(H_1)}$.
Similarly one can show $E[Y(d_1, 0)|X_1] - E[Y(0, 0)|X_1] = \mu_1(H_1, a_1)|_{a_1=d_1(H_1)}$.

Based on the arguments above, we can write:

$$\begin{aligned} E[Y(d_1, d_2)|X_1, X_2(d_1)] = {} & E[Y(d_1, d_2)|X_1, X_2(d_1)] - E[Y(d_1, 0)|X_1, X_2(d_1)] \\ & + E[Y(d_1, 0)|X_1, X_2(d_1)] - E\left[E[Y(d_1, 0)|X_1, X_2(d_1)]|X_1\right] \\ & + E[Y(d_1, 0)|X_1] - E[Y(0, 0)|X_1] \\ & + E[Y(0, 0)|X_1] \end{aligned}$$

and conclude that $E[Y(d_1, d_2)] = E[\mu_1(H_1, a_1)|_{a_1=d_1(H_1)} + \mu_2(H_2(a_1), a_2)|_{a_2=d_2(H_2(a_1)),a_1=d_1(H_1)} + Y(0, 0)]$. Thus $V_d = E\big[Y(A_1, A_2) - \mu_2(H_2(A_1), A_2) - \mu_1(H_1, A_1) + \mu_1(H_1, d_1(H_1)) + \mu_2(H_2(a_1), a_2)|_{a_2=d_2(H_2(a_1)),a_1=d_1(H_1)}\big]$. Finally, consider the transition from the degenerate distribution of A_1 that concentrates on $d_1(H_1)$ to the distribution of A_1 given by $p_1(\cdot|H_1)$, we then can rewrite $E[\mu_2(H_2(a_1), a_2)|_{a_2=d_2(H_2(a_1)),a_1=d_1(H_1)}]$ as
$E\left[\frac{I\{A_1=d_1(H_1)\}}{p_1(A_1|H_1)}\mu_2(H_2(A_1), d_2(H_2(A_1)))\right]$. This completes the proof of Lemma 2.2.1.

Proof of Lemma 2.2.2

First we prove the equality for the second-stage treatment effect. By sequential randomization of A_1, $E[Y(a_1, a_2)|H_2(a_1) = h_2] = E[Y(A_1, a_2)|H_2(A_1) = h_2]$ (note that a_1 is also part of h_2), which is then equal to $E[Y(A_1, A_2)|H_2(A_1) = h_2, A_2 = a_2]$ due to sequential randomization of A_2. Finally by consistency assumption, we conclude that $E[Y(a_1, a_2)|H_2(a_1) = h_2] = E[Y|H_2 = h_2, A_2 = a_2]$, thus the first equalilty for $\mu_2(h_2, a_2)$ holds.

Next we prove the equality for the first-stage treatment effect. By sequential randomization of A_1, $E[Y(a_1,0)|H_1=h_1]=E[Y(a_1,0)|H_1=h_1,A_1=a_1]$, which is then equal to $E[E[Y(a_1,0)|H_2(a_1),A_2=0]|H_1=h_1,A_1=a_1]$ due to sequential randomization of A_2. Re-using sequential randomization of A_1 for the inner conditional mean, this quantity can be written as $E[E[Y(A_1,A_2)|H_2(A_1),A_2=0]|H_1=h_1,A_1=a_1]$. Finally by consistency assumption, we conclude that $E[Y(a_1,0)|H_1=h_1]=E[E[Y|H_2,A_2=0]|H_1=h_1,A_1=a_1]$, thus the equality for $\mu_1(h_1,a_1)$ holds.

The equality of expressing the policy value V_d with the observed data directly follows from Lemma 2.2.1, due to the consistency assumption.

Proof of Lemma 2.2.3

(a) We know from Lemma 2.2.2 that $V_d=E\Big[Y-\mu_2(H_2,A_2)-\mu_1(H_1,A_1)+\mu_1(H_1,d_1(H_1))+\frac{I\{A_1=d_1(H_1)\}}{p_1(A_1|H_1)}\mu_2(H_2,d_2(H_2))\Big]$. Again using the transition from the degenerate distribution of A_1 that concentrates on $d_1(H_1)$ to the distribution of A_1 given by $p_1(\cdot|H_1)$, it is obvious to see that that $E[m(H_1,d_1(H_1))-\frac{I\{A_1=d_1(H_1)\}}{p_1(A_1|H_1)}m(H_1,A_1)]=0,\forall m$ that satisfies the integrable condition.

(b) The asymptotic variance of $\hat{V}_0(d;\hat{\beta})$ is equal to $Var(f_{\beta_0}+C_\varphi^T\varphi)$, and the asymptotic variance of $\hat{V}_m(d;\hat{\beta})$ is equal to $Var(f_{\beta_0}+C_\varphi^T\varphi+g_m)$, where the term $C_\varphi^T\varphi$ comes from the estimation of paramter β in the SNMM (φ is the influence function for β); $f_\beta(h_2,a_2,y)=y-\mu_1(h_1,a_1;\beta_1)-\mu_2(h_2,a_2;\beta_2)+\mu_1(h_1,d_1(h_1);\beta_1)+\omega_{d_1}(h_1,a_1)\mu_2(h_2,d_2(h_2);\beta_2)$, and $g_m(h_1,a_1)=m(h_1,d_1(h_1))-\omega_{d_1}(h_1,a_1)m(h_1,a_1)$. Then the difference in asymptotic variance between $\hat{V}_m(d;\hat{\beta})$ and $\hat{V}_0(d;\hat{\beta})$ is equal to $2Cov(f_{\beta_0}+C_\varphi^T\varphi,g_m)+Var(g_m)$.

We note that for $\hat{\beta}$ in subclass $\mathscr{B}$, $Cov(C_\varphi^T\varphi,g_m)=0$. More specifically, when $\hat{\beta}$ belongs to the subclass $\mathscr{B}$, it is the solution to an estimating equation with the nuisance function $q_1(h_1;\xi)$ chosen optimally (see the review of g-estimators), and one can show that $E[\varphi\cdot g_m]=0$. Thus for those $\hat{\beta}$'s, $Cov(C_\varphi^T\varphi,g_m)=0$, and we only need to focus on $2Cov(f_{\beta_0},g_m)+Var(g_m)$; i.e., the derivation of the optimal m function is the same as the arguments under a known β. For more general $\hat{\beta}$'s, $Cov(C_\varphi^T\varphi,g_m)$ would depend on the estimating equation that produces $\hat{\beta}$ as well as the policy d in a complicated way, thus affecting the choice of optimal m function; for simplicity, in this lemma we assume that $\hat{\beta}$ belongs to $\mathscr{B}$.

In addition, note that $E[(Y-\mu_1(H_1,A_1)-\mu_2(H_2,A_2))\cdot g_m]=0$ by taking the conditional mean with respect to (H_1,A_1). As a result, the optimal choice of m remains the same whether the estimator is for the value V_d, or for the contrast between policy d and a static policy that always assigns treatment 0.

Denote $m^*(h_1, a_1) \equiv E[\mu_2(H_2, d_2(H_2))|H_1 = h_1, A_1 = a_1]$; for simplicity, we write ω_{d_1}, m and m^* in short for $\omega_{d_1}(H_1, A_1), m(H_1, A_1)$ and $m^*(H_1, A_1)$, and write $m \circ d_1$ and $m^* \circ d_1$ in short for $m(H_1, d_1(H_1))$ and $m^*(H_1, d_1(H_1))$. Then, since $E[g_m(H_1, A_1)|H_1 = h_1] \equiv 0$, we could derive that $2Cov(f_{\beta_0}, g_m)+Var(g_m) = E[\omega_{d_1}(2m^* - m)(m \circ d_1 - \omega_{d_1} m)]$. Re-using the fact that $E[m \circ d_1 - \omega_{d_1} m|H_1 = h_1] \equiv 0$ for arbitrary function m, we have $2Cov(f_{\beta_0}, g_m) + Var(g_m) = E[\{(m - m^*) \circ d_1 - \omega_{d_1}(m - m^*)\}^2] - E[(m^* \circ d_1 - \omega_{d_1} m^*)^2]$. Thus the use of function m in the assisted estimator leads to efficiency improvement when $E[\{(m-m^*)\circ d_1-\omega_{d_1}(m-m^*)\}^2] < E[(m^*\circ d_1-\omega_{d_1}m^*)^2]$; in particular, the largest efficiency improvement is achieved when $m(h_1, d_1(h_1)) \equiv m^*(h_1, d_1(h_1))$. Note that, the values of function m at only $(h_1, d_1(h_1))$ have an impact on $\hat{V}_m(d; \hat{\beta})$.

Remark: To get more intuition about when the efficiency improvement that is achieved by using function m can be large, consider a simple scenario where treatments are binary and equally randomized in the data. Then $E[(m^* \circ d_1 - \omega_{d_1} m^*)^2]$, the maximal amount of variance reduction, is equal to $E[m^*(H_1, d_1(H_1))^2]$. This quantity can be large if, under the circumstance that d_1 is followed in stage one, on average the treatment recommended by d_2 at stage two has a large treatment effect.

Proof of Theorem 2.2.4

Under regularity conditions, the following class of functions is Glivenko-Cantelli:

$$\begin{aligned}
&\{y - \mu_2(h_2, a_2; \beta_2) - \mu_1(h_1, a_1; \beta_1) + \mu_1(h_1, d_1(h_1); \beta_1) \\
&\quad + \frac{I\{a_1 = d_1(h_1)\}}{p_1(a_1|h_1)} (\mu_2(h_2, d_2(h_2); \beta_2) - m(h_1, a_1; \alpha_m)) + m(h_1, d_1(h_1); \alpha_m) : \\
&\quad \|\beta_1 - \beta_{10}\| \le \delta, \|\beta_2 - \beta_{20}\| \le \delta, \|\alpha_m - \alpha_m^+\| \le \delta\}
\end{aligned}$$

The theorem then follows from Lemm 2.2.2 and Lemma 2.2.3 by applying Glivenko-Cantelli Theorem to this function class.

Proof of Theorem 2.2.5

Since $\hat{V}_0(d; \hat{\beta})$ is a special case of $\hat{V}_{\hat{m}}(d; \hat{\beta})$, we only prove the asymptotic normality of the latter one. We write $\hat{V}_d$ in short for $\hat{V}_{\hat{m}}(d; \hat{\beta})$ and V_d in short for the true policy value of d.

For ease of notation, define $\omega_{d_1}(H_1, A_1) = \frac{I\{A_1 = d_1(H_1)\}}{p_1(A_1|H_1)}$. Then

$$\begin{aligned}
\sqrt{n}(\hat{V}_d - V_d) =& \sqrt{n}\mathbb{P}_n\big\{Y - \mu_2(H_2, A_2; \hat{\beta}_2) - \mu_1(H_1, A_1; \hat{\beta}_1) + \mu_1(H_1, d_1(H_1); \hat{\beta}_1) \\
&\quad + \omega_{d_1}(H_1, A_1)\mu_2(H_2, d_2(H_2); \hat{\beta}_2)\big\} \\
&- \sqrt{n}P\big\{Y - \mu_2(H_2, A_2; \beta_{20}) - \mu_1(H_1, A_1; \beta_{10}) + \mu_1(H_1, d_1(H_1); \beta_{10}) \\
&\quad + \omega_{d_1}(H_1, A_1)\mu_2(H_2, d_2(H_2); \beta_{20})\big\} \\
&+ \sqrt{n}\mathbb{P}_n\big\{m(H_1, d_1(H_1); \hat{\alpha}_m) - \omega_{d_1}(H_1, A_1)m(H_1, A_1; \hat{\alpha}_m)\big\} \\
&- \sqrt{n}P\big\{m(H_1, d_1(H_1); \alpha_m^+) - \omega_{d_1}(H_1, A_1)m(H_1, A_1; \alpha_m^+)\big\}.
\end{aligned}$$

Under the regularity conditions specified in the theorem, $P\mu_1(H_1, A_1; \beta_1)$, as a function of β_1, is differentiable, and the order of differentiation and integration can be interchanged; moreover $P\dot{\mu}_1(H_1, A_1; \beta_1)$ is continuous in β_1 in a neighborhood of β_{10}. Combined with the fact that $\hat{\beta}_1$ converges in probability to β_{10}, we have:
$\sqrt{n}P\mu_1(H_1, A_1; \hat{\beta}_1) - \sqrt{n}P\mu_1(H_1, A_1; \beta_{10}) = (P\dot{\mu}_1(H_1, A_1; \beta_{10}) + o_p(1))\sqrt{n}(\hat{\beta}_1 - \beta_{10})$.
By similar arguments and the assumptions that $\sqrt{n}(\hat{\beta}_1 - \beta_{10}) = O_p(1)$, $\sqrt{n}(\hat{\beta}_2 - \beta_{20}) = O_p(1)$, we can get:

$$\begin{aligned}
\sqrt{n}(\hat{V}_d - V_d) =& \sqrt{n}(\mathbb{P}_n - P)\big\{Y - \mu_2(H_2, A_2; \hat{\beta}_2) - \mu_1(H_1, A_1; \hat{\beta}_1) + \mu_1(H_1, d_1(H_1); \hat{\beta}_1) \\
&\quad + \omega_{d_1}(H_1, A_1)\mu_2(H_2, d_2(H_2); \hat{\beta}_2)\big\} \\
&+ P\big[\omega_{d_1}(H_1, A_1)\dot{\mu}_2(H_2, d_2(H_2); \beta_{20}) - \dot{\mu}_2(H_2, A_2; \beta_{20})\big]\sqrt{n}(\hat{\beta}_2 - \beta_{20}) \\
&+ P\big[\dot{\mu}_1(H_1, d_1(H_1); \beta_{10}) - \dot{\mu}_1(H_1, A_1; \beta_{10})\big]\sqrt{n}(\hat{\beta}_1 - \beta_{10}) \\
&+ \sqrt{n}\mathbb{P}_n\big\{m(H_1, d_1(H_1); \hat{\alpha}_m) - \omega_{d_1}(H_1, A_1)m(H_1, A_1; \hat{\alpha}_m)\big\} \\
&- \sqrt{n}P\big\{m(H_1, d_1(H_1); \alpha_m^+) - \omega_{d_1}(H_1, A_1)m(H_1, A_1; \alpha_m^+)\big\} + o_p(1).
\end{aligned}$$

Under regularity conditions on $m(h_1, a_1; \alpha_m)$, we can derive

$$\begin{aligned}
&\sqrt{n}P\left(m(H_1, d_1(H_1); \hat{\alpha}_m) - \omega_{d_1}(H_1, A_1)m(H_1, A_1; \hat{\alpha}_m)\right) \\
&\quad - \sqrt{n}P\left(m(H_1, d_1(H_1); \alpha_m^+) - \omega_{d_1}(H_1, A_1)m(H_1, A_1; \alpha_m^+)\right) \\
&\quad = \left(P\dot{m}(H_1, d_1(H_1); \alpha_m^+) - P\omega_{d_1}(H_1, A_1)\dot{m}(H_1, A_1; \alpha_m^+) + o_p(1)\right)\sqrt{n}(\hat{\alpha}_m - \alpha_m^+).
\end{aligned}$$

Since $P\left[m(H_1, d_1(H_1); \alpha_m) - \omega_{d_1}(H_1, A_1)m(H_1, A_1; \alpha_m)\right] \equiv 0$, for all α_m, we derive the following equality, as long as $\sqrt{n}(\hat{\alpha}_m - \alpha_m^+) = O_p(1)$:

$$\begin{aligned}\sqrt{n}(\hat{V}_d - V_d) = \sqrt{n}(\mathbb{P}_n - P)\big\{&Y - \mu_2(H_2, A_2; \hat{\beta}_2) - \mu_1(H_1, A_1; \hat{\beta}_1) + \mu_1(H_1, d_1(H_1); \hat{\beta}_1) \\ &+ \omega_{d_1}(H_1, A_1)\mu_2(H_2, d_2(H_2); \hat{\beta}_2)\big\} \\ &+ P\left[\omega_{d_1}(H_1, A_1)\dot{\mu}_2(H_2, d_2(H_2); \beta_{20}) - \dot{\mu}_2(H_2, A_2; \beta_{20})\right]\sqrt{n}(\hat{\beta}_2 - \beta_{20}) \\ &+ P\left[\dot{\mu}_1(H_1, d_1(H_1); \beta_{10}) - \dot{\mu}_1(H_1, A_1; \beta_{10})\right]\sqrt{n}(\hat{\beta}_1 - \beta_{10}) \\ &+ \sqrt{n}(\mathbb{P}_n - P)\left\{m(H_1, d_1(H_1); \hat{\alpha}_m) - \omega_{d_1}(H_1, A_1)m(H_1, A_1; \hat{\alpha}_m)\right\} + o_p(1)\end{aligned}$$

Define functions indexed by $\beta_1, \beta_2, \alpha_m$ as $f_{\beta_1,\beta_2,\alpha_m}(x_1, a_1, x_2, a_2, y) = y - \mu_2(h_2, a_2; \beta_2) - \mu_1(h_1, a_1; \beta_1) + \mu_1(h_1, d_1(h_1); \beta_1) + \omega_{d_1}(h_1, a_1)\mu_2(h_2, d_2(h_2); \beta_2) + m(h_1, d_1(h_1); \alpha_m) - \omega_{d_1}(h_1, a_1)m(h_1, a_1; \alpha_m)$ and function class $\mathscr{F}_\delta = \left\{\tilde{f}_{\beta_1,\beta_2,\alpha_m} := 1_{\|\beta_1-\beta_{10}\|\le\delta, \|\beta_2-\beta_{20}\|\le\delta, \|\alpha_m-\alpha_m^+\|\le\delta} f_{\beta_1,\beta_2,\alpha_m}\right\}$. Since $\hat{\beta} \xrightarrow{p} \beta_0$ and $\hat{\alpha}_m \xrightarrow{p} \alpha_m^+$, $\sqrt{n}(\mathbb{P}_n - P)f_{\hat{\beta}_1,\hat{\beta}_2,\hat{\alpha}_m} = \sqrt{n}(\mathbb{P}_n - P)\tilde{f}_{\hat{\beta}_1,\hat{\beta}_2,\hat{\alpha}_m} + o_p(1)$. Under regularity conditions, $P\sup|\tilde{f}_{\beta_1,\beta_2,\alpha_m}|^2 < \infty$, thus $P[\tilde{f}_{\hat{\beta}_1,\hat{\beta}_2,\hat{\alpha}_m} - \tilde{f}_{\beta_{10},\beta_{20},\alpha_m^+}]^2 \xrightarrow{p} 0$. By assuming $\sum_{a_1} P\sup_{\|\beta_1-\beta_{10}\|\le\delta}|\dot{\mu}_1(H_1, a_1; \beta_1)|^2 + |\mu_1(H_1, a_1; \beta_1)|^2 < \infty$, $\sum_{a_2} P\sup_{\|\beta_2-\beta_{20}\|\le\delta}|\dot{\mu}_2(H_2, a_2; \beta_2)|^2 + |\mu_2(H_2, a_2; \beta_2)|^2 < \infty$ and $\sum_{a_1} P\sup_{\|\alpha_m-\alpha_m^+\|\le\delta}|\dot{m}_1(H_1, a_1; \alpha_m)|^2 + |m_1(H_1, a_1; \alpha_m)|^2 < \infty$, it can be shown that $\mathscr{F}_\delta$ is a $P-$Donsker class. By Lemma 19.24 in [79], $\sqrt{n}(\mathbb{P}_n - P)\tilde{f}_{\hat{\beta}_1,\hat{\beta}_2,\hat{\alpha}_m} = \sqrt{n}(\mathbb{P}_n - P)f_{\beta_{10},\beta_{20},\alpha_m^+} + o_p(1)$. Hence we have shown that

$$\begin{aligned}\sqrt{n}(\hat{V}_d - V_d) = \sqrt{n}(\mathbb{P}_n - P)&f_{\beta_{10},\beta_{20},\alpha_m^+} \\ &+ P\left[\omega_{d_1}(H_1, A_1)\dot{\mu}_2(H_2, d_2(H_2); \beta_{20}) - \dot{\mu}_2(H_2, A_2; \beta_{20})\right]\sqrt{n}(\hat{\beta}_2 - \beta_{20}) \\ &+ P\left[\dot{\mu}_1(H_1, d_1(H_1); \beta_{10}) - \dot{\mu}_1(H_1, A_1; \beta_{10})\right]\sqrt{n}(\hat{\beta}_1 - \beta_{10}) + o_p(1).\end{aligned}$$

This combined with the assumption that $\hat{\beta}$ is an asymptotically normal estimator for the parameter β in the SNMM, yields that $\hat{V}_d$ is an asymptotically normal estimator for V_d.

Therefore, the asymptotic variance of $\sqrt{n}\left(\hat{V}_d - V_d\right)$ is equal to $E\big(f_{\beta_{10},\beta_{20},\alpha_m^+}(X_1, A_1, X_2, A_2, Y) + P\left[\dot{\mu}_1(H_1, d_1(H_1); \beta_{10}) - \dot{\mu}_1(H_1, A_1; \beta_{10})\right]\varphi_1 + P\left[\omega_{d_1}(H_1, A_1)\dot{\mu}_2(H_2, d_2(H_2); \beta_{20}) - \dot{\mu}_2(H_2, A_2; \beta_{20})\right]\varphi_2\big)^2$, where φ_1, φ_2 are the influence functions for $\hat{\beta}_1$ and $\hat{\beta}_2$.

Proof of Lemma 2.3.1

With arguments similar to part (b) in Lemma 2.2.3, under the assumption that $\hat{\beta}$ belongs to the subclass $\mathscr{B}$ of g-estimators, the difference in asymptotic variance between the esti-

mators with function m and without function m is equal to $2Cov(f_{\tilde{d},\beta_0} - f_{d,\beta_0}, g_{m_{\tilde{d}}} - g_{m_d}) + Var(g_{m_{\tilde{d}}} - g_{m_d})$, where $f_{d,\beta}(x_1, a_1, x_2, a_2, y) = \mu_1(h_1, d_1(h_1); \beta_1) + \omega_{d_1}(h_1, a_1)\mu_2(h_2, d_2(h_2); \beta_2)$, and $g_{m_d}(x_1, a_1) = m_d(h_1, d_1(h_1)) - \omega_{d_1}(h_1, a_1)m_d(h_1, a_1)$.

Define $m^*_d(h_1, a_1) \equiv E[\mu_2(H_2, d_2(H_2))|H_1 = h_1, A_1 = a_1]$ and further define $\Delta_d = m_d - m^*_d$. It can be derived that $2Cov(f_{\tilde{d},\beta_0} - f_{d,\beta_0}, g_{m_{\tilde{d}}} - g_{m_d}) + Var(g_{m_{\tilde{d}}} - g_{m_d}) = E\big[\{(\Delta_{\tilde{d}} \circ \tilde{d}_1 - \omega_{\tilde{d}_1}\Delta_{\tilde{d}}) - (\Delta_d \circ d_1 - \omega_{d_1}\Delta_d)\}^2\big] - E\big[\{(m^*_{\tilde{d}} \circ \tilde{d}_1 - \omega_{\tilde{d}_1} m^*_{\tilde{d}}) - (m^*_d \circ d_1 - \omega_{d_1} m^*_d)\}^2\big]$. The second term in the previous formula is not dependent on m_d or $m_{\tilde{d}}$, and thus the lowest asymptotic variance is obtained when $(\Delta_{\tilde{d}} \circ \tilde{d}_1 - \omega_{\tilde{d}_1}\Delta_{\tilde{d}}) = (\Delta_d \circ d_1 - \omega_{d_1}\Delta_d)$, a.s.. The conclusions of the lemma are implied by this equality.

Proof of Lemma 2.3.2

Define $\hat{\Delta}(d, \tilde{d}) := \hat{V}_{\hat{m}_{\tilde{d}}}(\tilde{d}; \hat{\beta}) - \hat{V}_{\hat{m}_d}(d; \hat{\beta})$; since each assisted estimator for the value is asymptotically normal, $\hat{\Delta}(d, \tilde{d})$ is also asymptotically normal. For notational simplicity, assume that the treatment effect functions can be modeled as linear in unknown parameters, i.e., $\mu_1(h_1, a_1; \beta_1) = \phi_1(h_1, a_1)^T\beta_1$ and $\mu_2(h_2, a_2; \beta_2) = \phi_2(h_2, a_2)^T\beta_2$, where ϕ_t is some feature of (h_t, a_t). Denote $\Delta(d, \tilde{d}) := V_{\tilde{d}} - V_d$.

We first write the estimated value of m functions for each individual explicitly, assuming that m is a working estimate of $E[\mu_2(H_2, d_2(H_2))|H_1 = h_1, A_1 = a_1]$ obtained from least-squares. This assumption is made only for notational simplicity; in practice, more complicated approach can be taken to estimate m if considered necessary. Denote the predictors that are used to estimate m as $D_m = D_m(H_1, A_1)$, then the fitted value of m function for an individual with $(H_1, A_1) = (h_1, a_1)$ would be equal to:

$$m(h_1, a_1; \hat{\alpha}_m) = D_m(h_1, a_1)^T \left(\mathbb{P}_n D_m D_m^T\right)^{-1} \mathbb{P}_n D_m \phi_2(H_2, d_2(H_2))^T \hat{\beta}_2.$$

To simplify the notation, define $\hat{D} := \mathbb{P}_n D_m D_m^T$, $\hat{Z}_d := \mathbb{P}_n D_m \phi_2(H_2, d_2(H_2))$, then under the specified regularity conditions, we have:

$$\begin{aligned}
&\sqrt{n}(\hat{\Delta}(d, \tilde{d}) - \Delta(d, \tilde{d})) \\
&= \sqrt{n}(\mathbb{P}_n - P) f_{d,\tilde{d},\beta_{10},\beta_{20}} \\
&+ P\left[\phi_1(H_1, \tilde{d}_1(H_1)) - \phi_1(H_1, d_1(H_1))\right]^T \sqrt{n}(\hat{\beta}_1 - \beta_{10}) \\
&+ P\left[\omega_{\tilde{d}_1}(H_1, A_1)\phi_2(H_2, \tilde{d}_2(H_2)) - \omega_{d_1}(H_1, A_1)\phi_2(H_2, d_2(H_2))\right]^T \sqrt{n}(\hat{\beta}_2 - \beta_{20}) \\
&+ \sqrt{n}\mathbb{P}_n g_{\tilde{d}_1}^T P[D_m D_m^T]^{-1} P[D_m \phi_2(H_2, \tilde{d}_2(H_2))^T \beta_{20}] \\
&- \sqrt{n}\mathbb{P}_n g_{d_1}^T P[D_m D_m^T]^{-1} P[D_m \phi_2(H_2, d_2(H_2))^T \beta_{20}] + o_p(1),
\end{aligned}$$

in which $f_{d,\tilde{d},\beta_1,\beta_2}(h_2,a_2) = (\phi_1(h_1,\tilde{d}_1(h_1))-\phi_1(h_1,d_1(h_1)))^T\beta_1+\omega_{\tilde{d}_1}(h_1,a_1)\phi_2(h_2,\tilde{d}_2(h_2))^T\beta_2-\omega_{d_1}(h_1,a_1)\phi_2(h_2,d_2(h_2))^T\beta_2$, and $g_{d_1}(h_1,a_1) = D_m(h_1,d_1(h_1)) - \omega_{d_1}(h_1,a_1)D_m(h_1,a_1)$.

Thus if we denote the influence function of the estimator for parameters in the SNMM by (φ_1,φ_2), namely if $\sqrt{n}(\hat{\beta}_1-\beta_{10}) = \sqrt{n}\mathbb{P}_n\varphi_1 + o_p(1)$, and $\sqrt{n}(\hat{\beta}_2-\beta_{20}) = \sqrt{n}\mathbb{P}_n\varphi_2 + o_p(1)$, then the asymptotic variance of $\hat{\Delta}(d,\tilde{d})$ is equal to

$$\begin{aligned}\Sigma_\Delta = Var\Big(& f_{d,\tilde{d},\beta_{10},\beta_{20}} + P\left[\phi_1(H_1,\tilde{d}_1(H_1)) - \phi_1(H_1,d_1(H_1))\right]^T\varphi_1 \\ &+ P\left[\omega_{\tilde{d}_1}(H_1,A_1)\phi_2(H_2,\tilde{d}_2(H_2)) - \omega_{d_1}(H_1,A_1)\phi_2(H_2,d_2(H_2))\right]^T\varphi_2 \\ &+ P[D_m\phi_2(H_2,\tilde{d}_2(H_2))^T\beta_{20}]^T P[D_m D_m^T]^{-1}g_{\tilde{d}_1} \\ &- P[D_m\phi_2(H_2,d_2(H_2))^T\beta_{20}]^T P[D_m D_m^T]^{-1}g_{d_1}\Big).\end{aligned}$$

Next we provide the form of the plug-in estimator $\hat{\Sigma}_\Delta$ for Σ_Δ. Suppose we are able to estimate the influence function of $(\hat{\beta}_1,\hat{\beta}_2)$ evaluated at each data point by $(\hat{\varphi}_1,\hat{\varphi}_2) = (\varphi_1(Z;\hat{\beta},\hat{\xi}),\varphi_2(Z;\hat{\beta},\hat{\xi}))$ (Z includes all the observables from one individual; ξ is the nuisance parameter in estimating SNMM). Define $\hat{\Sigma}_\Delta =$

$$\begin{aligned}\mathbb{P}_n\Big(& f_{d,\tilde{d},\hat{\beta}_1,\hat{\beta}_2} + \mathbb{P}_n\left[\phi_1(H_1,\tilde{d}_1(H_1)) - \phi_1(H_1,d_1(H_1))\right]^T\hat{\varphi}_1 \\ &+ \mathbb{P}_n\left[\omega_{\tilde{d}_1}(H_1,A_1)\phi_2(H_2,\tilde{d}_2(H_2)) - \omega_{d_1}(H_1,A_1)\phi_2(H_2,d_2(H_2))\right]^T\hat{\varphi}_2 \\ &+ \mathbb{P}_n[D_m\phi_2(H_2,\tilde{d}_2(H_2))^T\hat{\beta}_2]^T\mathbb{P}_n[D_m D_m^T]^{-1}g_{\tilde{d}_1} \\ &- \mathbb{P}_n[D_m\phi_2(H_2,d_2(H_2))^T\hat{\beta}_2]^T\mathbb{P}_n[D_m D_m^T]^{-1}g_{d_1}\Big)^2.\end{aligned} \tag{2.8}$$

To show that $\hat{\Sigma}_\Delta$ converges in probability to Σ_Δ, we may use the result that the class of functions involved is a Glivenko-Cantelli class using arguments similar to the proof of Theorem 2.2.4.

Further Details about the Generative Model in Simulation

Here we provide more details about the generative model used in the simulation experiments. $\eta_0(\cdot),\eta_1(\cdot)$ and the variance of ϵ that we use are all based on the by-products of estimating the SNMM with the ExTENd data, using PACS as the primary outcome. More specifically, $\eta_0(\cdot)$ is the main effect of X_1, and it is set to $\eta_0(X_1) = (1, X_{11}, X_{12}, X_{13}, X_{11}X_{12}, X_{11}X_{13}, X_{12}X_{13}, X_{11}^2, X_{12}^2, X_{13}^2)\alpha_0$ where $\alpha_0 = (11.23, 0.3, 2.28, -0.25, 0.24, 0.73, 0.3, -0.74, -0.53, -0.47)$. $\eta_1(\cdot)$ is the main effect of X_2 conditional on (X_1,A_1), and it is set to $\eta_1(X_1,A_1,X_2) = 2(X_{21} - E[X_{21}|X_1,A_1]) - 2(X_{22} - E[X_{22}|X_1,A_1])$. The

standard deviation of ϵ is set to be 5.54.

Assisted Estimator with Missingness in the Outcome

Real data arising from SMART studies normally contains some missing data, due to participants' dropouts or missing some intermediate treatment sessions or research outcome measurement sesssions for various reasons. In this section we describe an approach to the adjustment of the proposed assisted estimator, in the simplified scenario where only the primary outcome variable Y contains missing values. In particular, this requires that patients do not leave the study before the second randomization.

First we denote our data from each participant as $(X_1, A_1, X_2, A_2, R_\pi, R_\pi Y)$, where R_π is an indicator of whether ($R_\pi = 1$) or not ($R_\pi = 0$) the outcome variable Y is observed for this participant. Let $\pi(h_2, a_2) = Pr[R_\pi = 1|H_2 = h_2, A_2 = a_2]$ be the conditional probability of observing Y given history (h_2, a_2). Estimator for the parameters in SNMM can be obtained following a similar least-squares procedure as the one introduced in Section 2.2.2.2:

1. Generalized linear regression to obtain $\pi(H_2, A_2; \hat{\alpha}_\pi)$ as an estimator for $\pi(H_2, A_2)$.

2. Weighted linear regression of Y on $(\phi_2(H_2, A_2) - E[\phi_2(H_2, A_2)|H_2], M_2)$ with weights $R_\pi/\pi(H_2, A_2; \hat{\alpha}_\pi)$ (note that only those observations with non-missing Y get non-zero weights); this regression outputs $\hat{\beta}_2$, which is the vector of the estimated coefficients for $\phi_2(H_2, A_2) - E[\phi_2(H_2, A_2)|H_2]$.

3. Weighted linear regression of $Y - \phi_2(H_2, A_2)^T\hat{\beta}_2$ on $(\phi_1(H_1, A_1) - E[\phi_1(H_1, A_1)|H_1], M_1)$ with weights $R_\pi/\pi(H_2, A_2; \hat{\alpha}_\pi)$ (again only those observations with non-missing Y get non-zero weights); this regression outputs $\hat{\beta}_1$, which is the vector of the estimated coefficients for $\phi_1(H_1, A_1) - E[\phi_1(H_1, A_1)|H_1]$.

Then one can use the following assisted estimator for the policy value:

$$\hat{V}_m(d;\hat{\beta}) = \mathbb{P}_n\Big\{\frac{R_\pi}{\pi(H_2, A_2; \hat{\alpha}_\pi)}Y - \mu_2(H_2, A_2; \hat{\beta}_2) - \mu_1(H_1, A_1; \hat{\beta}_1) + \mu_1(H_1, d_1(H_1); \hat{\beta}_1) + \frac{I\{A_1 = d_1(H_1)\}}{p_1(A_1|H_1)}\Big(\mu_2(H_2, d_2(H_2); \hat{\beta}_2) - m(H_1, A_1; \hat{\alpha}_m)\Big) + m(H_1, d_1(H_1); \hat{\alpha}_m)\Big\}.$$

Additional Results from Simulation 1

Here we present the simulation results for the same set of simulation experiments using an estimator for β that does not belong to the subclass $\mathscr{B}$ (simulation 1*); that is, the

particular nuisance function referred to in the definition of $\mathscr{B}$ is not correctly modeled (in fact, the true nuisance function includes linear terms and second-order terms of X_1; in this simulation, in the estimation of β we only model the nuisance function by the linear terms). Our conjecture is that the assisted estimator $\hat{V}_{\hat{m}_d}(d;\hat{\beta})$ with $\hat{m}_d$ is still slightly more efficient than the assisted estimator with $m_d \equiv 0$. Since $\hat{\beta}$ no longer belongs to $\mathscr{B}$, we do not compare the assisted estimators with an "oracle" estimator. Results are shown in Table 2.5. We found that, as expected, the resulting assisted estimators are unbiased. The MSEs of the two different assisted estimators are similar; yet the one using a good working estimate $\hat{m}_d$ seems to be slightly more efficient in some cases. In general, the results are very similar to the results from the experiments using a $\hat{\beta}$ that belongs to the subclass $\mathscr{B}$.

Simulation of the Relative Efficiency of Assisted Estimators

In this section we further investigate the extent to which the assisted estimator with a working estimate of the optimal m_d improves efficiency over the assisted estimator with $m_d = 0$. We apply the two types of assisted estimators to estimate each of the two policy contrasts: (1) contrast between embedded policies (1, 1, 1) and (0, 0, 0); (2) contrast between embedded policies (1, 1, 0) and (0, 0, 0). Motivated by the remark in the proof of Lemma 2.2.3 about the magnitude of the achievable variance reduction by adopting a good choice of m_d, the experiments are conducted with data from a series of generative models, in which the standardized effect size (SES) of the coordinate in β_2 that corresponds to the A_2 main effect varies from 0.0 to 3.0, and all the other coordinates in β_1 and β_2 have an SES equal to 0.2. We focus on the relative mean squared errors of the assisted estimator with a working estimate of the optimal m_d as compared to that with $m_d = 0$, and for both estimands we plot the trend of the relative mean squared error as the A_2 main effect grows.

The simulation results are shown in Figure 2.2. As expected, the benefit of using a working estimate m_d in the assisted estimator increases when the stage two treatment effect amplifies. However, under the generative model we consider, the A_2 main effect needs to be as large as having an effect size of 1.5 so that the efficiency improvement is about 20%. In practice, we suspect whether such a huge treatment effect would ever be present in a SMART; thus in general using $m_d = 0$ in the assisted estimator may perform just as well as the assisted estimator with a working estimate of m_d. We also notice that, the extent to which using a working estimate of m_d is more efficient than using $m_d = 0$ varies with the estimand.

Table 2.5: Simulation 1*: Statistical properties of the assisted estimators of the contrast between values of policies (1,1,1) and (0,0,0), when $\hat{\beta}$ does not belong to $\mathscr{B}$. Assist = contrast estimator based on $\hat{V}_{\hat{m}_d}(d;\hat{\beta})$ with a working estimate of the optimal m_d. Assist ($m_d = 0$) = contrast estimator based on $\hat{V}_0(d;\hat{\beta})$. The displayed numbers for confidence interval coverage are the coverage proportion $\times$ 100. An Asterisk indicates that the MSE of Assist ($m_d = 0$) is significantly different from MSE of Assist (at 0.05 level).

$N = 100$						
Scenario	Bias / SD		MSE		ASE Coverage	
	Assist	Assist ($m_d = 0$)	Assist	Assist ($m_d = 0$)	Assist	Assist ($m_d = 0$)
(none,none)	0.04	0.04	3.48	3.54*	95.1	95.2
(none,low)	0.02	0.01	4.33	4.41	94.3	94.6
(none,med)	0.03	0.01	3.89	4.28*	95.6	95.5
(low,none)	-0.02	-0.02	3.38	3.39	95	95.4
(low,low)	0.01	0.01	4.15	4.14	95	95.6
(low,med)	0.03	0.02	3.97	4.13*	95.3	95.6
(med,none)	0.05	0.05	3.93	3.98	95.2	95
(med,low)	-0.02	-0.02	4.42	4.43	94.9	94.7
(med,med)	0	0	4.04	4.25*	94.8	95.5
$N = 250$						
Scenario	Bias / SD		MSE		ASE Coverage	
	Assist	Assist ($m_d = 0$)	Assist	Assist ($m_d = 0$)	Assist	Assist ($m_d = 0$)
(none,none)	0.01	0.01	1.36	1.36	93.8	94
(none,low)	0.01	0.02	1.51	1.53*	95.4	95.6
(none,med)	0.03	0.03	1.44	1.53*	94.7	95.2
(low,none)	-0.01	-0.01	1.35	1.35	95.9	95.9
(low,low)	-0.01	-0.01	1.73	1.73	94	94.1
(low,med)	-0.01	-0.01	1.45	1.57*	95.4	94.8
(med,none)	-0.03	-0.03	1.47	1.47	94.9	94.8
(med,low)	0	0.01	1.66	1.68	94.8	95.3
(med,med)	-0.02	-0.01	1.53	1.61*	94.7	95.1

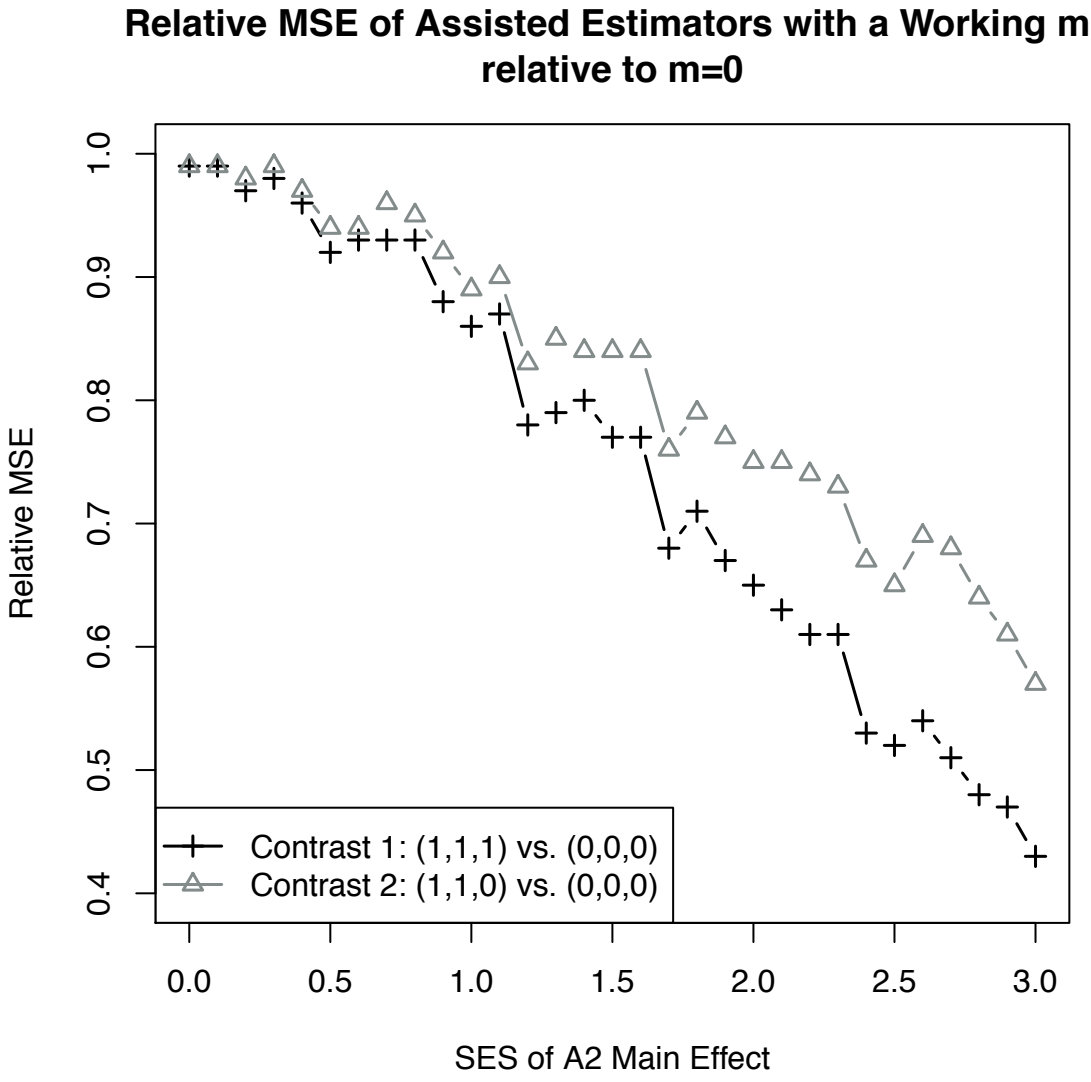

Figure 2.2: Relative mean squared error of the two assisted estimators, as a function of the SES of A_2 main effect in the generative model.

CHAPTER 3

Comparing Dynamic Treatment Regimes Using Repeated-Measures Outcomes: Modeling Considerations in SMART Studies

3.1 Introduction

Dynamic treatment regime (DTR) adapts the type or dosage of treatments to patients' changing needs. DTRs are particularly useful for the treatment of chronic diseases and mental disorders, where the status of the individual is often waxing and waning, or in settings in which no one treatment is effective for most individuals.

An example DTR for improving spoken communication in children with autism spectrum disorder [73, 20] is "Begin with a therapist-delivered behavioral language intervention (BLI) for 12 weeks. At the end of week 12, if a child is a slow responder, augment BLI with an augmentative or alternative communication (AAC) approach, most often a speech-generating device; otherwise, if the child shows early signs of response, continue with the first stage BLI for an additional 12 weeks." See [20] for more details concerning the intervention procedures and the definition of early signs of response at week 12.

This chapter focuses on statistical methods for comparing DTRs on the basis of a repeated-measures outcome observed across the multiple stages of treatment in a sequential, multiple assignment, randomized trial (SMART; [25, 24, 40]). In the context of the autism example, a researcher may be interested in comparing two DTRs, say, based on the trajectory of the number of socially communicative utterances collected at baseline and weeks 12, 24 and 36.

Study features that are unique to SMARTs make repeated-measures modeling a challenge. In this setting, repeated-measures models must account appropriately for (i) the temporal ordering of treatments relative to outcome measurement occasions and (ii) the fact that participants may transition from one stage of treatment to the next at different time

points. In this chapter we discuss how to accommodate the features of various forms of SMART designs in modeling repeated-measures outcomes. For illustration, we use data from three SMART case studies in autism, child attention deficit hyperactivity disorder (ADHD), and adult alcohol dependence. We do this because each case study presents a progressively more complex SMART design. ADHD SMART was conducted to investigate the effect of sequential implementation of two forms of ADHD treatment: medication and behavioral modification. The unique feature of ADHD SMART is that participants were evaluated early response status at each month after the first two months, if they had not yet become slow responders; they transitioned to stage two and were re-randomized immediately after being classified as slow responders. Thus, for different participants, measurements taken at the same calendar time, say, measurement taken at the end of month 4, may belong to different treatment stages. However, measurements were taken monthly for 8 months, for all the participants. On the other hand, ExTENd is a SMART trial of alcohol dependence in which all participants went through two stages of treatments. They had varying lengths of stage one treatments depending on initial randomization and participant's response status to first-line treatment, but roughly the same length of stage two treatment (4 months).

A secondary contribution is the extension and application of an inverse probability of treatment weighted (IPTW) estimator for the repeated-measures models. The IPTW estimator was earlier introduced for estimating time-varying treatment effect in observational studies and for the evaluation of one specific DTR (i.e., marginal mean model). Later it was developed and illustrated for the Marginal Structural Models (MSMs) that compare DTRs based on an end-of-study outcome [47, 3, 50, 46]. On the other hand, there is also works concerning MSMs for the marginal effect of time-varying treatments or static treatment regimes (rather than a DTR) on a repeated-measures outcome [60, 13, 61]; these papers discussed the possibility of using a working covariance matrix in the estimator to improve the statistical efficiency. In this chapter, we describe an easy-to-implement estimator for the repeated-measures model that generalizes this estimator to the comparison of repeated measures among DTRs, that permits analysts to efficiently use the data to estimate the mean trajectories associated to all embedded DTRs simultaneously, and to take advantage of within-person correlations in repeated measures with an attempt to improve statistical efficiency.

This chapter is organized as follows. In Section 3.2, we give a brief review about the existing works that investigate the effect of time-varying treatment and regime on repeated-measures outcomes. In Section 3.3, we describe the three SMART studies that will be used to illustrate the proposed modeling principles and methodology. In Section 3.4, we present

and discuss general principles for modeling repeated-measures outcomes in a SMART and illustrate these principles with the three SMART studies. A weighted-and-replicated estimator for the parameters in the repeated-measures marginal model is proposed in Section 3.5. In Section 3.6 we present the data analysis results for the three SMART studies. In Section 3.7 we report results of simulation studies that investigate both the modeling and estimation aspects of the methodology. Finally, a discussion, including other possible ideas for modeling repeated-measures outcomes from a SMART, is presented in Section 3.8.

3.2 Existing Works Regarding Repeated-Measures Outcome

[60] presented MSMs for a repeated-measures outcome under a particular treatment sequence. This paper discussed a class of IPTW estimators for such models that generalize the GEE methodology; in particular, in some discussions about the practical choice of an estimator in this class of IPTW estimators, Robins recommended using a working covariance matrix of the repeated measures conditional on the treatment sequence, to achieve reasonably high efficiency. [13] more deeply investigated MSMs for repeated measures that correspond to pre-specified static treatment regimes, with the motivation that standard methodology produces biased causal results when there are time-varying confounders that are predicted by previous treatments. Interestingly, we note that there are two different weighting strategies proposed in these two papers. [60] uses an identical weight for all time points for an individual, which is equal to the inverse probability of the individual receiving the assigned/observed entire treatment sequence. On the other hand, [13] uses time-varying weights for the repeated measures of each individual, and the weight for each time point is the cumulative inverse probability of treatment up to that specific time point. The first weighting scheme is one that is applicable for general outcome types, and it allows for using a working covariance matrix in the estimator without impairing the consistency of the estimator. The second weighting scheme is more specific to the methodology for analyzing repeated measures. When the working covariance matrix is taken to be independence and there are no additional augmentation terms in the estimator, there have been empirical results showing that it is more efficient than the first weighting scheme. In the methodology proposed in this chapter, we choose to adopt the first weighting scheme and allow for a non-independence working correlation structure.

[36] compares repeated-measures outcome among DTRs by focusing on repeated measures that occur only after all the re-randomizations. For example, if the data arises from

a two-stage SMART, the repeated measures that are analyzed in this paper would be those measured after the second randomization. Thus it has the limitation that no conclusion can be made about the effect of the treatment regimes on the repeated measures over a longer period of time that begins from the initial treatment stage. They propose methods that are based on mixed models and multiple imputations. The working paper by Li proposes a methodology to compare DTRs in terms of repeated measures that span through multiple treatment stages. In this paper, the author focuses on the estimation perspective of the repeated-measures analysis under various SMART designs, and discusses in detail the sample size calculation based on such estimators; modeling considerations specific to the analysis of repeated measures from SMART studies are not as emphasized as the estimation component.

3.3 Three SMART Studies for Case Study

In a SMART participants proceed through multiple treatment stages, and at each treatment stage the participant may be randomized to one of several treatment options available at that stage. Often, subsequent randomized treatment options in a SMART are restricted depending on the participant's response to prior treatment.

In this section we describe the three SMART studies that we use for illustration in this chapter: the autism, ADHD and ExTENd studies. These designs vary in complexity, with the autism study being the least complex and the ExTENd study being the most complex of the three. The complexity in study design is in terms of the number of DTRs that are embedded in the design and the number of time points at which participants can transition from one treatment stage to another. In Section 3.4, we will discuss how these varying design features have implications on the choice of models for repeated measures arising from the SMART studies.

Figure 3.1 provides the design of the autism SMART (C. Kasari, P.I.; [20]), for the treatment of minimally verbal children with autism, aged 5 to 8 years. In this SMART, at the first stage children were randomized to BLI or BLI+AAC. This stage lasted for 12 weeks for all children. After 12 weeks, children were classified as either early responders or slow responders and made the transition to the second stage. In the second stage, early responders continued with the treatments that were assigned in the first stage; slow responders to initial BLI+AAC received intensified BLI+AAC (more sessions per week), and slow responders to initial BLI were randomly assigned to either intensifying the initial treatment (BLI) or to augmenting the initial treatment with AAC (i.e., BLI+AAC). The second stage treatment lasted for 12 weeks.

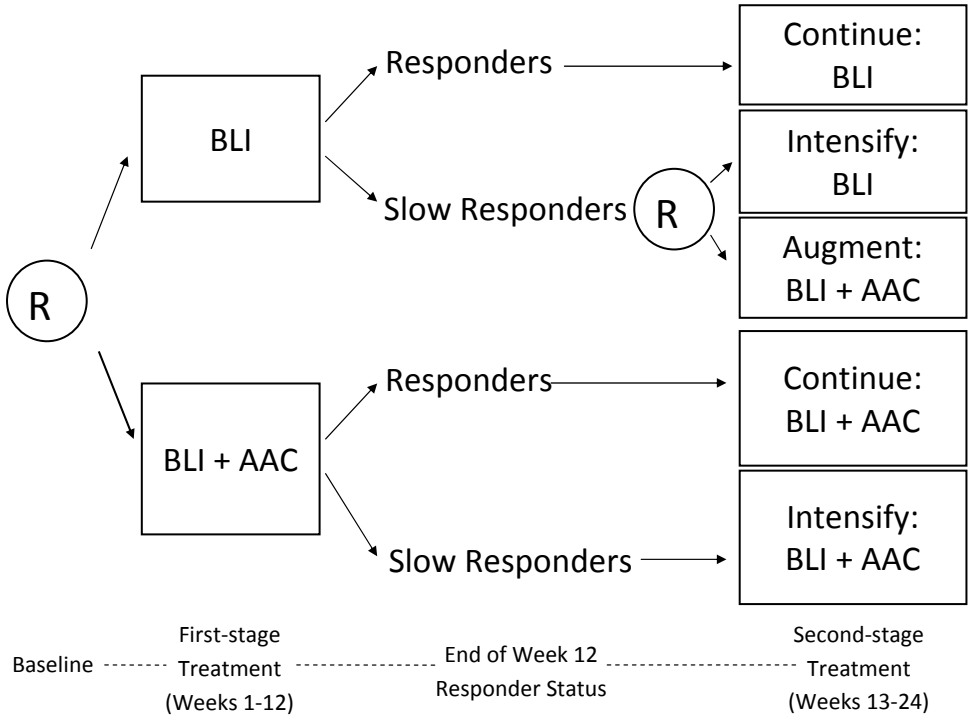

Figure 3.1: A SMART study for developing a DTR for children with autism who are minimally verbal. R = randomization. BLI = behavioral language intervention. AAC = augmentative or alternative communication approach.

SMARTs, such as the one shown in Figure 3.1, have a set of DTRs embedded within them, by design. The phrase "embedded by design" is used to express that the variables used to "tailor" the treatment in these DTRs can only be those used in the SMART to restrict randomized treatment options. These embedded DTRs are pre-determined in the design phase of the study; all participants are expected to provide data that is consistent with at least one of these DTRs. These embedded DTRs correspond to different strategies to managing the disease/disorder over time.

In the autism SMART there are three embedded two-stage DTRs; they are listed in Table 3.1. These three embedded DTRs reflect different strategies towards improving spoken communication skills, with varying levels of the provision of AAC (in the context of BLI). For example, DTR#1 uses AAC only for those who show slow response to initial BLI; in contrast, DTR#3 uses AAC from entry to study for every participant. Note that some

Table 3.1: Embedded DTRs in the autism SMART.

	Label	Treatment decision rule
DTR #1	$(1, -1)$	Begin treatment with BLI for 12 weeks. At the end of week 12, if the child does not show early signs of response, augment BLI with AAC for 12 weeks. Otherwise, continue with BLI for another 12 weeks.
DTR #2	$(1, 1)$	Begin treatment with BLI for 12 weeks. At the end of week 12, if the child does not show early signs of response, intensify BLI for 12 weeks. Otherwise, continue with BLI for another 12 weeks.
DTR #3	$(-1, \cdot)$	Begin treatment with BLI+AAC for 12 weeks. At the end of week 12, if the child does not show early signs of response, intensify BLI+AAC for 12 weeks. Otherwise, continue with BLI+AAC for another 12 weeks.

participants in a SMART have treatment sequences that are consistent with more than one DTR. For example, early responders to BLI have a treatment sequence that is consistent with both DTR#1 and DTR#2.

In the analysis of the autism SMART data, we focus on the repeated measures of the number of socially communicative utterances at baseline, week 12, 24 and 36. The repeated measure at baseline is prior to the first-stage treatment; the repeated measure at week 12 is prior to the second-stage treatment.

Figure 3.2 shows the design of the ADHD SMART for the treatment of children (aged 5 to 13 years with mean of 8 years) with ADHD (W. Pelham, P.I.). In this SMART, at the first stage children were randomly assigned to begin with low-intensity behavioral modification (BMOD) or with low-dose medication (MED; methylphenidate). Starting at the end of month two, children were assessed monthly for response/non-response to the initial treatment. See [28] and [43] for more details concerning the definition of response/non-response. Children who met the criteria for non-response were immediately re-randomized to either an intensified version of the initial treatment (INT) or to augmenting the initial treatment with the alternative treatment (MED+BMOD). Children who continued to respond remained on their initial treatment. Treatment duration was eight months in total for all children in the study.

The ADHD SMART has four embedded DTRs, as a result of two treatment options in the initial randomization and two treatment options in the re-randomization of non-responders. The ADHD SMART differs from the autism SMART in that the duration of stage one varied among participants. Those who met the non-response criteria at later time points transitioned to the second treatment stage later during the study, and the duration of

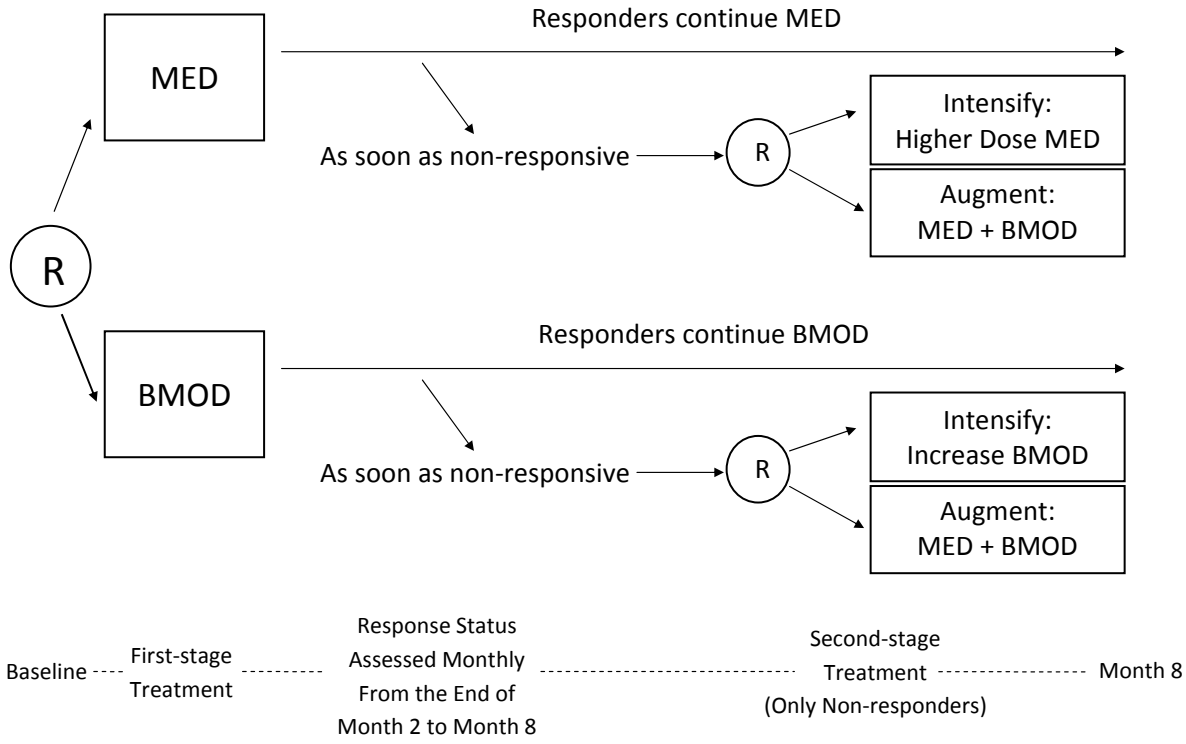

Figure 3.2: A SMART study for developing a DTR for children with attention deficit/hyperactivity disorder. R = randomization. MED = medication. BMOD = behavioral modification.

the first treatment stage was an outcome of the initial treatment. Those who continued to respond to the initial treatment did not transition to the second stage. In the analysis of the ADHD SMART data, we focus on the repeated measures of classroom performance rating that is part of the teachers' Impairment Rating Scale (IRS). This measure is available at the end of each month until the end of the study (i.e., month eight). Note that, for different participants, the classroom performance rating at a certain time point may belong to different treatment stages.

Figure 3.3 shows the design of a third SMART study, the ExTENd SMART aiming to develop a DTR for individuals with alcohol dependence. This study was used as an illustrating example for the assisted estimator in Chapter 2, where more details about this study were provided.

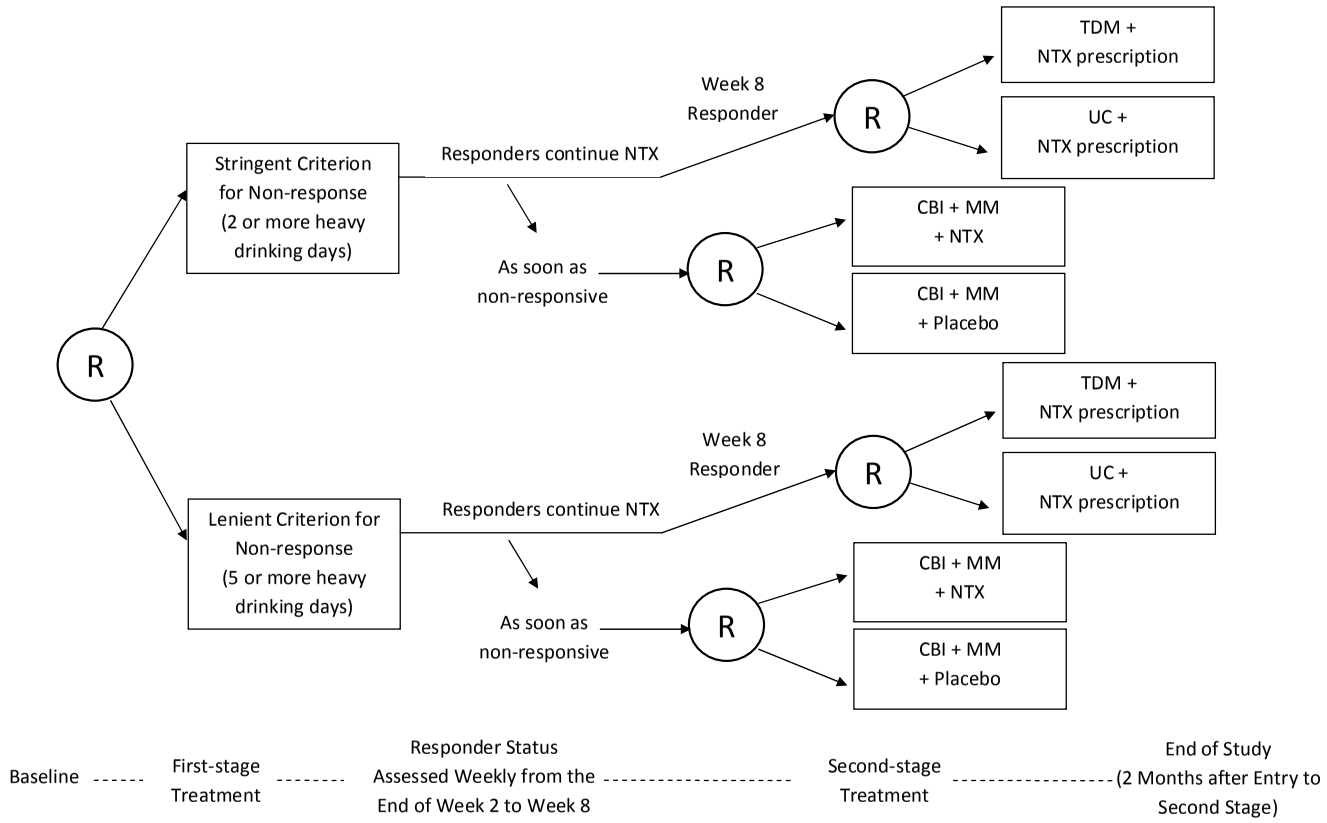

Figure 3.3: A SMART study for developing a DTR for adults with alcohol dependence. R = randomization. NTX = Naltrexone. TDM = telephone disease management. UC = usual care. CBI = combined behavioral interventions. MM = medical management.

The ExTENd SMART design is more complex than the autism and ADHD studies. Similar to the ADHD study, in ExTENd the duration of stage one varied among participants. Specifically, non-responders transitioned to stage two at any of a variety of weeks prior to week eight whereas all responders transitioned to stage two at week eight. However, in ExTENd both responders and non-responders to the initial treatment were re-randomized to subsequent treatment options. As a result of the initial randomization and the re-randomization , the ExTENd SMART has eight embedded DTRs. In the analysis of the ExTENd SMART data, we focus on the repeated measures of an alcohol craving scale. This craving measurement is available weekly, from baseline to the end of the study. Note that the entire duration of the study varies among the participants, depending on how they respond to the initial treatment. Thus we analyze the repeated measures of alcohol craving scale from baseline to week 16; these measurement occasions are applicable to all participants in the study.

3.4 Repeated-Measures Marginal Model

In this section we develop marginal models for comparing the embedded DTRs in a SMART based on repeated measures. For simplicity, we focus on two-stage SMARTs; all ideas can be extended readily to SMARTs with more than two stages. By examining the three SMART studies introduced above, we will illustrate modeling considerations by varying degrees of complexity in SMART design. There has been work that compares DTRs by focusing on repeated measures that occur only after all the re-randomizations [36]; in our work we allow the repeated measures to span across multiple treatment stages.

We label each embedded DTR in a two-stage SMART by the pair (a_1, a_2), where a_j is used to denote a treatment option at stage j. For example, in the autism SMART, we let $a_1 = 1$ denote BLI and let $a_1 = -1$ denote BLI+AAC. We let $a_2 = 1$ denote assigning intensified BLI to slow responders to first stage BLI and let $a_2 = -1$ denote assigning BLI+AAC to slow responders to first stage BLI. Note that in the autism SMART, a_2 is nested within $a_1 = 1$ because only slow responders to BLI were re-randomized. See Table 3.1 for the labels of all three embedded DTRs in the autism study.

X denotes baseline, pre-randomization covariates, such as age, gender and ethnicity. In all models below, the variables in X are mean-centered to facilitate model interpretations. Y_t denotes the repeated-measures primary outcome that is of scientific interest, observed at time t, $t \in \mathcal{T}$. In the autism study, Y_t is the number of socially communicative utterances at week $t = 0$ (baseline), 12, 24, 36. For this outcome, higher values of Y_t are more favorable.

$E_{(a_1,a_2)}[Y_t|X]$ is the marginal mean of the repeated-measures outcome Y_t under the em-

bedded DTR defined by (a_1, a_2), conditional on the baseline variable X. Note that under the potential outcome framework [45, 68], this is the mean of the repeated-measures outcomes had all participants followed the DTR (a_1, a_2). Thus a model $\mu_t(X, a_1, a_2; \beta)$ for $E_{(a_1,a_2)}[Y_t|X]$ is a repeated-measures marginal structural mean model [63, 46]; in this chapter for conciseness we will call $\mu_t(X, a_1, a_2; \beta)$ a marginal mean model. The primary focus of this chapter is on developing parametric models $\mu_t(X, a_1, a_2; \beta)$ for $E_{(a_1,a_2)}[Y_t|X]$ under various forms of SMART designs. We will also discuss the estimation of the unknown parameter β.

3.4.1 A Traditional yet Naïve Approach to Modeling Repeated Measures in a SMART

To appreciate the need to accommodate the specific features of a SMART design in repeated-measures modeling, we first consider using a traditional approach to comparing the mean trajectories between two DTRs in the autism SMART. For simplicity, suppose we are interested in comparing DTR#1 (labeled (1, -1) in Table 3.1) versus DTR#2 (labeled (1, 1) in Table 3.1) using only data from children who began with BLI $(a_1 = 1)$. A traditional model in this case might be

$$E_{(a_1,a_2)}[Y_t|X] = \eta^T X + \beta_0 + \beta_1 t + \beta_2 1_{a_1=1,a_2=1} t.$$

This is a traditional approach in that it is often used in the analysis of two-arm RCTs. In this model, the trajectories associated with the two DTRs are modeled as two straight lines that start with the same intercept at $t = 0$: the marginal mean of Y_t under DTR $(1, -1)$ is $(\beta_0 + \beta_1 t)$, whereas the marginal mean of Y_t under DTR $(1, 1)$ is $(\beta_0 + (\beta_1 + \beta_2)t)$. In this example of a traditional approach, therefore, the difference between the marginal mean trajectories is given by the single parameter β_2. This model will incur bias if either one of the two DTRs does not have a linear mean trajectory. However, in a study such as the autism SMART, it may be important to accommodate a possible deflection at week 12 in the mean trajectory because this is the point at which the treatment is modified for slow responders. Further, since neither participants nor staff were aware of the randomly assigned second-stage treatment during the first stage of treatment (this is a typical feature of SMART designs), these two DTRs should not differ, on average, from $t = 0$ to $t = 12$. An example of an improved model is presented in the next section. In Section 3.7 we investigate, via simulations, the bias that occurs when adopting a traditional slope or quadratic model to analyze repeated measures from a SMART.

In general, depending on what treatment is practically administered in one stage under each of the embedded DTRs, it might be possible that more than one DTR should share the same marginal mean until some critical decision time point, where participants advance to a new treatment stage. Typically, this constraint is valid in the analysis of SMART due to the lack of anticipatory behavior resulting from sequential randomization of the treatments.

In addition, when it is no longer reasonable to adopt simple models such as the slope model above, the comparison between DTRs based on a repeated-measures outcome would require alternative estimands. In the following subsections, we discuss (a) modeling considerations for $\mu_t(X, a_1, a_2; \beta)$ for various forms of SMART designs, and (b) options for estimands in the comparison of the embedded DTRs in a SMART.

3.4.2 Repeated-Measures Modeling Considerations: The Autism Example

As noted earlier, modeling of a repeated-measures outcome arising from a SMART should be guided by two key principles: (a) properly accommodate the timing of repeated measures in relation to the treatment stages in a SMART; and (b) properly accommodate the restrictions applied on the randomizations by design. The autism SMART provides a relatively simple example to illustrate these modeling principles.

In the autism SMART, all participants had the same duration of stage one treatment (12 weeks) and stage two treatment (12 weeks), and they all advanced to stage two after week 12. Additionally, only slow responders to BLI were re-randomized.

The primary outcome, the number of socially communicative utterances, was measured on four occasions. Baseline measurement Y_0 was pre-treatment; Y_{12} was measured right before the second treatment stage (re-randomization, if applicable, happened right after Y_{12}); Y_{24} was measured at the end of treatment; Y_{36} was measured post treatment and treatment ended at the end of week 24. Since all participants transitioned at $t = 12$, one approach to modeling the repeated measures in the autism SMART is using a continuous, piecewise marginal model with a knot at week 12. For example, consider the following marginal model for Y_t:

$$\begin{aligned} E_{(a_1,a_2)}[Y_t|X] = \eta^T X + \beta_0 + 1_{t\leq 12}\{\beta_1 t + \beta_2 t a_1\} \\ + 1_{t>12}\{12\beta_1 + 12\beta_2 a_1 + \beta_3(t-12) + \beta_4(t-12)a_1 + \beta_5(t-12)1_{a_1=1}a_2\}, \end{aligned} \tag{3.1}$$

where the unknown parameters $\beta = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5)$ model the effect of the three em-

bedded DTRs over time; and η captures the association between the time-varying outcome and baseline covariates X; β are of primary interest.

This example model entails two main restrictions. The first restriction is that Y_0 is modeled to have the same marginal mean for all three embedded DTRs. This is a common restriction used in the analysis of longitudinal randomized trials [30] since, by design, treatment groups are not expected to differ at baseline (prior to randomization). The second restriction is that the marginal mean trajectory is assumed to be the same between embedded DTRs $(1, 1)$ and $(1, -1)$ until week 12. This restriction is unique to SMARTs. It is consistent with the study design, in that (1) these two DTRs are identical up to week 12 and (2) re-randomization to second stage treatment does not occur until week 12 (i.e., there can be no expectancy or anticipatory effects due to knowledge of second stage treatments during stage one).

For simplicity, the example model above assumes a piecewise linear trend. In practice, a quadratic mean trajectory (or some other trend) may be more appropriate.

Note that in the model proposed above, we implicitly assume that the marginal mean of Y_t under each of the embedded DTRs is continuous in time. In practice, it might be more reasonable to allow for a "jump", i.e., an abrupt change, in the marginal model when there is treatment stage transition, because it is possible that actions such as informing the patients their initial response status or informing the patients their re-randomized treatments may have an momentary effect on the outcome. However, learning a more flexible model like this requires more frequent measurement of the outcome. Given the scheme of outcome measurement in the autism SMART, we choose to impose the continuity assumption on the marginal mean model.

3.4.3 Repeated-Measures Modeling Considerations: The ADHD Example

In analyzing the ADHD SMART, we focus on comparing the four embedded DTRs based on the repeated measures of classroom performance rating measured on eight occasions – at the end of each month of the study (i.e., $Y_1, ..., Y_8$). Note that unlike in the autism SMART, the repeated-measures outcome in the ADHD SMART is unavailable at baseline. This outcome is coded so that higher values are more favorable. Each of the four embedded DTRs is labeled by a pair (a_1, a_2). Let $a_1 = 1$ denote starting with low-intensity BMOD and let $a_1 = -1$ denote starting with low-dose MED. Let $a_2 = 1$ denote intensifying the initial treatment for non-responders and let $a_2 = -1$ denote augmenting the initial treatment with the alternative treatment for non-responders.

As discussed previously, the design of the ADHD SMART study is more complex relative to the autism study. The duration of the first treatment stage varied among participants; it could be as short as two months (for the children who became non-responders at the end of month two), or as long as eight months (for the children who continued to respond throughout the entire study). This has implications for modeling the marginal mean under a DTR in that, for a fixed $t > 2$, the marginal mean of Y_t is a weighted average of the mean for participants who have transitioned and the mean for participants who have yet to transition; as a result, there may be deflections in the marginal mean at any given month, starting at month 2 ($t = 2$) and ending at month 7 ($t = 7$).

Additionally, the initial treatment (BMOD versus MED) had an impact on participants' performance, which determined whether or when the participants transitioned to the second stage as non-responders. For example, among the 75 participants who were assigned to MED initially, only 19 transitioned to stage two (as non-responders) at month two; whereas among the 75 participants who were assigned to BMOD initially, 36 transitioned to stage two (as non-responders) at month two. Therefore, we may allow the pattern of deflection in the mean trajectory to differ between DTRs that start with BMOD and those starting with MED (see the exploratory plot in the appendix).

Based on the discussions above, as well as exploratory analysis aimed at refining the modeling assumptions, we propose to model the repeated measures from the ADHD study as shown below:

$$\begin{aligned} E_{(a_1,a_2)}[Y_t|X] =& \eta^T X + \beta_0 + \beta_1 a_1 + 1_{a_1=1} 1_{t\leq 2} \beta_2 (t-1) \\ &+ 1_{a_1=1} 1_{t>2} (\beta_2 + 1_{a_2=1}(\beta_3 (t-2) + \beta_4 (t-2)^2) + 1_{a_2=-1}\beta_5 (t-2)) \\ &+ 1_{a_1=-1} 1_{t\leq 3} \beta_6 (t-1) \\ &+ 1_{a_1=-1} 1_{t>3} (2\beta_6 + 1_{a_2=1}\beta_7 (t-3) + 1_{a_2=-1}\beta_8 (t-3)). \end{aligned} \tag{3.2}$$

Here, the DTR (BMOD, BMOD+MED) (i.e., $(a_1, a_2) = (1, -1)$) is assumed to have a piecewise linear trajectory with a knot at $t = 2$, whereas (BMOD, INT) (i.e., $(a_1, a_2) = (1, 1)$) has the same mean trajectory as (BMOD, BMOD+MED) until $t = 2$ and then develops a quadratic trajectory. The two DTRs that begin with MED are assumed to have piecewise linear trajectories with a knot at $t = 3$ and they share the same mean trajectory until $t = 3$.

3.4.4 Repeated-Measures Modeling Considerations: The ExTENd Example

The greater complexity in the ExTENd SMART necessitates more careful modeling considerations. In analyzing the ExTENd SMART, we focus on comparing the eight embedded DTRs based on a repeated-measures outcome of alcohol craving. Alcohol craving was collected on 17 occasions: at baseline (Y_0) and the end of each week for 16 weeks ($Y_1, ..., Y_{16}$). This outcome is re-coded so that higher values are more favorable. Each of the eight DTRs is denoted by a triplet (a_1, a_{2R}, a_{2NR}), where a_1 is used to denote whether the stringent definition or the lenient definition of early non-response is adopted, a_{2R} is used to denote a treatment option for responders at stage two, and a_{2NR} is used to denote a treatment option for non-responders at stage two.

The transition time to the second treatment stage ranged from the end of week two to the end of week eight. As a result, similar to the ADHD study, for a fixed t, Y_t may come from different treatment stages for different participants. In addition, note that DTRs that begin with the same a_1 might differ only in a_{2R} (how responders are treated in the second stage), only in a_{2NR} (how non-responders are treated in the second stage), or both. The impact of differing a_{2NR} can take place from the end of week two (non-responders could start to transition to stage two as early as the end of week two); however the impact of differing a_{2R} can only take place from the end of week eight (responders could only transition to stage two at the end of week eight).

Because of the features illustrated above, and given the relatively frequent repeated measures, we do not model each of the mean trajectories by simple parametric form; instead, we adopt flexible spline-based models with constraints that are consistent with the SMART design. First, we allow two DTRs that differ only in a_{2NR} to start to differ in the mean trajectories after $t = 2$, because participants could become non-responders and, therefore, receive salvage treatment options specified by a_{2NR} on or after week two. Second, we allow two DTRs that differ only in a_{2R} to start to differ in the mean trajectories after $t = 8$, because on week eight participants could become responders and, therefore, receive the maintenance treatment options specified by a_{2R}. Aside from forcing all DTRs to have the same mean of Y_0 and these two constraints, we allow the trajectories of the DTRs to be regression splines. In the appendix we provide additional details about building the regression splines model based on these considerations.

Table 3.2: Design features of ExTENd study and their implications on the repeated-measures modeling.

Design features	Implications for repeated-measures modeling
Randomization is (or should be) stratified on baseline measurements; there is no difference in anticipatory effect among the eight DTRs.	Trajectories of all the eight DTRs have the same intercept.
Patients became responders only if they stayed in stage one for eight weeks without meeting the assigned criterion for non-response. There can be no expectancy effects due to knowledge of second stage treatments during stage one.	A pair of DTRs that only differ in a_{2R} (the second-stage treatment for responders) should share the same trajectory until the end of week eight, and may differ from then on.
Patients transitioned to stage two as non-responders as early as week two. There can be no expectancy effects due to knowledge of second stage treatments during stage one.	A pair of DTRs that only differ in a_{2NR} (the second-stage treatment for non-responders) should share the same trajectory until the end of week two, and may differ from then on.

3.4.5 Estimands

In the repeated-measures analysis of SMARTs, a variety of interesting estimands are possible for the comparisons among embedded DTRs. Here, we present two that are clinically important and easy to communicate: change score comparisons and area under curve (AUC). The first approach, change score comparisons, measures the differences among embedded DTRs in terms of change in response from t_1 to t_2. A change score estimand is $\Delta_{t_1,t_2} = E_{(a_1,a_2)}[Y_{t_2} - Y_{t_1}] - E_{(a_1^*,a_2^*)}[Y_{t_2} - Y_{t_1}]$, where (a_1, a_2) and (a_1^*, a_2^*) are two embedded DTRs. In the autism example, a change score comparison from week 0 to week 36 compares the embedded DTRs in terms of the mean increase from baseline to the end of follow-up in the number of socially communicative utterances.

The second approach, AUC, summarizes the cumulative amount of Y_t within a time range (t_1, t_2); it provides an alternative single number summary of the overall mean trajectory under each embedded DTR. In the autism study, the AUC of Y_t from $t = 0$ to $t = 36$ for a specific embedded DTR has a clinically relevant interpretation as the average total number of socially communicative utterances from $t = 0$ to $t = 36$ under this DTR.

Note that if the investigator believes that the change score comparison is more relevant for the subject-specific area, one may choose to adopt a statistical methodology that compares the DTRs in terms of only the end-of-study outcome, which ideally takes advantage of the data of other time-varying variables (including the Y_t's at earlier time points) in some way (e.g., via estimating the weights in the weighted-and-replicated estimator using the co-

variates). However, these methodologies cannot simultaneously answer research questions that are related to other estimands such as AUCs. The methodology we propose models the entire trajectory of the marginal mean, which can later be used to address multiple research questions.

The AUC is a more informative summary of the marginal mean trajectory than the change score, because it captures not only change from the start to the end point, but also characteristics of the progression in the mean outcome during the period. In the data analysis, we mainly report the AUC for the embedded DTRs.

3.5 Estimator for Repeated-Measures Marginal Model

3.5.1 Observed Data

For simplicity, we present the proposed estimator for the repeated-measures marginal model with the autism example. Details about how this estimator is implemented to analyze the ADHD and ExTENd SMARTs can be found in the appendix.

The structure of the data is as follows. For individual i ($i = 1, ..., N$), we observe $X_i, A_{1,i}, R_i, A_{2,i}$ and $Y_{t,i}, t \in \mathcal{T}$. X includes a set of mean-centered baseline covariates; A_1 denotes the first-stage treatment to which an individual is randomized; R is the indicator of early response; A_2 denotes the second-stage treatment to which the individual is re-randomized. Y_t is the observed value of the repeated-measures outcome at time t.

For example, in the autism SMART, we have $(X_i, Y_{0,i}, A_{1,i}, Y_{12,i}, R_i, A_{2,i}, Y_{24,i}, Y_{36,i})$. A_1 denotes whether the child was randomized to BLI ($A_1 = 1$) or BLI+AAC ($A_1 = -1$) during the first 12 weeks. For slow responders to BLI ($A_1 = 1, R = 0$), A_2 denotes whether the child was re-randomized to intensified BLI ($A_2 = 1$) or BLI+AAC ($A_2 = -1$).

3.5.2 A Review of the Weighted-and-Replicated Estimator

This section is a review of a weighted-and-replicated (WR) estimator for comparing the DTRs with respect to an end-of-study outcome [47, 78, 50, 46, 43], illustrated with the autism example. In the next section, we extend this estimator to repeated-measures outcomes.

Suppose that one is interested in comparing the mean of Y_{36} among the embedded DTRs, and assume that $\mu_{36}(X, a_1, a_2; \beta)$ is a parametric model for the marginal mean of Y_{36} under embedded DTR (a_1, a_2), which takes a linear form in β and has derivative $d(X, a_1, a_2)$ with respect to β. The WR estimator for β is obtained by solving the follow-

ing estimating equation:

$$0 = \sum_{i=1}^{N} \sum_{(a_1,a_2)} I\{\text{treatment sequence of individual } i \text{ consistent with DTR } (a_1, a_2)\} \cdot d(X_i, a_1, a_2) W_i \cdot (Y_{36,i} - \mu_{36}(X_i, a_1, a_2; \beta)),$$

where $I\{\text{treatment sequence of individual } i \text{ consistent with DTR } (a_1, a_2)\}$ is a binary indicator that the individual i was assigned to treatment sequence that would be observed under the DTR (a_1, a_2); and W_i is the product of stage-specific weights, each being the inverse probability of receiving the observed treatment at that stage, conditional on the observed covariate and treatment history. In a SMART, W is known, by design. For example in the autism study, $W = 1/\left(Pr(A_1|X, Y_0) \cdot Pr(A_2|X, Y_0, A_1, Y_{12}, R)\right)$. Slow responders to BLI receive a weight equal to 4; all the other participants receive a weight equal to 2.

To appreciate why weighting is necessary, note that by design, BLI slow responders are randomized twice, whereas other participants are randomized only once; thus, slow responders to BLI would have a 1/4 chance of following the sequence of treatments they were offered, whereas other participants would have a 1/2 chance of following the treatments they were offered. Therefore, slow responders to BLI are under-represented in the data. To account for this imbalance, weights inversely proportional to the probability of being assigned to a particular treatment sequence are employed in the estimating equation.

Next, note that this estimating equation is an aggregate of estimating equations for each of the embedded DTRs. In a SMART, each individual may be consistent with one or more embedded DTRs depending on the study design. For example, in the autism SMART, responders to initial BLI are consistent with DTRs (1, 1) and (1, -1); that is, their treatment sequences are identical to the treatment sequences that would be recommended if embedded DTRs (1, 1) or (1, -1) were followed. To account for this "sharing" of observations, those observations contribute to the estimating equations for multiple DTRs.

3.5.3 An Extension for Repeated Measures

For the estimation of the repeated-measures marginal model, we use a longitudinal version of the WR estimator reviewed above. This estimator builds on works by Robins and Vansteelandt concerning the estimation of the effect of time-varying treatment on a repeated-measures outcome in observational studies [60, 13, 61, 80].

Let $Y_i = (Y_{0,i}, Y_{1,i}, ..., Y_{T,i})^T$ denote the vector of a repeated-measures outcome for individual i. Denote the vector of the model for the marginal mean as $\mu(X_i, a_1, a_2; \beta, \eta)$,

where $\mu = (\mu_0, \mu_1, ..., \mu_T)^T$. Recall $\mu_t(x, a_1, a_2; \beta, \eta)$ is a parametric model for the marginal mean of Y_t among participants that have pre-treatment baseline covariates equal to x, under the embedded DTR labeled (a_1, a_2). Denote the derivative of $\mu(X_i, a_1, a_2; \beta, \eta)$ with respect to (η, β) as $D(X_i, a_1, a_2)$. $D(X_i, a_1, a_2)$ is a $(T+1)$-by-p matrix, where p is the dimension of (η, β).

An estimator for (η, β) for a general SMART design is

$$0 = \sum_{i=1}^{N} \sum_{(a_1,a_2)} I\{\text{treatment sequence of individual } i \text{ consistent with DTR } (a_1, a_2)\} \cdot D(X_i, a_1, a_2)^T V(a_1, a_2)^{-1} W_i \cdot (Y_i - \mu(X_i, a_1, a_2; \beta, \eta)). \quad (3.3)$$

$V(a_1, a_2)$ is a working variance-covariance matrix of $(Y_0, Y_1, ..., Y_T)^T$ conditional on the baseline X, under the embedded DTR labeled (a_1, a_2). The weight W is used to account for the fact that participants received the observed treatment sequences with different probabilities. In the autism example, $W = 1/(Pr(A_1|X, Y_0) \cdot Pr(A_2|X, Y_0, A_1, Y_{12}, R))$. The choice of the working variance-covariance matrix $V(a_1, a_2)$ does not have an impact on the unbiasedness of the estimating equation above, assuming that the marginal model $\mu(X, a_1, a_2; \beta, \eta)$ and the weight W are correctly specified. In the analysis of SMART studies, the weight W can be correctly specified because the treatments are assigned according to known probabilities. Asymptotics of this estimator is provided in the appendix; this includes the formula for the asymptotic standard error of $\hat{\beta}$.

The estimator in (3.3) is an extension of the WR estimator to accommodate a repeated-measures outcome. Each patient now has a vector-valued outcome. Moreover, this estimator uses a working variance-covariance matrix for the vector of repeated measures, which is a strategy that is usually taken when performing longitudinal analysis, for the purpose of improving statistical efficiency [86]. Note that the weighting scheme here is consistent with the weighting scheme used in [59, 80], but differs from the weighting scheme adopted in [13]

Furthermore, known weights W_i in (3.3) can be estimated, for example, using covariates thought to be correlated with the repeated-measures outcome [13, 14, 5]. This approach can asymptotically improve efficiency of the estimator. We take this approach in our analyses of the data arising from three SMART studies.

3.5.4 Implementation of the Estimator for Repeated-Measures Marginal Model

To facilitate using the estimator shown in (3.3) with over-the-counter statistical software, here we conceptualize the estimating equation in (3.3) as an estimating equation based on an augmented data set, as follows:

$$0 = \sum_{j=1}^{M} D(X_j, A_{1,j}, A_{2,j})^T V(A_{1,j}, A_{2,j})^{-1} W_j \cdot (Y_j - \mu(X_j, A_{1,j}, A_{2,j}; \beta, \eta)). \quad (3.4)$$

Here, an augmented data set of size $M = N + K$ is used, with the additional K rows arising from K participants who are consistent with more than one embedded DTR. These individuals are replicated in the augmented data set. For example, in the autism study, K is the number of responders to first-stage BLI, because responders to BLI are consistent with both DTR (1, 1) and DTR (1, -1). In the augmented data set, one of the two replicated rows for a BLI responder is given the value $A_2 = 1$ and the other is given the value $A_2 = -1$; the two rows are identical in all the other components. Therefore, in this augmented data set, unlike in the original data set, each observation is associated with only one embedded DTR (i.e., the j-th observation is associated with DTR $(A_{1,j}, A_{2,j})$). In Table 3.3 we show a chunk of fake data before and after augmenting.

The estimator, written in this form, can be readily implemented on the augmented data set, in any standard statistical software that implements GEE methodology [86] in `R` [15]. For example, using the function `geeglm` in `geepack`, one can obtain both estimators and standard errors for the parameters in the repeated-measures marginal model. However, note that unlike in (3.3), the M observations in (3.4) cannot be considered as independent; this has implications on how one should use the statistical software to obtain valid standard errors. Therefore, To acquire valid standard errors that take into account replicates in the augmented data set, we need to inform `geeglm` which rows in the augmented data set are associated with the same individual (i.e., the same "cluster" in `R` terminology).

To do this, the augmented data set should include an identifier for replicates that come from the same participant; see the use of `id` in Table 3.3. For example, participant 10001 in Table 3.3 does not have replicate and thus there are four observations associated with him (because measurements are collected at four time points). Participant 10002 is replicated in the augmented data set and thus there are eight observations associated with him. This is all that is needed if an independence working correlation structure is used.

When non-independence working correlation structures are used, the preceding step also ensures that each subject's individual working correlation matrix has the appropri-

ate dimension. For example, `geeglm` expects a 4×4 correlation matrix for participant 10001 and an 8×8 correlation matrix for participant 10002. However, in the case of non-independence working correlation structures, additional work is necessary so that the software understands how to appropriately construct the user-specified working correlation matrix. To understand why this additional work is necessary, assume that we would like to analyze the repeated measures using an exchangeable working correlation structure, i.e., we want to let $V(a_1, a_2) \equiv R(\alpha)$ in (3.4), where $R(\alpha)$ is a 4×4 correlation matrix with all the off-diagonal entries equal to α. Here, in order to match the estimating equation in (3.4), participant 10001 should be assigned a correlation identical to $R(\alpha)$, whereas participant 10002 should be assigned a correlation identical to the following block matrix:

$$\begin{pmatrix} R(\alpha) & 0_{4\times4} \\ 0_{4\times4} & R(\alpha) \end{pmatrix}.$$

That is, we must inform the software that the working correlation between an observation in the original copy and an observation in the replicated copy is zero. Without additional work, the software would now know how to create an appropriate block matrix such as this; instead, the software would create an 8×8 exchangeable correlation structure with all off-diagonal entries equal to α. In `geeglm`, specialized `R` code (available from the author's webpage: `http://www-personal.umich.edu/~luxi/`) is necessary to accomplish this. This `R` code utilizes `geeglm`'s "wave" argument, which also requires additional data pre-processing step (see the `wave` column in the augmented data set in Table 3.3).

With the prepared data set and a properly specified user-defined correlation structure as illustrated above, implementing `geeglm` gives correct estimates of the parameters and their (robust) standard errors.

As discussed previously, when estimated weights rather than known weights are used, there is a potential for efficiency gains. Additional work is also necessary in order to reflect such potential efficiency gains in the estimates of the standard errors. The `R` code provided allows users to obtain more accurate standard errors that account for the estimation of weights.

3.6 Data Analysis

Here, we present the results of the data analysis of the three SMART studies. For all three SMARTs, prior to analysis, a sequential type of multiple imputation algorithm was used to

Table 3.3: Example of a chunk of data before and after augmenting. The augmented data set is ready to be analyzed by `geeglm` in `geepack`.

Before augmenting:

id	time	A_1	R	A_2	Y
10001	0	-1	0	·	23
10001	12	-1	0	·	28
10001	24	-1	0	·	32
10001	36	-1	0	·	30
10002	0	1	1	·	20
10002	12	1	1	·	25
10002	24	1	1	·	30
10002	36	1	1	·	30

After augmenting:

id	time	A_1	R	A_2	Y	waves
10001	0	-1	0	·	23	1
10001	12	-1	0	·	28	2
10001	24	-1	0	·	32	3
10001	36	-1	0	·	30	4
10002	0	1	1	1	20	1
10002	12	1	1	1	25	2
10002	24	1	1	1	30	3
10002	36	1	1	1	30	4
10002	0	1	1	-1	20	5
10002	12	1	1	-1	25	6
10002	24	1	1	-1	30	7
10002	36	1	1	-1	30	8

replace missing values in the data set [70]. This was implemented using the `mice` package in `R` [76]. All estimates and standard errors reported are calculated using standard rules for combining identical analyses performed on each of the imputed data sets [67]. Data are analyzed using the approach outlined in Section 3.5.3, with an auto-regressive working correlation structure.

3.6.1 Analysis of the Autism SMART Data

We first present the analysis of data arising from the autism SMART ($N = 61$). The weight at the first stage is estimated using age, gender, indicator of African American, indicator of Caucasian, number of socially communicative utterances at baseline; the weight (for slow responders to the first-stage BLI) at the second stage is estimated using number of socially communicative utterances at baseline and number of socially communicative utterances at week 12. Figure 3.4 displays a plot of the estimated marginal mean trajectories of the

number of socially communicative utterances for each of the three embedded DTRs. Estimates and standard errors for the parameters in the repeated-measures model and pairwise comparisons among the three embedded DTRs based on the AUCs are given in Table 3.4. To enhance interpretation we report estimates of AUC/36, which can be interpreted as the average number of socially communicative utterances over the entire course of the 36-week study.

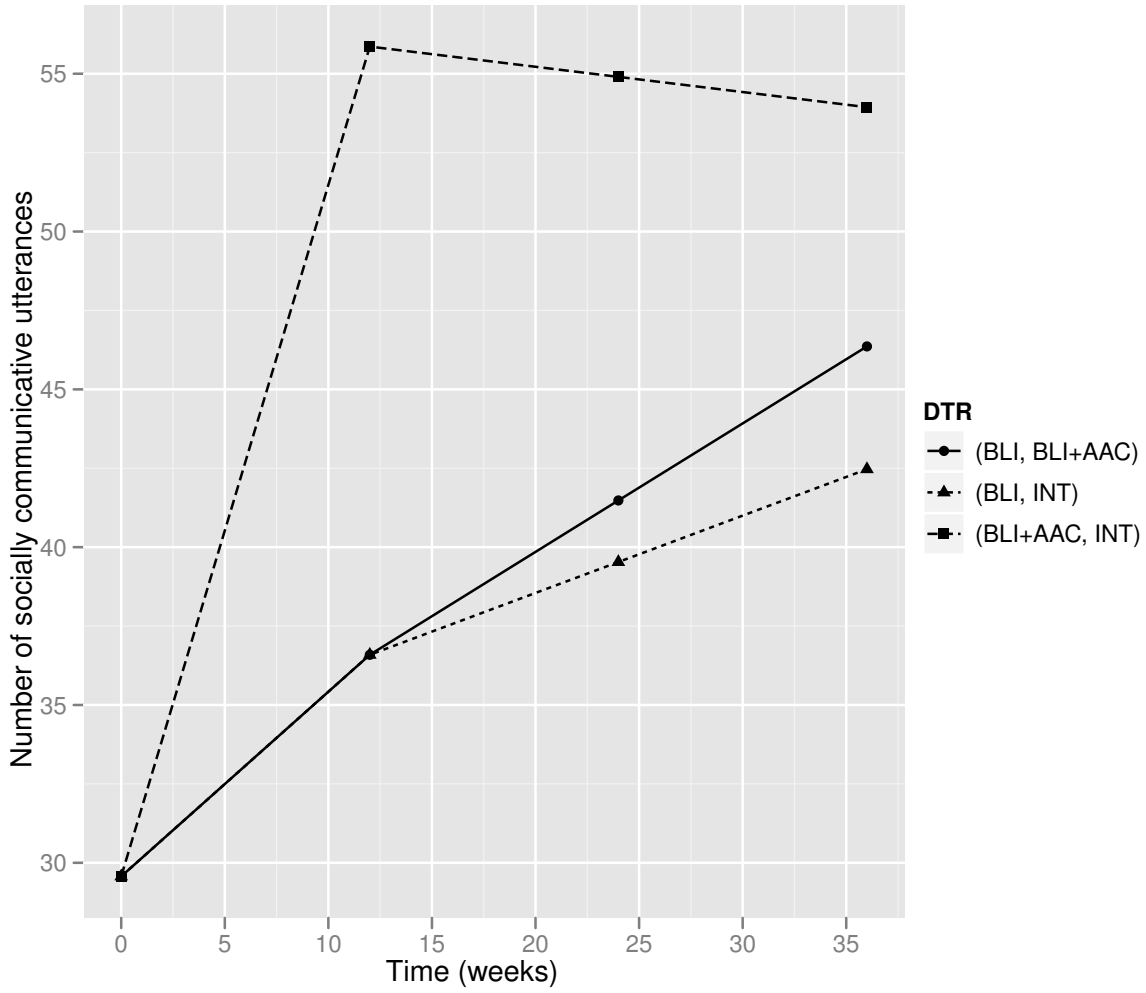

Figure 3.4: Estimated mean trajectories under the embedded DTRs of the autism SMART.

The DTR (labeled (-1, ·)) that assigns BLI+AAC at the first stage and intensifies BLI+AAC for slow responders, appears to outperform the other two embedded DTRs, in terms of the AUC (e.g., 95% CI of the contrast of (BLI+AAC, ·) versus (BLI, INT) is (2.52, 24.40)). Under this DTR, the average number of socially communicative utterances during the 36-week study is estimated to be 50.84 (95% CI (42.29, 59.39)), whereas it is smaller than 40 for the other two DTRs.

Table 3.4: An analysis of the repeated-measures outcome from the autism SMART. The reported summary of each DTR and the comparison between DTRs is regarding AUC/36.

	Estimate	SE	p-value
β_0 (Intercept)	29.57	4.53	$<$0.01
η_1 (age)	-3.63	3.59	0.32
η_2 (male)	-9.39	15.73	0.55
η_3 (AfricanAmerican)	1.04	13.92	0.94
η_4 (Caucasian)	1.09	8.18	0.89
β_1 (time; stage one)	1.39	0.40	$<$0.01
β_2 (time$\times A_1$; stage one)	-0.80	0.32	0.01
β_3 (time; stage two)	0.12	0.20	0.54
β_4 (time$\times A_1$; stage two)	0.20	0.20	0.33
β_5 (time$\times A_1A_2$; stage two)	-0.08	0.14	0.57
(BLI, INT)	37.38	4.87	$<$0.01
(BLI, BLI+AAC)	38.68	4.75	$<$0.01
(BLI+AAC, $\cdot$)	50.84	4.36	$<$0.01
(BLI, INT) vs (BLI, BLI+AAC)	-1.30	2.24	0.57
(BLI+AAC, $\cdot$) vs (BLI, BLI+AAC)	12.16	5.44	0.03
(BLI+AAC, $\cdot$) vs (BLI, INT)	13.46	5.58	0.02

Interestingly, while the DTR that begins with BLI+AAC is superior in terms of AUC, it does not maintain the positive trend from week 12 to week 36 (change score = -1.91, 95% CI (-14.21, 10.39)), while the other two DTRs seem to show marginally an average increasing trend during the same period (e.g., change score from week 12 to week 36 under (BLI, BLI+AAC) = 9.76, 95% CI (-6.89, 26.42)). These findings suggest that, in a study where the participants are followed for a longer period, the DTR that starts with BLI+AAC might be less advantageous than the other two DTRs; an additional study with a longer follow-up period would be needed to confirm this hypothesis.

3.6.2 Analysis of the ADHD SMART Data

Analysis of the ADHD SMART data ($N = 150$) is based on the repeated-measures model proposed in (3.2). The repeated-measures outcome is the classroom performance rated by teachers; higher values indicate better classroom performance. The weight at the first stage is estimated using age, indicator of being previously medicated at home, indicator of being diagnosed with oppositional defiant disorder (ODD), baseline ADHD severity of symptoms, classroom performance rating at baseline; the weight (for non-responders) at the second stage is estimated using age, stage one treatment, time to re-randomization, classroom performance rating at baseline and immediately prior to re-randomization.

Table 3.5: An analysis of the repeated-measures outcome from the ADHD SMART. The reported summary of each DTR and the comparison between DTRs is regarding AUC/7.

	Estimate	SE	p-value
β_0 (Average Intercept of BMOD and MED)	2.31	0.12	< 0.01
β_1 (A_1; baseline)	-0.25	0.11	0.03
β_2 (time; $t \leq 2$ under BMOD)	-0.01	0.16	0.93
β_3 (time; $t > 2$ under (BMOD,INT))	0.34	0.13	0.01
β_4 (time^2; $t > 2$ under (BMOD,INT))	-0.04	0.02	0.03
β_5 (time; $t > 2$ under (BMOD,BMOD+MED))	0.09	0.03	0.01
β_6 (time; $t \leq 3$ under MED)	0.11	0.07	0.13
β_7 (time; $t > 3$ under (MED,INT))	-0.03	0.04	0.41
β_8 (time; $t > 3$ under (MED,MED+BMOD))	-0.05	0.03	0.17
(BMOD, BMOD+MED)	2.29	0.14	< 0.01
(BMOD, INT)	2.48	0.13	< 0.01
(MED, MED+BMOD)	2.67	0.13	< 0.01
(MED, INT)	2.69	0.12	< 0.01
(BMOD, BMOD+MED) vs (BMOD, INT)	-0.19	0.13	0.15
(BMOD, BMOD+MED) vs (MED, MED+BMOD)	-0.38	0.19	0.05
(BMOD, BMOD+MED) vs (MED, INT)	-0.41	0.19	0.03
(BMOD, INT) vs (MED, MED+BMOD)	-0.19	0.18	0.30
(BMOD, INT) vs (MED, INT)	-0.21	0.18	0.24
(MED, MED+BMOD) vs (MED, INT)	-0.03	0.07	0.71

Table 3.5 presents the estimated AUCs for the four embedded DTRs and their comparisons. AUC/7 can now be interpreted as the average classroom performance rating from the end of month one until the end of month eight. The estimated mean trajectories of the classroom performance under the four embedded DTRs are shown in Figure 3.5.

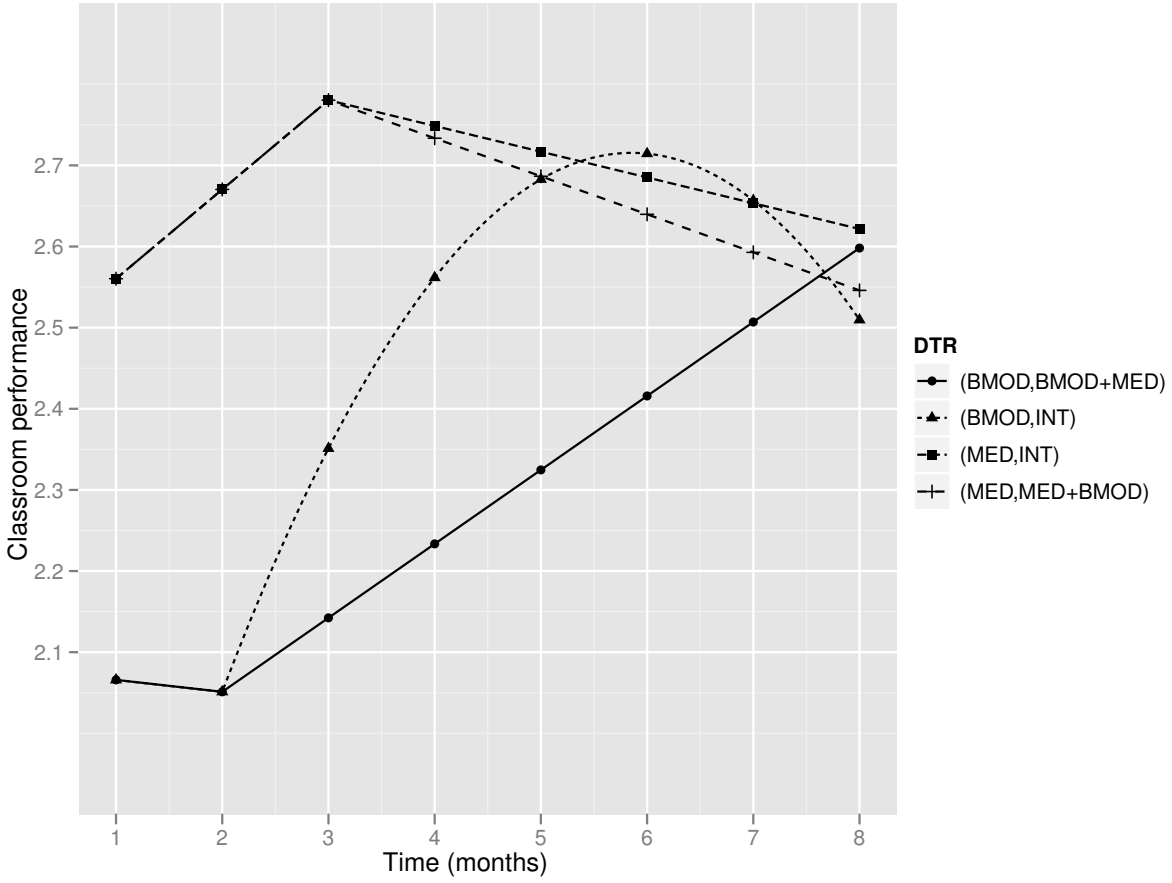

Figure 3.5: Estimated mean trajectories under the embedded DTRs of the ADHD SMART.

The DTR (BMOD, BMOD+MED) is estimated to have the smallest AUC among the four embedded DTRs, and it differs significantly from the two DTRs that start with MED. The two MED DTRs are identical in terms of AUC. However, the two DTRs starting with BMOD seem to differ. Specifically, as suggested by the estimated coefficients, the slope of DTR (BMOD, BMOD+MED) after $t = 2$ is significantly positive (0.09; 95%CI (0.03, 0.15)); on the other hand, (BMOD, INT) has a quadratic trajectory with the second-order coefficient significantly negative (-0.04; 95%CI (-0.08, 0)), and the two MED DTRs both have a slope not significantly different from zero after $t = 3$. The data suggest that (BMOD, BMOD+MED) is the only embedded DTR that maintains a trend of improvement after $t = 2$. In summary, assigning MED initially seems to yield a more positive outcome than

assigning BMOD initially in the short term, but the performance of children who initially receive BMOD improves within a wider range of time. In addition, there is no evidence that the two DTRs starting with MED differ in terms of their second-stage trajectories, but the two DTRs beginning with BMOD differ markedly in terms of their second-stage trajectories.

3.6.3 Analysis of the ExTENd SMART Data

Our analysis of the ExTENd SMART data ($N = 250$) is based on the flexible regression splines model discussed in Section 3.4.4 (details are presented in the appendix). The repeated-measures outcome is alcohol craving, assessed using the Penn Alcohol Craving Scale (PACS; [11]); values of this variable are reverse coded (ranging from 0 to 30), such that higher values indicate less alcohol craving, which is more favorable. Recall that in this study there were two definitions of non-response: stringent and lenient definitions. The weight at the first stage is estimated using age, gender, pre-study percent days heavy drinking, alcohol craving at the screening visit and the first stage one visit; the weight at the second stage is estimated using age, alcohol craving at the first and the last stage one visits, the assigned non-response definition, indicator of response to the first stage treatment, duration of stage one.

Recall that for the purpose of modeling the repeated measures, the entire 16 weeks can be conceptualized to have three periods: in the first two weeks, each group of four DTRs that are identical in the definition of non-response share one trajectory; from week two to week eight, each pair of DTRs that are identical in the non-response definition and the treatment for non-responders share one trajectory; from week eight on, each DTR has a distinct trajectory.

The estimated mean trajectories for alcohol craving under the eight embedded DTRs are shown in Figure 3.6. The estimated AUCs for the eight embedded DTRs are reported in Table 3.6. AUC/16 can be interpreted as the average alcohol craving from entry to study to the end of week 16. The estimated trajectories imply that outcomes improve over time, on average across all eight embedded DTRs. DTRs that utilize the lenient definition for non-response seem to lead to less alcohol craving than DTRs that use the stringent definition. In particular, the DTR with the highest AUC (21.19; 95%CI (20.1, 22.3)) utilizes the lenient definition and assigns UC to responders and Placebo+CBI to non-responders. However, there were no significant differences between the eight DTRs in terms of the AUCs.

The repeated-measures analysis of the ExTENd study should be considered exploratory in nature. The estimated mean trajectories are non-parametric with some constraints that

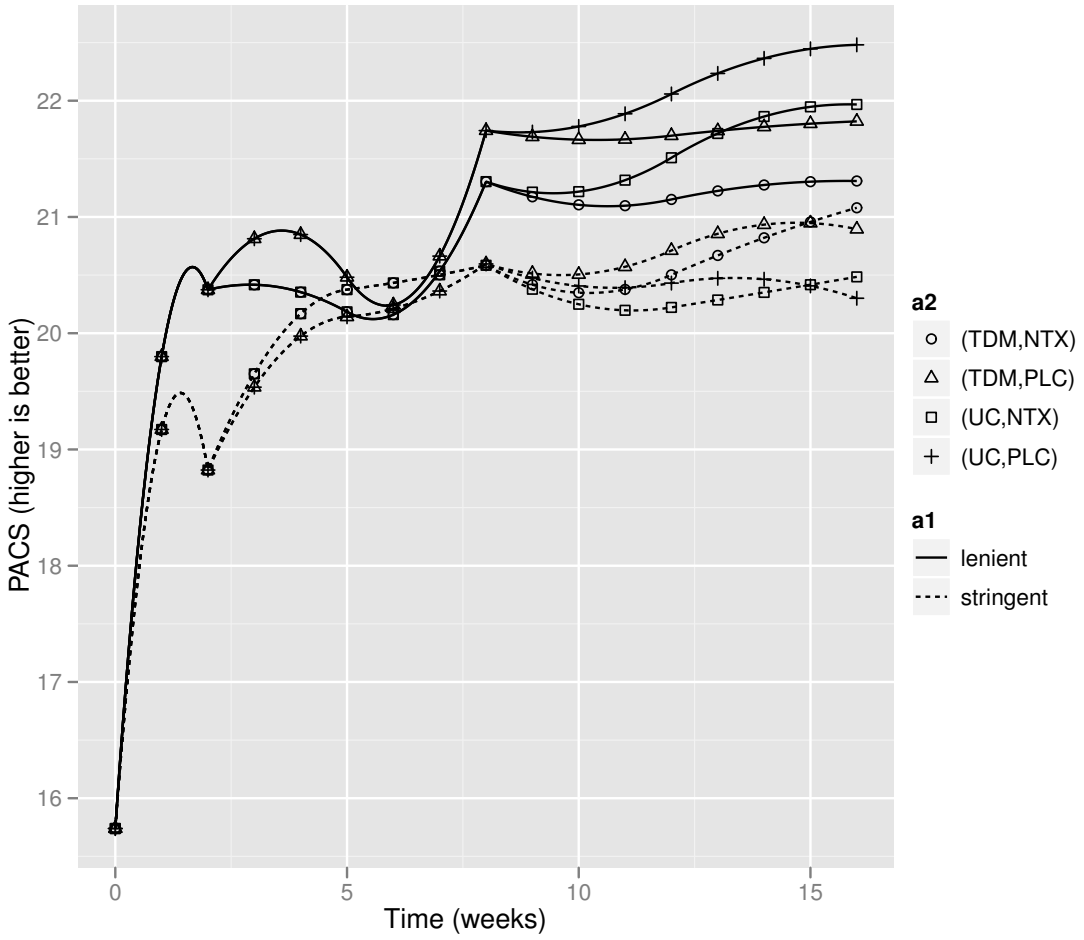

Figure 3.6: Estimated mean trajectories under the embedded DTRs of the ExTENd SMART. a1 (the definition for non-response) and a2 (stage two treatment regime) jointly specify the eight embedded DTRs.

are consistent with the design of the study, and we did not impose any smoothing constraints at $t = 2, 8$ (i.e., where the two consecutive regression splines are connected), thus the plot may present some artificial patterns that are not interpretable. Moreover, the repeated-measures outcome PACS is moderately noisy (standard deviation of the outcome at each time point ranges from 6 to 8). However, such analysis is useful for generating hypotheses regarding the developmental pattern of the repeated measures.

Table 3.6: An analysis of the repeated-measures outcome from the ExTENd SMART. LNT=lenient non-response definition. STRGT=stringent non-response definition. NTX=naltrexone+CBI. PLC=placebo+CBI.

Embedded DTR	Estimate of AUC/16	SE of AUC/16
(LNT, TDM, NTX)	20.65	0.56
(LNT, TDM, PLC)	21.02	0.55
(LNT, UC, NTX)	20.83	0.57
(LNT, UC, PLC)	21.19	0.55
(STRGT, TDM, NTX)	20.18	0.59
(STRGT, TDM, PLC)	20.18	0.56
(STRGT, UC, NTX)	20.04	0.57
(STRGT, UC, PLC)	20.03	0.58

3.7 Simulation

3.7.1 Importance of Modeling Considerations

We conduct a small set of simulation experiments to investigate the importance of incorporating the unique features of a SMART in repeated-measures models comparing embedded DTRs. In particular, we compare the bias and relative efficiency of estimators from a repeated-measures model that incorporates the features of a SMART versus traditional repeated-measures models that ignore these features. Data $(X, Y_0, A_1, Y_{12}, R, A_2, Y_{24}, Y_{36})$ were generated to mimic the autism SMART study. Notation is the same as described in Section 3.5. In particular, X is a 4-dimensional pre-treatment covariate for age, gender, indicator of African American, indicator of Caucasian.

It is well known that bias in the estimated comparison between the DTRs is expected to occur under misspecified models [46]. Here we focus on a type of model misspecification that is specific to the analysis of repeated-measures data in a SMART. We adopt a series of data-generative models under which the mean trajectory of DTR (-1, $\cdot$) is maintained to be linear, and the average of the two mean trajectories of DTRs (1, 1) and (1, -1) is maintained to be linear. Recall that DTRs (1, 1) and (1, -1) ought to share trajectories up to $t = 12$. We create a series of models by varying the extent to which the trajectories of (1, 1) and (1, -1) deviate from the average between them, thus deviating from being linear. More specifically, we let the mean trajectories of (1, 1) and (1, -1) be two piecewise linear curves that share the path from $t = 0$ to $t = 12$. To quantify the magnitude of the deviation from linear, we conceptualize an effect size in terms of the comparison of AUCs between DTRs (1, 1) and (1, -1); this is operationalized as the true difference between the two AUCs divided by the pooled standard deviation in person-specific AUCs in each DTR group. Data sets with

effect sizes equal to 0, 0.2, 0.5 and 0.8 and with sample sizes $N = 100$ and $N = 300$ are generated (details provided in the appendix). Note that a zero effect size corresponds to the case where the marginal mean trajectories of the two DTRs (1, 1) and (1, -1) do not differ over the entire course of the study; thus in this case both DTRs have a linear mean trajectory.

For each data-generative scenario, we fit three models: (a) the model shown in Equation (3.1); (b) a linear slope model, in which the mean trajectory of each embedded DTR is assumed to be linear; (c) a quadratic model, in which the mean trajectory of each embedded DTR is assumed to be quadratic. The slope and quadratic models do not impose the constraint that DTRs (1, 1) and (1, -1) share the same trajectory until the end of the first treatment stage; in other words, the treatment stage transition is not explicitly accounted for in those two models. In all cases, the estimator for the repeated-measures models utilizes an independence working correlation.

We present results for two pairwise comparisons: Δ_1^{AUC} (the difference in AUC between DTRs (1, 1) and (1, -1)) and Δ_2^{AUC} (the difference in AUC between DTRs (-1, $\cdot$) and (1, -1)). We report the bias in the estimates when using the slope model and quadratic model, and the ratio of MSE of estimators arising from the slope and quadratic models over the MSE of estimators arising from the model (a). As the slope and quadratic models are correctly specified models only in the scenario with zero effect size, we expect to see bias in all scenarios except zero effect size. On the other hand, model (a) is a correctly specified model across all simulation scenarios. However, since the slope model is more parsimonious than model (a), for small effect sizes we expect the slope model to have smaller MSE than model (a).

Table 3.7: Bias and Relative MSE (in relation to model (a) that respects the design features of a SMART) of estimates from the slope model and the quadratic model. Δ_1^{AUC} = the constrast in AUC between DTRs (1, 1) and (1, -1); Δ_2^{AUC} = the contrast in AUC between DTRs (-1, $\cdot$) and (1, -1). Bias that significantly differs from zero is in bold.

	$N = 100$							
	Δ_1^{AUC}				Δ_2^{AUC}			
	Bias x 100		RMSE		Bias x 100		RMSE	
Effect Size	Slope	Quad	Slope	Quad	Slope	Quad	Slope	Quad
0	0.1	1.7	1.67	1.87	-5.2	-1.9	0.71	1.55
0.2	**-35.4**	**13.5**	1.89	1.73	**-19.7**	-0.7	0.77	1.48
0.5	**-98.6**	**33.6**	2.7	1.6	**-50.1**	**14.9**	0.98	1.49
0.8	**-161.8**	**65.6**	2.91	1.35	**-79**	**33.9**	1.32	1.37
	$N = 300$							
	Bias x 100		RMSE		Bias x 100		RMSE	
Effect Size	Slope	Quad	Slope	Quad	Slope	Quad	Slope	Quad
0	0.1	0	1.71	1.71	3.4	5.4	0.72	1.43
0.2	**-41**	**10.7**	2.57	1.69	**-20.9**	**6.2**	0.85	1.45
0.5	**-105.5**	**25.7**	5.17	1.57	**-54.5**	**10.7**	1.53	1.44
0.8	**-189.2**	**36**	7.87	1.34	**-96**	**17**	2.97	1.38

Results are shown in Table 3.7. We notice that, as expected, the slope and quadratic models produce biased estimates when they are not correctly specified (i.e., effect size not equal to zero). However, the slope model has smaller MSE than model (a) in some non-zero effect size cases; in particular, when the sample size is small ($N = 100$), the slope model has smaller MSE than model (a) for the estimation of Δ_2^{AUC} unless the effect size is large (this is when there is severe mis-specification when assuming the slope model). This is due to the bias-variance tradeoff; the slope model is more parsimonious thus may have smaller MSE when the induced bias is larger than model (a). This tradeoff can also be appreciated by noticing that, as the sample size increases, model (a) starts to outperform the slope model under conditions with small effect sizes. Interestingly, for the estimation of Δ_1^{AUC}, model (a) is better than the slope model uniformly under all simulation scenarios. This can be intuitively explained by the fact that, the information that DTRs (1, 1) and (1, -1) share trajectory from $t = 0$ to $t = 12$ is particularly useful for the estimation of the AUC contrast between these two DTRs; this information is explicitly imposed in model (a) but not in the slope model. We also notice that the quadratic model always leads to a larger MSE than model (a), for the estimation of both contrasts and across sample sizes.

These results suggest that it is important to account for unique features of a SMART in the analysis of repeated-measures data. More traditional models such as the slope or quadratic model (these are the types of models often used in the analysis of three-arm RCTs) do not effectively utilize known information about the SMART study design and may result in bias and efficiency loss. The efficiency loss for certain estimands appears to occur even in settings where the true mean trajectories do not deviate much from the slope model or the quadratic model.

3.7.2 Efficiency Gain by Utilizing Within-person Correlation

As discussed in Section 3.5.4, the estimator for the repeated-measures model can be implemented using standard GEE software. Here we explore the extent to which use of a non-independent working correlation structure improves the statistical efficiency of the estimator. For the experiments, we generate data $(X, Y_0, A_1, Y_{12}, R, A_2, Y_{24}, Y_{36})$ to mimic the autism SMART study. For the purpose of investigating the efficiency gain due to the use of a non-independent working correlation, we adopt a series of data-generative models under which the marginal mean trajectories of the three embedded DTRs remain the same, yet the magnitude of the within-person correlation among Y_0, Y_{12}, Y_{24} and Y_{36} varies in the context of an exchangeable correlation structure. In particular, we vary the within-person correlation over 0, 0.3, 0.6, 0.9. We also vary the sample size over $N = 100, 300$.

Table 3.8: Comparison between two implementations of the proposed estimator (GEE-I uses an independent working correlation; GEE-exch uses an exchangeable working correlation), concerning the estimation of two estimands: Δ_1^{AUC} = the constrast in AUC between DTRs (1, 1) and (1, -1); Δ_2^{AUC} = the contrast in AUC between DTRs (-1, $\cdot$) and (1, -1).

	$N = 100$					
	Δ_1^{AUC}			Δ_2^{AUC}		
	CI coverage		RMSE	CI coverage		RMSE
	GEE-I	GEE-exch		GEE-I	GEE-exch	
$\rho = 0$	95.2	95.4	0.92	95.8	95.5	1.01
$\rho = 0.3$	94.1	94.7	1.01	96.3	95.8	0.89
$\rho = 0.6$	95.2	96.2	0.83	96.6	97.1	0.62
$\rho = 0.9$	94.2	95.2	0.44	95.0	96.7	0.29
	$N = 300$					
	Δ_1^{AUC}			Δ_2^{AUC}		
	CI coverage		RMSE	CI coverage		RMSE
	GEE-I	GEE-exch		GEE-I	GEE-exch	
$\rho = 0$	95.8	95.5	0.94	96.8	96.6	1.01
$\rho = 0.3$	93.7	94.5	1.03	94.0	94.4	0.88
$\rho = 0.6$	94.6	95.4	0.74	95.9	95.4	0.58
$\rho = 0.9$	96.1	96.7	0.44	94.8	97.5	0.26

Additional details concerning the data-generative models are given in the appendix.

For each data-generative scenario, we compare two estimators: they both estimate the repeated-measures model shown in (3.1); the first estimator uses an independent working correlation, and the second estimator uses an exchangeable working correlation. We present results for two pairwise comparisons: Δ_1^{AUC} (the difference in AUC between DTRs (1, 1) and (1, -1)) and Δ_2^{AUC} (the difference in AUC between DTRs (-1, $\cdot$) and (1, -1)). We report the relative mean squared error (RMSE) between the estimator with exchangeable correlation and the estimator with independent correlation in terms of Δ_1^{AUC} and Δ_2^{AUC}. We also report the coverage of the confidence interval based on the asymptotic standard error. We hypothesize that, similar to the regular GEE [30], when the true correlation level is low, using an exchangeable working correlation is almost as efficient as using an independent working correlation; however, when the true correlation among the repeated measures is at some moderate level, using an exchangeable correlation in the estimator will lead to improved efficiency.

Results are shown in Table 3.8. We observe that as expected, as the underlying within-person correlation among the repeated measures increases, the advantage of adopting an exchangeable working correlation structure in terms of the efficiency, as compared to

adopting an independent working correlation, is more remarkable. In particular, under a practically reasonable within-person correlation level $\rho = 0.6$ (e.g., this was the within-person correlation observed in the autism SMART example), for both sample sizes, using exchangeable correlation in the estimation lowers the MSE by about 40% for estimating Δ_2^{AUC}, and about 25% for estimating Δ_1^{AUC}. For correlation levels below $\rho = 0.3$, using the exchangeable correlation in the estimator does not seem to improve the efficiency.

We also observe that the confidence intervals are sometimes conservative (i.e., the coverage probability greater than the nominal level). This is because in the estimation of the standard error we use a degree-of-freedom type of correction in the sandwich estimator to account for the use of estimated coefficients as a surrogate for the unknown true values of the coefficients. Additional work is needed to better correct for small sample bias in the estimation of standard errors [34], in particular with the complication of weighting and replication.

3.8 Discussion

This chapter provides modeling guidelines for comparing DTRs based on a repeated-measures outcome arising from a SMART. Three distinct SMART study designs were used for illustration. The autism SMART has a relatively simple design, with only three embedded DTRs, and all patients transitioned to the second stage at the same time. In addition, there are only four measurement occasions during the entire study. Therefore, we suggested the piecewise linear model. In the ADHD SMART, non-responders transitioned to the second stage at different time points, and the transition times vary between two initial treatment groups on average. Thus we recommended a parametric model that accommodates these features. The ExTENd SMART differs from the other two SMARTs in that both responders and non-responders were re-randomized, but with different transition times to second stage. There are more DTRs embedded in this study (i.e., eight DTRs) and more frequent measurements of the repeated-measures outcome. Thus we modeled the trajectories of all embedded DTRs using regression splines that are properly constrained to respect the relationship among the embedded DTRs. In practice, decisions about how to appropriately model repeated measures arising from SMARTs should be based on when patients transition between treatment stages, the timing of outcome measurement occasions relative to treatment stages, and any additional area specific knowledge about the developmental pattern of the repeated-measures outcome under the assigned treatments.

In additional simulations not reported here we discovered that including the repeated measures before re-randomization in the model for estimating the true known weight seems

to play a similar role as specifying a non-independent working correlation structure in the GEE implementation, for the purpose of improving the efficiency of the estimator. However, this was in simulations mimicking the autism study with just two measurement occasions in the second stage. One advantage of the approach of using non-independent working correlation is that it allows scientists to capitalize on utilizing correlation among repeated measures that belong to the second treatment stage, which cannot be included in the model for estimating the weights.

There has been debate in the field about whether the repeated measure at baseline should be considered a covariate or a dependent variable [31]. In this chapter we chose to treat baseline as part of the repeated-measures outcome, when the measurement is available at baseline (in the ADHD SMART, it was not). We think this approach provides researchers with a more complete picture of the developmental trajectory associated with each DTR, because we are able to capture the change in the repeated measures from entry to the study.

Model selection for modeling the repeated measures under the embedded DTRs in a SMART is a challenging task and a direction for future research. In this chapter, we mainly focused on the general principles of a repeated-measures model that takes into account the specific design features of a SMART study. However, there might be multiple parametric models that are in accordance with the design features of a SMART study. Evaluation of the goodness of fit in the context of the weighted-and-replicated estimation procedure requires novel statistical methods.

The ExTENd SMART contains more subtle features that may have implications on modeling repeated measures, which are beyond the scope of this chapter. For example, the initial randomization is between two distinct criteria for non-response, instead of between two distinct treatments, as in most other trials. This implies that two DTRs that differ in the criterion for non-response can only start to differ, after the participant meets the more stringent non-response criterion. In other words, there is a chance for all the embedded DTRs to share the same mean trajectory during the first few weeks of the study. In addition, non-responders were blinded to the re-randomization, but responders were not (due to the nature of the treatments); this might have implications for modeling repeated-measures data in ExTENd. In future work we will extend the guidelines provided here to accommodate other unique features of SMART designs like ExTENd.

This work can also be extended readily in a number of directions. One natural extension is to consider different link functions in the marginal model to examine how DTRs differ based on trajectories of categorical, count or ordinal outcomes. A second extension is to the analysis of cluster- (or group-) randomized SMARTs in which clusters are randomized sequentially, yet the primary outcome is measured at the level of individuals nested within

clusters [22]; this is a setting where GEE methods are often used to account for clustering of individuals (patients) within clusters (groups).

3.9 Appendix

Asymptotics of the Weighted-and-replicated estimator for Repeated Measures

In this section we show the consistency of the estimator in Equation (3.3), and then derive the asymptotic standard error of the estimator.

The estimator given in Equation (3.3) is based on known weights and a pre-specified working variance-covariance matrix $V(a_1, a_2)$. In practice, the known weights might be estimated using covariates to improve efficiency [13, 14, 5], and $V(a_1, a_2)$ might be estimated based on a specified working correlation structure [30] (e.g., in the data analysis we use an auto-regressive correlation structure). In this case, the estimating equation is

$$\begin{aligned} 0 =& \frac{1}{N}\sum_{i=1}^{N}\sum_{(a_1,a_2)} I\{\text{treatment sequence of individual } i \text{ consistent with DTR } (a_1,a_2)\} \\ & \cdot D(X_i, a_1, a_2)^T V(a_1, a_2; \hat{\alpha})^{-1} W_i(\hat{\gamma})(Y_i - \mu(X_i, a_1, a_2; \beta, \eta)), \end{aligned} \tag{3.5}$$

where X includes a set of mean-centered baseline covariates. Consistency of the estimator arising from this estimating equation is shown in the theorem below:

Theorem 3.9.1. *Assume that the marginal model for the repeated-measures outcome is correctly specified, that is, $E_{(a_1,a_2)}[Y|X] = \mu(X, a_1, a_2; \beta_0, \eta_0)$, where (β_0, η_0) is the true value for the parameter (β, η) in the repeated-measures model. Also assume that there exist α^+, γ_0 such that $\sqrt{N}(\hat{\alpha} - \alpha^+) = O_p(1)$, and $\sqrt{N}(\hat{\gamma} - \gamma_0) = O_p(1)$, where $W(\gamma_0) \equiv W$, the true inverse-probability weight. Then the estimator $(\hat{\beta}, \hat{\eta})$ obtained by solving* (3.5) *is consistent for (β_0, η_0).*

Proof. For notational simplicity, we use $\theta = (\beta, \eta)$ to denote the parameter in the repeated-measures model and $\theta_0 = (\beta_0, \eta_0)$ to denote its true value. Denote the estimating equation by $0 = \sum_{i=1}^{N} U(Z_i; \theta, \hat{\alpha}, \hat{\gamma})/N$, where Z contains all the observed covariates for an individual. We will show that $E[U(Z; \theta_0, \alpha^+, \gamma_0)] = 0$, and then the consistency of our estimator can be established in the same way as the standard GEE estimator [30].

Note that $I\{\text{treatment sequence of the individual consistent with DTR } (a_1, a_2)\}/W$ is the Radon-Nikodym derivative between P_{obs} and $P_{(a_1,a_2)}$, where P_{obs} is the distribution

of the observed data, $P_{(a_1,a_2)}$ is the distribution of data in the population where all the individuals follow the embedded DTR (a_1, a_2). Hence,

$$\begin{aligned} E[U(Z;\theta_0,\alpha^+,\gamma_0)] &= \sum_{(a_1,a_2)} E_{(a_1,a_2)} D(X,a_1,a_2) V(a_1,a_2;\alpha^+)^{-1} \left(Y - \mu(X,a_1,a_2;\theta_0)\right) \\ &= \sum_{(a_1,a_2)} E_X D(X,a_1,a_2) V(a_1,a_2;\alpha^+)^{-1} E_{(a_1,a_2)}[Y - \mu(X,a_1,a_2;\theta_0)|X] \\ &= 0. \end{aligned}$$

The last equality follows because the repeated-measures model is correctly specified. □

The theorem below concerns the asymptotic distribution of the estimator obtained from (3.5).

Theorem 3.9.2. *Assume mild regularity conditions and the same assumptions as in Theorem 3.9.1, and assume that the parameter γ in the weight is obtained from a maximum likelihood estimator for the treatment assignment probabilities, with a score function S_γ. Then $\sqrt{N}\left((\hat{\beta},\hat{\eta}) - (\beta_0,\eta_0)\right)$ is asymptotically multivariate normal with zero mean and covariance matrix $\Sigma = J^{-1}IJ^{-1}$, where I, J are given by*

$$I = E[UU^T] - E[US_\gamma^T]E[S_\gamma S_\gamma^T]^{-1}E[S_\gamma U^T],$$

and

$$\begin{aligned} J =& E \sum_{(a_1,a_2)} I\{\textit{treatment sequence of the individual consistent with DTR } (a_1,a_2)\} \\ & W D(X,a_1,a_2)^T V(a_1,a_2;\alpha^+)^{-1} D(X,a_1,a_2) \end{aligned}$$

where

$$\begin{aligned} U :=& \sum_{(a_1,a_2)} I\{\textit{treatment sequence of individual consistent with DTR } (a_1,a_2)\} \\ & \cdot D(X,a_1,a_2)^T V(a_1,a_2;\alpha^+)^{-1} W(Y - \mu(X,a_1,a_2;\beta_0,\eta_0)). \end{aligned}$$

Proof. Again use θ to denote the unknown parameter in the repeated-measures model. Denote the proposed estimating equation by $0 = \sum_{i=1}^N U(Z_i;\theta,\hat{\alpha},\hat{\gamma})/N$. Using the same argument as for standard GEE estimator, we can derive that

$$\sqrt{N}(\hat{\theta}-\theta_0) = -E\left[\frac{\partial U(Z;\theta_0,\alpha^+,\gamma_0)}{\partial\theta}\right]^{-1}\Big\{\frac{1}{\sqrt{N}}\sum_{i=1}^{N}U(Z_i;\theta_0,\alpha^+,\gamma_0) + E\left[\frac{\partial U(Z;\theta_0,\alpha^+,\gamma_0)}{\partial\gamma}\right]\sqrt{N}(\hat{\gamma}-\gamma_0)\Big\} + o_p(1)$$

Let $U := U(Z;\theta_0,\alpha^+,\gamma_0)$ and use the fact that S_γ is the score function for $\hat{\gamma}$, then we could further write:

$$\sqrt{N}(\hat{\theta}-\theta_0) = -E\left[\frac{\partial U(Z;\theta_0,\alpha^+,\gamma_0)}{\partial\theta}\right]^{-1} \cdot\Big\{\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left[U_i - E[US_\gamma^T]E[S_\gamma S_\gamma^T]^{-1}S_{\gamma,i}\right]\Big\} + o_p(1).$$

Thus the asymptotic variance of $\sqrt{N}(\hat{\theta}-\theta_0)$ is equal to

$$J^{-1}\left(E[UU^T] - E[US_\gamma^T]E[S_\gamma S_\gamma^T]^{-1}E[S_\gamma U^T]\right)J^{-1}.$$

□

Remark: In particular, the asymptotic variance of $\sqrt{N}((\hat{\beta},\hat{\eta})-(\beta_0,\eta_0))$ does not depend on the choice of the estimator for α, i.e., the parameter in the working covariance matrix, among those that have a $\sqrt{N}$-rate for a same limit α^+. This is because $E\left[\frac{\partial U(Z;\theta_0,\alpha^+,\gamma_0)}{\partial\alpha}\right]=0$ holds when the repeated-measures model and the model for weights are correctly specified. This property is same as the standard GEE estimator. On the other hand, estimating the known weights should provide a reduction in the asymptotic variance of $\hat{\beta}$, particularly when the projection of U on the vector space of S_γ is not close to zero. Intuitively, this might be the case if the repeated-measures outcome Y is correlated with the covariates used in the model for estimating weights.

To obtain an estimate of the standard error of $(\hat{\beta},\hat{\eta})$, we use plug-in estimators for J and I. Namely, we set

$$\hat{J} = 1/N\sum_{i=1}^{N}\sum_{(a_1,a_2)} I\{\text{treatment sequence of the individual } i \text{ consistent with DTR } (a_1,a_2)\} W_i D(X_i,a_1,a_2)^T V(a_1,a_2;\hat{\alpha})^{-1}D(X_i,a_1,a_2);$$

and we set

$$\hat{I} = 1/N \sum_{i=1}^{N} \hat{U}_i \hat{U}_i^T - \left(1/N \sum_{i=1}^{N} \hat{U}_i \hat{S}_{\gamma,i}^T \right) \left(1/N \sum_{i=1}^{N} \hat{S}_{\gamma,i} \hat{S}_{\gamma,i}^T \right)^{-1} \left(1/N \sum_{i=1}^{N} \hat{S}_{\gamma,i} \hat{U}_i^T \right),$$

where

$$\hat{U}_i := \sum_{(a_1,a_2)} I\{\text{treatment sequence of individual } i \text{ consistent with DTR } (a_1, a_2)\}$$
$$\cdot D(X_i, a_1, a_2)^T V(a_1, a_2; \hat{\alpha})^{-1} W_i (Y_i - \mu(X_i, a_1, a_2; \hat{\beta}, \hat{\eta}))$$

and $\hat{S}_{\gamma,i} := S_{\hat{\gamma},i}$. The plug-in estimator for Σ is $\hat{\Sigma} = \hat{J}^{-1} \hat{I} \hat{J}^{-1}$.

Exploratory Plot of Repeated Measures in ADHD SMART

To guide the repeated-measures modeling of the ADHD SMART, we made an exploratory plot of the repeated measures. More specifically, we plot the empirical mean of the classroom performance rating under each of the four embedded DTRs separately at each time point. The exploratory plot is shown in Figure 3.7.

We notice that the two MED regimes do not seem to differ much before $t = 3$, whereas the two BMOD regimes start to differ notably after $t = 2$. There is no evident trend in the two MED regimes after $t = 3$. Repeated measures under (BMOD, BMOD+MED) seem to improve at a constant rate after $t = 2$, whereas repeated measures under (BMOD, INT) seem to improve rapidly immediately after $t = 2$, but the rate of improvement is not maintained through the end of the study. Guided by these observations, we propose the repeated-measures model that is shown in Equation (3.2).

Details Concerning the Data-generative Models for Simulation 1

In the simulation in Section 3.7.1, we illustrate the importance of considering the special features of SMART designs in the modeling of repeated measures from SMART trials, by comparing a repeated-measures model that incorporates SMART features with more traditional longitudinal models such as slopes and quadratic models. Here we provide additional details about the data-generative models used in this simulation. We adopted a series of data-generative models under which the mean trajectory of DTR (-1, $\cdot$) is maintained to be linear, and the average of the two mean trajectories of DTRs (1, 1) and (1, -1) is maintained to be linear; each data-generative model in the series is indexed by a parameter $\theta > 0$, which quantifies the extent to which the trajectories of DTRs (1, 1) and (1, -1)

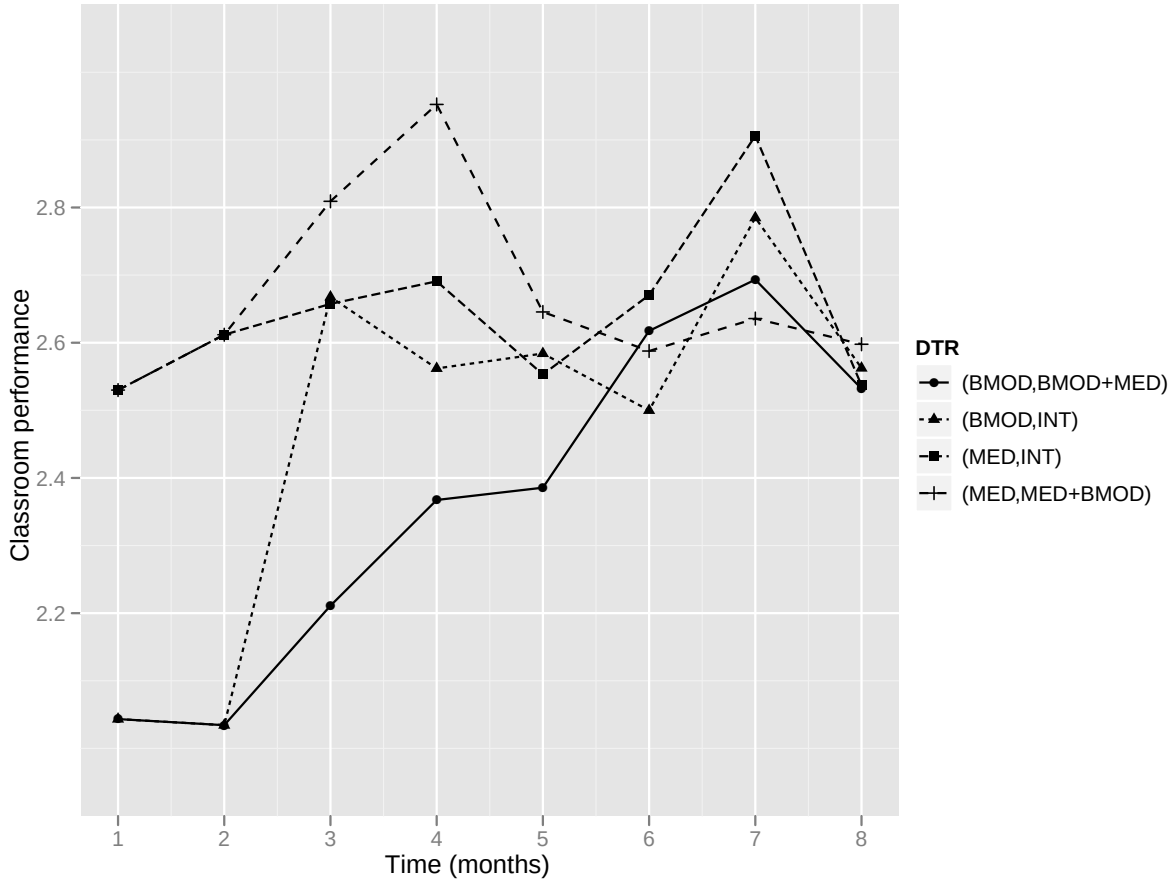

Figure 3.7: Exploratory plot of ADHD SMART: empirical mean of the repeated-measures outcome under each embedded DTR, at each time point.

deflect at $t = 12$.

We generate data $(X, Y_0, A_1, Y_{12}, R, A_2, Y_{24}, Y_{36})$ for each individual in a sample of size N.

- X includes six mean-centered baseline covariates: age, gender, indicator of African American, indicator of Caucasian, indicator of Hispanic, indicator of Asian. (Note that in the simulation experiments, the marginal model we fit for repeated measures only includes the first four covariates in X to avoid rank deficient problem in the estimation) X is sampled (with replacement) from the real autism SMART data. For notational simplicity, we let X below always contain intercept as the first coordinate.
- Generate $Y_0 = \eta_0^T X + \epsilon_0$, where $\epsilon_0 \sim N(0, \sigma^2)$.
- Generate A_1 to be -1 or 1 with equal probability.
- Generate $Y_{12} = \eta_{11}^T X + \eta_{12} Y_0 + \beta_{11} A_1 + \epsilon_1$, where $\epsilon_1 \sim N(0, \sigma^2)$.
- Generate A_2 to be -1 or 1 with equal probability, among individuals with $A_1 = 1$ and $R = 0$; otherwise set $A_2 = 0$.
- Generate $Y_{24} = \eta_{21}^T X + \eta_{22} Y_0 + \eta_{23}^T A_1 + \eta_{24} Y_{12} + \beta_{21}(1 - R)(A_1 + 1)A_2 + \epsilon_2$, where $\beta_{21} = -\theta$ and $\epsilon_2 \sim N(0, \sigma^2)$.
- Generate $Y_{36} = \eta_{31}^T X + \eta_{32} Y_0 + \eta_{33}^T A_1 + \eta_{34} Y_{12} + \beta_{31}(1 - R)(A_1 + 1)A_2 + \epsilon_3$, where $\eta_{31} = 2\eta_{21}$, $\eta_{32} = 2\eta_{22}$, $\eta_{33} = 2\eta_{23}$, $\eta_{34} = 2\eta_{24} - 1$, $\beta_{31} = 2\beta_{21} = -2\theta$ and $\epsilon_3 \sim N(0, \sigma^2)$.
- The values of the coefficients mentioned above: $\eta_0 = (29.5, -5.1, -16.3, 0, 14.3, -11.8, 0.5)$, $\sigma = 10$, $\eta_{11} = (23.46, 1.4, -3.0, 16.6, 11.1, 6.5, 22.5)$, $\eta_{12} = 0.3$, $\beta_{11} = -1$, $\eta_{21} = (22.758, 1.20, 4.33, 12.33, 4.00, 7.53, 7.47)$, $\eta_{22} = 0.2$, $\eta_{23} = -1.8$, $\eta_{24} = 0.2$.

In order to have data-generative models that are reasonable, we conceptualize an effect size in terms of the contrast in AUC between two embedded DTRs. We define the effect size of the comparison between DTR (1, 1) and DTR (1, -1) as the ratio of the difference in their AUCs over the pooled standard deviation of "a person-specific AUC" between the two DTR groups. More specifically, we operationalize the person-specific AUC as $12(Y_0/2 + Y_{12} + Y_{24} + Y_{36}/2)$ for each individual. Let $\sigma_{(1,1)}$ denote the standard deviation of this person-specific AUC under DTR (1, 1) and $\sigma_{(1,-1)}$ denote the standard deviation of this person-specific AUC under DTR (1, -1). Then the effect size mentioned above can be written as $(AUC_{(1,1)} - AUC_{(1,-1)})/\sqrt{(\sigma^2_{(1,1)} + \sigma^2_{(1,-1)})/2}$. This measure quantifies the extent to which DTRs (1, 1) and (1, -1) differ throughout the entire study period.

Figure 3.8 shows the true mean trajectories of the repeated measures under the three embedded DTRs, in each of the four data-generative models (with the effect size defined earlier equal to 0, 0.2, 0.5, 0.8) that we use in our simulation experiments. Across all of the four data-generative models, the effect size in terms of the comparison between DTR (-1, ·) and the average of DTRs (1, 1) and (1, -1) is kept at around 0.4.

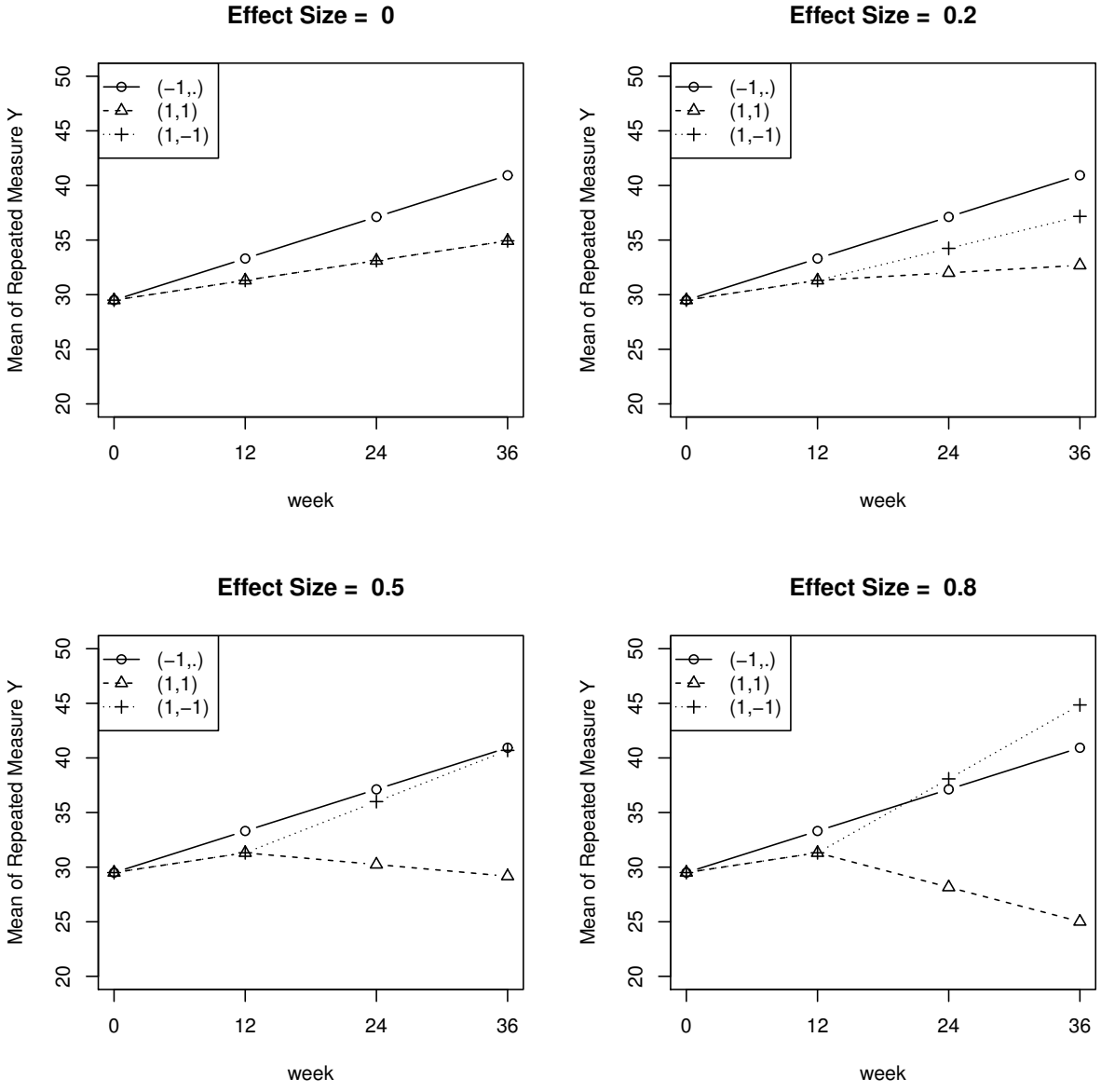

Figure 3.8: True mean trajectories of the repeated measures under the embedded DTRs, under four data-generative models corresponding to effect size (of the contrast in AUC between DTR (1, 1) and (1, -1)) = 0, 0.2, 0.5, 0.8.

Details concerning the Data-generative Models for Simulation 2

Here we provide additional details concerning the data-generative models adopted for the simulation experiments that investigate the efficiency gain due to a working correlation structure. The data-generative models here differ from those in the previous section; the goal here is to provide a series of data-generative models that all imply the same marginal mean model for the repeated-measures outcome Y_t, yet the within-person correlation among Y_t's varies across different data-generative models.

We generate data $(X, Y_0, A_1, Y_{12}, R, A_2, Y_{24}, Y_{36})$ for each individual in a sample of size N. The data-generating process below is indexed by $(\alpha_1, \alpha_2, \alpha_3, k_1, k_2, k_3)$; tuning these parameters would not change the marginal mean model for Y_t (conditional on baseline X), but would change the within-person correlation and variances of Y_t.

- X includes six mean-centered baseline covariates: age, gender, indicator of African American, indicator of Caucasian, indicator of Hispanic, indicator of Asian. X is generated in a way identical to the previous section.

- Generate $Y_0 = \eta_0^T X + \epsilon_0$, where $\epsilon_0 \sim N(0, \sigma^2)$.

- Generate A_1 to be -1 or 1 with equal probability.

- Generate $Y_{12} = \eta_{11}^T X + \eta_{12} Y_0 + \beta_{11} A_1 + \epsilon_1$, where $\epsilon_1 \sim N(0, \sigma_1^2)$. $\eta_{12} = \eta_{12}^* \alpha_1$; α_1 controls the correlation between Y_0 and Y_{12}. $\eta_{11} = \eta_{11}^* + (1 - \alpha_1)\eta_{12}^* \eta_0$; η_{11} is chosen to maintain the same marginal mean $E[Y_{12}(a_1)|X]$ across various data-generative models. $\sigma_1^2 = k_1 \sigma^2$; k_1 controls the variance of Y_{12}.

- Generate A_2 to be -1 or 1 with equal probability, among individuals with $A_1 = 1$ and $R = 0$; otherwise $A_2 = 0$.

- Generate (e_2, e_3) jointly from a bivariate mean zero normal distribution with variance 1 and correlation 0.25. They will be used in the generation of error terms of (Y_{24}, Y_{36}).

- Generate $Y_{24} = \eta_{21}^T X + \eta_{22} Y_0 + \eta_{23} A_1 + \eta_{24} Y_1 + \eta_{25} R + \eta_{26} A_1 R + \beta_{21}(1 - R)(A_1 + 1)A_2 + \epsilon_2$. $\eta_{22} = \eta_{22}^* \alpha_2$; α_2 controls the correlation between Y_0 and Y_{24} (and Y_{36}). $\eta_{24} = \eta_{24}^* \alpha_3$; α_3 controls the correlation between Y_{12} and Y_{24} (and Y_{36}). η_{21} and η_{23} are accordingly adjusted to maintain the same marginal mean $E[Y_{24}(a_1, a_2)|X]$; in particular, $\eta_{21} = \eta_{21}^*$ and $\eta_{23} = \eta_{23}^*$ when $\alpha_2 = \alpha_3 = 1$. $\epsilon_2 = \sigma_2 e_2$; $\sigma_2^2 = k_2 \sigma^2$; k_2 controls the variance of Y_{24}.

- Generate $Y_{36} = \eta_{31}^T X + \eta_{32} Y_0 + \eta_{33}^T A_1 + \eta_{34} Y_1 + \eta_{35} R + \eta_{36} A_1 R + \beta_{31}(1-R)(A_1 + 1)A_2 + \epsilon_3$, where $\eta_{31} = 2\eta_{21}$, $\eta_{32} = 2\eta_{22}$, $\eta_{33} = 2\eta_{23}$, $\eta_{34} = 2\eta_{24} - 1$, $\eta_{35} = 2\eta_{25}$, $\eta_{36} = 2\eta_{26}$, $\beta_{31} = 2\beta_{21}$. $\epsilon_3 = \sigma_3 e_3$; $\sigma_3^2 = k_3\sigma^2$; k_3 controls the variance of Y_{36}.

- The values of the coefficients mentioned above: $\eta_0 = (29.5, -5.1, -16.3, 0, 14.3, -11.8, 0.5)$. $\eta_{11}^* = (23.46, 3.5, 1.4, -3.0, 16.6, 11.1, 6.5, 22.5)$, $\eta_{12}^* = 0.7$, $\beta_{11} = -2.8$. $\eta_{21}^* = (18.4, 1.20, 4.33, 12.33, 4.00, 7.53, 7.47)$, $\eta_{22}^* = 0.33$, $\eta_{23}^* = -7.2$, $\eta_{24}^* = 0.5$, $\eta_{25} = -10.1$, $\eta_{26} = 5.1$, $\beta_{21} = -3$.

As can be seen above, each data-generating model is indexed by a set of parameters $(\alpha_1, \alpha_2, \alpha_3, k_1, k_2, k_3)$. In the simulation, we focus on four scenarios with the following choices of those parameters (we were only able to control the within-person correlation among Y_0, Y_{12} and Y_{24}):

- $(\alpha_1, \alpha_2, \alpha_3, k_1, k_2, k_3) = (0, 0, 0, 1, 1, 1.5)$. As a result, the correlation among Y_0, Y_{12}, Y_{24} is around 0.

- $(\alpha_1, \alpha_2, \alpha_3, k_1, k_2, k_3) = (0.43, 0.72, 0.48, 0.91, 0.85, 1.19)$. As a result, the correlation among Y_0, Y_{12}, Y_{24} is around 0.3.

- $(\alpha_1, \alpha_2, \alpha_3, k_1, k_2, k_3) = (0.86, 1.17, 0.77, 0.64, 0.52, 0.72)$. As a result, the correlation among Y_0, Y_{12}, Y_{24} is around 0.6.

- $(\alpha_1, \alpha_2, \alpha_3, k_1, k_2, k_3) = (1.28, 1.48, 0.97, 0.19, 0.10, 0.42)$. As a result, the correlation among Y_0, Y_{12}, Y_{24} is around 0.9.

Details Concerning the Analysis of Repeated-Measures Data in the ADHD Study

For individual i, we observe $X_i, A_{1,i}, R_i, M_i, A_{2,i}$ and repeated measures $Y_{1,i}, ..., Y_{8,i}$. $A_1 = 1$ denotes that the individual received low-intensity BMOD and $A_1 = -1$ denotes that the individual received low-dose MED. R indicates whether the individual continued to respond until the end of the study. When $R = 0$ (i.e., the individual became a non-responder during the study), M denotes the time (in months) of non-response and A_2 denotes whether $(A_2 = 1)$ the initial treatment was intensified or $(A_2 = -1)$ the initial treatment was augmented with the alternative treatment.

The repeated-measures model proposed in (3.2) was estimated using the estimator presented in (3.3). In particular, the treatment sequence of each individual is consistent with either one or two of the embedded DTRs. An individual's treatment sequence is consistent

with only one embedded DTR if this individual was a non-responder (e.g., a non-responder to BMOD who was later re-randomized to INT is only consistent with DTR (BMOD, INT)). An individual's treatment sequence is consistent with two embedded DTRs if this individual was a responder (e.g., a responder to BMOD is consistent with both (BMOD, INT) and (BMOD, BMOD+MED)). The weight W in (3.3) is the inverse probability of an individual receiving the treatment sequence that was assigned to him/her. Therefore, responders receive a weight equal to 2 (they were randomized only once, to two options) and non-responders receive a weight equal to 4 (they were randomized twice, each time to two options). In our data analysis, however, we estimate these known weights using covariates specified in Section 3.6.2 to improve the statistical efficiency.

Details Concerning the Analysis of Repeated-Measures Data in the ExTENd Study

Given the many measurement occasions of the repeated-measures outcome, we adopted a piecewise splines model. Here we describe the details. From $t = 0$ to $t = 2$, we let the mean trajectory under DTR (a_1, a_{2R}, a_{2NR}) be a regression spline that can only vary with a_1 and has the identical intercept regardless of a_1. From $t = 2$ to $t = 8$, we let the mean trajectory under the DTR (a_1, a_{2R}, a_{2NR}) be a regression spline that continously connects to the trajectory between $t = 0$ and $t = 2$, and can vary with different values of (a_1, a_{2NR}). From $t = 8$ to $t = 16$, we let the mean trajectory under the DTR (a_1, a_{2R}, a_{2NR}) be a regression spline that continuously connects to the trajectory up to $t = 8$, and the trajectory can vary with different values of (a_1, a_{2R}, a_{2NR}). For model simplicity, all the b-spline bases are of degree 2. We apply internal knots at $t = 5$ (midway from $t = 2$ to $t = 8$) and $t = 12$ (midway from $t = 8$ to $t = 16$).

A regression splines model can be viewed as a linear model, with properly chosen functions of b-spline bases as predictors. Therefore, the estimator presented in (3.3) can be readily applied. More specifically, in the ExTENd study, each individual's treatment sequence is consistent with two embedded DTRs. For example, a patient who was assigned the lenient early non-response definition and later transitioned to stage two as a responder and received TDM was consistent with the following two DTRs: (a_1, a_{2R}, a_{2NR})=(lenient, TDM, NTX+CBI) and (a_1, a_{2R}, a_{2NR})=(lenient, TDM, Placebo+CBI). The weight in (3.3) is equal to 4 for every individual, because in the ExTENd study each individual was randomized twice, each time to one of two options. In our data analysis, we estimate these known weights using the covariates specified in Section 3.6.3 to improve the statistical efficiency.

CHAPTER 4

Small-Sample Considerations in the Comparison of Dynamic Treatment Regimes Using SMART Data

4.1 Introduction

The Sequential Multiple Assignment Randomized Trial (SMART), was developed for the purpose of building high-quality dynamic treatment regimes (DTRs). A common aim in a SMART is to compare the mean of an end-of-study outcome between two or more of the DTRs embedded within it (see Section 3.3 for introduction about the embedded DTRs in a SMART). A regression-based inverse probability-of-treatment weighting (IPTW) method has been proposed for comparing the DTRs with respect to an end-of-study outcome [47, 78, 50, 46, 43], based on marginal mean models for the DTRs; in a SMART, these weights are known, by design. We call it a weighted-and-replicated (WR) estimator due to a simple way to implement it using existing software [43]. Briefly, weighting adjusts for the fact that, by design, participants in a SMART may differ in their probability of being offered their sequence of treatments; whereas replication is used to take advantage of the fact that some SMART participants are consistent with more than one of the embedded DTRs being compared.

This chapter focuses on small sample considerations in the use of WR estimator with data arising from a SMART. This work is motivated by a SMART in autism (shown in Figure 3.1 and described in more detail in Section 3.3) which has three DTRs embedded within in and a sample size of $n = 61$ (considered small). Little is known concerning the small sample properties of the WR estimator, particularly the performance of its corresponding asymptotic variance estimator. Not all SMARTs are expected to have sample sizes that are "sufficiently large". This may be particularly true in settings in which (i) the SMART may not have been powered for the mean comparisons of the embedded DTRs because this is a

secondary aim of the SMART, or (ii) recruitment difficulties prohibited investigators from achieving their planned sample size goals.

In this chapter, we develop a small-sample variance estimator of the WR estimator by extending the work of [34]. Moreover, it is well-known that the asymptotic statistical efficiency of IPTW estimators is improved when estimated, rather than known, weights are used [62, 14, 5, 84]. Hence, we also develop a small-sample variance estimator for the case when estimated, as opposed to known, weights are used in the WR estimator. We investigate via simulation studies the performance of the proposed small-sample variance estimator. The ongoing work of Almirall et al. provides more complete discussions about the contents of this chapter. In particular, there the authors also investigate the efficiency gains that can be achieved when varying the level of correlation among covariates used in the weight model and the outcome, via simulation studies in a small-sample setting; also presented there are the analyses of the autism SMART data that motivates this project. However, this chapter will mainly focus on presenting the small-sample variance estimator of the WR estimator.

In Section 4.2, we briefly review the marginal mean model and the form of WR estimator. In Section 4.3, we propose a small-sample bias-corrected estimator for the variance of WR estimator. The empirical performance of the proposed variance estimator is investigated in a simulation study in Section 4.4. Concluding remarks and discussions are in Section 4.5.

4.2 Model and Estimator

The marginal mean model and the WR estimator for the model were introduced in Section 3.5. For completeness, here we briefly review them in the context of the autism SMART.

For each SMART study participant and each one of the dynamic treatment regimes (a_1, a_2) embedded in the SMART, we envision a primary end-of-study outcome $Y(a_1, a_2)$. a_1 denotes the first-stage treatment, and a_2 denotes the second-stage treatment. We use contrast coding (i.e., (-1, +1) coding) to facilitate the interpretation of the parameters in the marginal mean model. The three DTRs embedded in the autism SMART can be denoted in the identical way to that in Section 3.4. Specifically, the DTR that starts with BLI and augments BLI with AAC for slow responders is labeled (1, -1); the DTR that starts with BLI and intensifies BLI for slow responders is labeled (1, 1); the DTR that starts with BLI+AAC and intensifies BLI+AAC for slow responders is labeled (-1, $\cdot$). Here in the analysis of data arising from the autism study, we focus on a primary outcome Y that is

the total number of socially communicative utterances at the end of treatment (week 24); higher values are more desirable.

Our overarching goal is to compare values of $E[Y(a_1, a_2)]$, which is the marginal mean of the primary outcome Y for each of the SMART-embedded DTRs. As with models used in the analysis of standard randomized clinical trials, the marginal modeling approach also includes models for quantities such as $E[Y(a_1, a_2)|X]$, in which X includes pre-specified pre-treatment covariates.

The comparison of mean outcomes between the DTRs embedded in a SMART is facilitated by parametric marginal mean models $\mu(a_1, a_2, X; \beta, \eta)$ for $E[Y(a_1, a_2)|X]$ of the form

$$\mu(a_1, a_2, X; \beta, \eta) = \beta^T f(a_1, a_2) + \eta^T X,$$

where $\beta = (\beta_1, ..., \beta_p)^T$ and $\eta = (\eta_1, ..., \eta_q)$ are, respectively, p- and q-dimensional column vectors of unknown parameters; and, for simplicity, we assume X is mean centered. The form of $f(a_1, a_2)$ and the covariates in X are pre-specified in advance of collecting the SMART data. η quantifies the association between the baseline covariates X and the outcome Y, but is not necessarily of scientific interest because it does not carry information about the comparison of the embedded DTRs. The form of $f(a_1, a_2)$ will depend on the design of the SMART. In the autism study, an example model is $f(a_1, a_2) = (1, a_1, I(a_1 = 1)a_1a_2)$ so that

$$\mu(a_1, a_2, X; \beta, \eta) = \beta_1 + \beta_2 a_1 + \beta_3 I(a_1 = 1)a_1a_2 + \eta^T X. \tag{4.1}$$

We next review the WR estimator, using the autism SMART as an example. Let $\theta = (\beta, \eta)$ denote the complete set of $p+q$ unknown parameters. Let $O = (X, L_1, A_1, R, L_2, A_2, Y)$ denote the observed SMART data used to estimate the unknown parameters θ. L_t denotes auxiliary data collected prior to first-stage treatment assignment (L_1), as well as data collected after first-stage treatment assignment but prior to second-stage treatment assignment (L_2). R is a binary variable denoting responder $(R = 1)$ versus non-responder/slow responder $(R = 0)$ to first-stage treatment. A_t is the randomly assigned treatment at each stage t.

In the autism example, WR estimator $\hat{\theta}$ is the solution for θ to the following weighted

estimating equations:

$$\begin{aligned}0 =&\mathbb{P}_n 2RI(A_1=1)(Y-\beta_1-\beta_2-\beta_3-\eta^T X)D_{(1,1)}^T\\&+\mathbb{P}_n 2RI(A_1=1)(Y-\beta_1-\beta_2+\beta_3-\eta^T X)D_{(1,-1)}^T\\&+\mathbb{P}_n 2I(A_1=-1)(Y-\beta_1+\beta_2-\eta^T X)D_{(-1,\cdot)}^T\\&+\mathbb{P}_n 4(1-R)I(A_1=1)(Y-\beta_1-\beta_2-\beta_3 A_2-\eta^T X)D_{(1,A_2)}^T,\end{aligned}$$

where $D_{(1,1)} = (f^T(1,1),X)$, $D_{(1,-1)} = (f^T(1,-1),X)$ and $D_{(-1,\cdot)} = (f^T(-1,\cdot),X)$ are the $(p+q)$-dimensional model design row-vectors associated with each of the three embedded DTRs.

Next we present the WR estimator in a different form, for general SMART designs. Mathematically, the estimators presented in the previous display is equivalent to this alternative form; however, this alternative form will later facilitate the derivation of the small-sample variance of the WR estimator more easily. For a general SMART design with K embedded two-stage DTRs $\{(a_1^k,a_2^k)\}_{k=1}^K$, the marginal mean model estimator can be formalized as the solution to the estimating equations

$$0=\sum_{i=1}^{n} w(X_i,\bar{L}_{2,i},R_i,\bar{A}_{2,i})D_i^T\epsilon_i(\theta), \tag{4.2}$$

where

$$D=\begin{pmatrix}D_1\\ \vdots\\ D_k\\ \vdots\\ D_K\end{pmatrix} \quad\text{and}\quad \epsilon(\theta)=\begin{pmatrix}I_1\cdot(Y-\mu(a_1^1,a_2^1,X;\theta))\\ \vdots\\ I_k\cdot(Y-\mu(a_1^k,a_2^k,X;\theta))\\ \vdots\\ I_K\cdot(Y-\mu(a_1^K,a_2^K,X;\theta))\end{pmatrix},$$

where $D_k = \partial\mu(a_1^k,a_2^k,X;\theta)/\partial\theta^T$ is the model design vector under the k-th embedded DTR denoted by (a_1^k,a_2^k), and I_k is shorthand for the indictor $I(\bar{A}_2$ is consistent with DTR $(a_1^k,a_2^k))$. The weight $w(X,\bar{L}_2,R,\bar{A}_2)$ is the inverse of the product of the probability density function of A_1 given (X,L_1) and A_2 given (X,L_1,A_1,R,L_2). These weights are known, by design; for notational simplicity, we denote $w(X,\bar{L}_2,R,\bar{A}_2)$ as W. While the expression for weights is written in general terms as a function of $(X,\bar{L}_2,R,\bar{A}_2)$, for the autism SMART, the weights are only a function of (A_1,R,A_2). Note that for each individual, we have conceptualized design vector and error for each of all the embedded DTRs; however, only those errors associated to the DTRs that this individual is consistent with will have non-zero values.

4.3 The Variance Estimator for WR Estimator

Standard Taylor series arguments (around θ_0, the true value for parameter θ) can be used to obtain the asymptotic covariance matrix of $\hat{\theta}$:

$$M_n^{-1} E\left(\sum_{i=1}^{n} W_i D_i^T \epsilon_i(\theta_0)\epsilon_i(\theta_0)^T D_i W_i\right) M_n^{-1}, \tag{4.3}$$

where $M_n = -E\left(\sum_{i=1}^{n} W_i D_i^T \frac{\partial \epsilon_i(\theta_0)}{\partial \theta^T}\right)$. The expectation in the formula is with respect to the randomness of the data conditional on the baseline X; in particular, D_i is not considered random, but W_i, ϵ_i are considered random. In practice, M_n is estimated by $-\sum_{i=1}^{n} W_i D_i^T \frac{\partial \epsilon_i(\hat{\theta})}{\partial \theta^T}$, and $E\left(\sum_{i=1}^{n} W_i D_i^T \epsilon_i(\theta_0)\epsilon_i(\theta_0)^T D_i W_i\right)$ is estimated by $\sum_{i=1}^{n} W_i D_i^T \epsilon_i(\hat{\theta})\epsilon_i(\hat{\theta})^T D_i W_i$.

When the sample size is relatively small, $\hat{\theta}$ can deviate from θ_0 by a non-ignorable amount. More often, the estimated $\hat{\theta}$ may result in $\epsilon(\hat{\theta})$ that vary less than $\epsilon(\theta_0)$. This phenomenon is well-known in the GEE literature [34]. Therefore, using $\hat{\theta}$ to replace the unknown θ_0 in the sandwich estimator of the asymptotic variance of $\hat{\theta}$ may induce considerable amount of small-sample bias. Below we focus on correcting the small-sample bias of using $\sum_{i=1}^{n} W_i D_i^T \epsilon_i(\hat{\theta})\epsilon_i(\hat{\theta})^T D_i W_i$ as an estimator for $E\left(\sum_{i=1}^{n} W_i D_i^T \epsilon_i(\theta_0)\epsilon_i(\theta_0)^T D_i W_i\right)$. We will discover later that, in the context of WR estimator, the small-sample correction approach proposed by [34] cannot be applied in a straightforward way, because (i) data from one individual may contribute to the estimation of multiple DTRs, and (ii) the weight used in the WR estimator is a function of post-treatment covariates, thus must be considered as "random", unlike the predictors in a GEE.

From now on we re-define the "error vector" $e_i = W_i \epsilon_i(\theta_0)$ and the "residual vector" $r_i = W_i \epsilon_i(\hat{\theta})$. The way we define the errors and the residuals differs from GEE, in that they incorporate weighting and replication; weighting and replication is random for each subject, because the treatments and the response status are random. Then the middle piece in (4.3) is $\sum_{i=1}^{n} D_i^T E[e_i e_i^T] D_i$, which is normally estimated by $\sum_{i=1}^{n} D_i^T r_i r_i^T D_i$. Below we attempt to quantify the gap between them.

Consider a first-order Taylor series expansion of r_i about θ: $r_i = e_i + \frac{\partial e_i}{\partial \theta^T}(\hat{\theta} - \theta_0)$. A first-order approximation gives

$$(\hat{\theta} - \theta_0) \approx M_n^{-1} \sum_{i=1}^{n} D_i^T e_i. \tag{4.4}$$

Thus we can derive

$$E[r_i r_i^T] \approx E[e_i e_i^T] + E\left[\frac{\partial e_i}{\partial \theta^T}\left(M_n^{-1}\sum_{j=1}^{n} D_j^T e_j\right) e_i^T\right] \tag{4.5}$$
$$+ E\left[e_i\left(M_n^{-1}\sum_{j=1}^{n} D_j^T e_j\right)^T \frac{\partial e_i^T}{\partial \theta}\right] + E\left[\frac{\partial e_i}{\partial \theta^T}\left(M_n^{-1}\sum_{j=1}^{n} D_j^T e_j\right)\left(M_n^{-1}\sum_{j=1}^{n} D_j^T e_j\right)^T \frac{\partial e_i^T}{\partial \theta}\right].$$

This equation quantifies the difference between $e_i e_i^T$ and $r_i r_i^T$, on average. Some further approximations are needed to simplify the calculation. First, we ignore all the terms in (4.5) that involve interaction between e_i and e_j where $i \neq j$. Then we get

$$E[r_i r_i^T] \approx E[e_i e_i^T] + E\left[\frac{\partial e_i}{\partial \theta^T} M_n^{-1} D_i^T e_i e_i^T\right] \tag{4.6}$$
$$+ E\left[e_i e_i^T D_i M_n^{-1} \frac{\partial e_i^T}{\partial \theta}\right] + E\left[\frac{\partial e_i}{\partial \theta^T} M_n^{-1} \sum_{j=1}^{n} D_j^T e_j e_j^T D_j M_n^{-1} \frac{\partial e_i^T}{\partial \theta}\right].$$

Notice that $\left(-\frac{\partial e_i}{\partial \theta^T} M_n^{-1} D_i^T\right)$ plays a crucial role in forming the gap between $E[r_i r_i^T]$ and $E[e_i e_i^T]$. Although M_n is unknown, we can use its estimate, $-\sum_{i=1}^{n} W_i D_i^T \frac{\partial \epsilon_i(\hat{\theta})}{\partial \theta^T}$, to construct an approximation to $\left(-\frac{\partial e_i}{\partial \theta^T} M_n^{-1} D_i^T\right)$. Thus we define

$$H_{ij} := \frac{\partial e_i}{\partial \theta^T}\left(\sum_{k=1}^{n} W_k D_k^T \frac{\partial \epsilon_k(\hat{\theta})}{\partial \theta^T}\right)^{-1} D_j^T = W_i \frac{\partial \epsilon_i(\hat{\theta})}{\partial \theta^T}\left(\sum_{k=1}^{n} W_k D_k^T \frac{\partial \epsilon_k(\hat{\theta})}{\partial \theta^T}\right)^{-1} D_j^T,$$

in particular,

$$H_{ii} := \frac{\partial e_i}{\partial \theta^T}\left(\sum_{k=1}^{n} W_k D_k^T \frac{\partial \epsilon_k(\hat{\theta})}{\partial \theta^T}\right)^{-1} D_i^T = W_i \frac{\partial \epsilon_i(\hat{\theta})}{\partial \theta^T}\left(\sum_{k=1}^{n} W_k D_k^T \frac{\partial \epsilon_k(\hat{\theta})}{\partial \theta^T}\right)^{-1} D_i^T. \tag{4.7}$$

Note that H_{ii} is a K-by-K square matrix, where K is the total number of embedded DTRs in the SMART (e.g., $K = 3$ in the autism example). (4.6) is then rewritten as

$$E[r_i r_i^T] \approx E[e_i e_i^T] - E\left[H_{ii} e_i e_i^T\right] - E\left[e_i e_i^T H_{ii}^T\right] + E\left[\sum_{j=1}^{n} H_{ij} e_j e_j^T H_{ij}^T\right].$$

We further make the approximation that $H_{ij} \approx 0$ for $i \neq j$; the similar type of approximation was also made in [34] in the GEE setting. Then we have

$$E[r_i r_i^T] \approx E[e_i e_i^T] - E\left[H_{ii} e_i e_i^T\right] - E\left[e_i e_i^T H_{ii}^T\right] + E\left[H_{ii} e_i e_i^T H_{ii}^T\right]. \tag{4.8}$$

Here, H_{ii} can not be extracted out of the expectation, because H_{ii} defined in (4.7) not only depends on the baseline covariates, but also depends on the treatment sequences and response status that are post treatment. Thus it is not straightforward to recover $E[e_i e_i^T]$ from $E[r_i r_i^T]$.

We propose an *ad hoc* approach to correcting the bias, suggested by (4.8). The gap between $E[e_i e_i^T]$ and $E[r_i r_i^T]$ is approximately $E\left[H_{ii} e_i e_i^T\right] + E\left[e_i e_i^T H_{ii}^T\right] - E\left[H_{ii} e_i e_i^T H_{ii}^T\right]$. Although e_i is unknown, it can be approximated by r_i, and the remaining bias would be of higher order. That is, in place of $r_i r_i^T$ in the original estimator for the covariance for $\hat{\theta}$, we use $r_i r_i^T + H_{ii} r_i r_i^T + r_i r_i^T H_{ii}^T - H_{ii} r_i r_i^T H_{ii}^T$. Therefore, the adjusted estimator for the middle piece in (4.3) is $\sum_{i=1}^n D_i^T (r_i r_i^T + H_{ii} r_i r_i^T + r_i r_i^T H_{ii}^T - H_{ii} r_i r_i^T H_{ii}^T) D_i$.

4.3.1 The Variance Estimator for WR Estimator with Estimated Weights

When the weights in the WR estimator are estimated rather than the known ones, some further adjustments are necessary. The estimating equation for θ can now be written as $0 = \sum_{i=1}^n W_i(\hat{\alpha}) D_i^T \epsilon_i(\theta)$. $W(\alpha)$ is a model for the weight and $\hat{\alpha}$ is the maximum likelihood estimator; we assume that it is the solution to the estimating equation $0 = \sum_{i=1}^n S_{\alpha,i}(\alpha)$. Now, the asymptotic covariance matrix of $\hat{\theta}$ can be written as

$$M_n^{-1} E\left[\sum_{i=1}^n \left(W_i D_i^T \epsilon_i(\theta_0) - \Pi[W_i D_i^T \epsilon_i(\theta_0) | S_{\alpha,i}]\right)^{\otimes 2}\right] M_n^{-1}, \tag{4.9}$$

where S_α is $S_\alpha(\alpha)$ evaluated at the true value α_0; $\Pi[V|S_\alpha]$ is the projection of V on the space of S_α; $V^{\otimes 2} = VV^T$.

Using the notation defined previously, by taking D_i outside the expectation, the middle piece in (4.9) can be written as $\sum_{i=1}^n D_i E\left(e_i - \Pi[e_i|S_{\alpha,i}]\right)^{\otimes 2} D_i^T$. Thus the goal now is to quantify the bias that is introduced when estimating this quantity by $\sum_{i=1}^n D_i \left(r_i - \Pi[r_i|S_{\alpha,i}]\right)^{\otimes 2} D_i^T$. Since $\Pi[e_i|S_{\alpha,i}]$ is relatively negligible compared to e_i, we will focus on correcting the bias incurred by replacing $e_i e_i^T$ with $r_i r_i^T$, and ignore the bias incurred by replacing $\Pi[e_i|S_{\alpha,i}]$ with $\Pi[r_i|S_{\alpha,i}]$.

Note that, when the weights are estimated, the deviation of $\hat{\theta}$ from the true value θ_0 is no longer characterized by (4.4). Instead, we have

$$(\hat{\theta} - \theta_0) \approx M_n^{-1} \sum_{i=1}^n \left(D_i^T e_i - \Pi[D_i^T e_i | S_{\alpha,i}]\right). \tag{4.10}$$

Thus we have

$$E[r_i r_i^T] \approx E[e_i e_i^T] + E\left[\frac{\partial e_i}{\partial \theta^T}\left(M_n^{-1}\sum_{j=1}^{n}\left(D_j^T e_j - \Pi[D_j^T e_j|S_{\alpha,j}]\right)\right)e_i^T\right] \tag{4.11}$$
$$+ E\left[e_i\left(M_n^{-1}\sum_{j=1}^{n}\left(D_j^T e_j - \Pi[D_j^T e_j|S_{\alpha,j}]\right)\right)^T\frac{\partial e_i^T}{\partial \theta}\right]$$
$$+ E\left[\frac{\partial e_i}{\partial \theta^T}\left(M_n^{-1}\sum_{j=1}^{n}\left(D_j^T e_j - \Pi[D_j^T e_j|S_{\alpha,j}]\right)\right)\left(M_n^{-1}\sum_{j=1}^{n}\left(D_j^T e_j - \Pi[D_j^T e_j|S_{\alpha,j}]\right)\right)^T\frac{\partial e_i^T}{\partial \theta}\right].$$

Here, in addition to the approximations made before in the case of using known weights, we further approximate the third line in (4.11) with $E[H_{ii}e_i e_i^T H_{ii}^T]$ (i.e., some lower order terms are approximated by zero). After all the simplifications, we have

$$E[r_i r_i^T] \approx E[e_i e_i^T] - E\left[H_{ii}e_i e_i^T\right] - E\left[e_i e_i^T H_{ii}^T\right] + E\left[H_{ii}e_i e_i^T H_{ii}^T\right] \tag{4.12}$$
$$- E\left[\frac{\partial e_i}{\partial \theta^T}M_n^{-1}\Pi[D_i^T e_i|S_{\alpha,i}]e_i^T\right]$$
$$- E\left[\frac{\partial e_i}{\partial \theta^T}M_n^{-1}\Pi[D_i^T e_i|S_{\alpha,i}]e_i^T\right]^T.$$

Motivated by this approximation, we propose the following *ad hoc* bias-corrected estimator for the middle piece in (4.9): $\sum_{i=1}^{n} D_i^T\Big((r_i - \hat{\Pi}[r_i|S_{\alpha,i}])^{\otimes 2} + H_{ii}r_i r_i^T + r_i r_i^T H_{ii}^T - H_{ii}r_i r_i^T H_{ii}^T - \left[W_i\frac{\partial \epsilon_i(\hat{\theta})}{\partial \theta^T}\left(\sum_{j=1}^{n}W_j D_j^T\frac{\partial \epsilon_j(\hat{\theta})}{\partial \theta^T}\right)^{-1}\hat{\Pi}[D_i^T r_i|S_{\alpha,i}]r_i^T\right]$
$- \left[W_i\frac{\partial \epsilon_i(\hat{\theta})}{\partial \theta^T}\left(\sum_{j=1}^{n}W_j D_j^T\frac{\partial \epsilon_j(\hat{\theta})}{\partial \theta^T}\right)^{-1}\hat{\Pi}[D_i^T r_i|S_{\alpha,i}]r_i^T\right]^T\Big)D_i$, where we calculate $\hat{\Pi}[V_i|S_{\alpha,i}]$ by taking the fitted values in the regression of V on $S_\alpha(\hat{\alpha})$.

4.4 Simulation Studies for the Small-sample Variance Estimator

A small set of simulation experiments were conducted to investigate the performance of the proposed small-sample bias-corrected variance estimator for the WR estimator. In particular, we compared the confidence intervals constructed based on the bias-corrected variance estimator, and confidence intervals constructed based on the simple plug-in sandwich variance estimator. The comparison was made with both WR estimator with known weights and WR estimator with estimated weights. Data $(X, L_1, A_1, R, L_2, A_2, Y)$ were generated

to mimic the autism SMART study. Here, X includes age, gender and ethnicity; L_1 and L_2 are the numbers of socially communicative utterances at baseline and week 12, respectively. Note that L_1, L_2 and Y are the repeated measures of the number of socially communicative utterances of a subject; although L_1, L_2 are not in the marginal model, they will be used in the weight model for the WR estimator with estimated weights.

The marginal mean model (4.1) is estimated in all the experiments. Since the goal of the simulations is to evaluate the performance of standard error estimators, rather than the performance of the estimators, we use generative models which all imply that (4.1) is a correctly specified marginal mean model for the outcome. More specifically, we create four simulation scenarios in which the marginal means of the outcome under the embedded DTRs do not vary, but the within-person correlations among L_1, L_2, Y are 0, 0.3, 0.6, 0.9. Therefore, under these simulation scenarios, estimating the weights using L_1, L_2 is expected to provide different levels of efficiency improvement. Data sets with a sample size $n = 100$ are generated.

For each data-generative scenario, we apply two estimators: (a) WR estimator with known weights; (b) WR estimator with estimated weights (predictors in stage one weight model are X and L_1; predictors in stage two weight model are L_1 and L_2). For each estimator, we implement two variance estimators: the sandwich estimator without bias correction and the sandwich estimator with the proposed bias correction. In the implementation of the former, we use a naive degree-of-freedom adjustment for the middle piece of the sandwich estimator, i.e., to obtain the plug-in estimate, we use a denominator equal to $(n - (p + q))$ (when known weights are used) or $(n - (p + q) - p_\alpha)$ (when estimated weights are used; p_α is the total number of parameters in the estimated weight model).

We present results for two pairwise comparisons: Δ_1 (the mean difference between DTRs (1, 1) and (1, -1)) and Δ_2 (the mean difference between DTRs (-1, $\cdot$) and (1, -1)). We report the coverage of the confidence intervals constructed based on two different variance estimators, for each of the two WR estimators.

Table 4.1: Coverage of confidence intervals constructed by plug-in sandwich estimators (Plug-in) and sandwich estimators with small-sample bias correction (BC), for the variances of WR estimators, in four simulation scenarios where the within-person correlation (ρ) among L_1, L_2, Y varies. Δ_1 = the mean difference between DTRs (1, 1) and (1, -1); Δ_2 = the mean difference between DTRs (-1, $\cdot$) and (1, -1).

	Δ_1				Δ_2			
	Known wts		Est wts		Known wts		Est wts	
ρ	Plug-in	BC	Plug-in	BC	Plug-in	BC	Plug-in	BC
0	93.6	94.7	89.3	91.8	94.2	95.2	96.2	94.7
0.3	94.8	95.4	90.8	91.9	95.9	96.5	96.2	95.8
0.6	94.5	95.2	90.5	93.9	94.1	94.1	94.3	96.4
0.9	93.1	94.0	88.7	94.3	94.6	95.1	96.9	97.8

Results are shown in Table 4.1. We notice that, confidence intervals based on the sandwich estimator without bias correction have some under coverage, even with the degree-of-freedom adjustment. In particular, when weights are estimated in the WR estimator, the confidence intervals for the difference between DTRs (1, 1) and (1, -1) have coverage probabilities that are below 90%. Using the small-sample bias correction, the coverage for this estimand is improved to at least 92%. We also notice that, for the other estimand, i.e., the difference between DTRs (-1, $\cdot$) and (1, -1), both variance estimators seem to be somewhat conservative in some simulation scenarios, especially for the WR estimator with estimated weights.

4.5 Conclusion and Discussion

Comparison among the embedded DTRs of a SMART study in terms of an end-of-study outcome can be a primary aim of a SMART study. This aim can be achieved by proposing a marginal mean model and estimating the model with a weighted-and-replicated estimator. There are some research questions regarding the WR estimator for relatively small sample sizes, when the asymptotics may not accurately characterize the performance of the estimators. This chapter proposes and investigates the performance of a small-sample bias-corrected variance estimator for the WR estimator. In the simulations we have found that the proposed variance estimator gives a confidence interval with no worse performance than the traditional sandwich estimator, and can improve the coverage probability in some scenarios. Because of the special properties of a WR estimator (i.e., the estimator involves weighting and replication, both of which are "random", conditional on the baseline covariates), the approach proposed by [34] is not directly applicable to the WR estimator. The bias-corrected variance estimator proposed in this chapter extends the idea in [34], and uses some approximations to simplify the form of the estimator. In Almirall et al., we also examine via simulations, the variance estimator obtained by naively adopting the method in [34] without acknowledging the special properties of a WR estimator for the marginal mean model.

Currently the confidence intervals of the estimands are constructed using the estimated standard errors and a z-score. Alternatively, one may construct the confidence interval using a test statistic that has a chi-squared distribution with certain degrees of freedom under the null. Properly identifying this degree-of-freedom may further improve the performance of the confidence interval. We will investigate this in future research.

CHAPTER 5

Regularized Search within a Restricted Class of Treatment Policies

5.1 Introduction

One natural extension to the assisted estimator for evaluating and comparing competing treatment policies, is to estimate the optimal one among a pre-specified class of parametrized treatment policies, that is, the one that yields the highest mean of the primary outcome. The pre-specified class of treatment policies, for example, may be a class of linear decision rules that involve a set of variables suggested by the clinical scientists; it is believed that these variables, when properly used to tailor the treatment assignment, lead to personalized treatments that will yield higher value of the outcome. In addition to identifying the optimal policy within such a pre-specified policy class, it is also of interest to investigate to what extent one variable is useful for decision-making, in the context of all the other variables that are already in the decision rule. This is because including one additional variable in the decision rule given a set of variables in the decision rule may not further increase the highest achievable value; or including this additional variable may only further increase the highest achievable value by a small amount that is comparable to the noise level of the data. In particular, we would like to detect this type of scenario if this additional variable is expensive or difficult to collect or measure in practice. These thoughts are closely related to the concept of "value of information" in the decision theory literature [16, 17]. Value of information of one variable in the decision making procedure, roughly speaking, is how much it is worth to observe this variable, in terms of the average benefit in the utility function by making decision based on the more complete information with the value of this variable revealed. Therefore, in most cases, value of information of a variable would be the amount one is willing to pay to reveal this variable. In this framework, benefits and costs are calculated with the same scale (e.g., in terms of the economic impact), whereas we will

concentrate more on the benefit in terms of a pre-specified primary outcome and omit the economic cost-benefit analysis.

There have been relatively limited works related to determining the important decision-making variables in the context of optimizing treatment policies. In particular, all the existing methods do not consider a pre-specified policy class. [32] considers selecting decision-making variables via a penalized regression framework that is based on modeling and estimation of the effect of treatment interacting with prognostic factors, i.e., A-learning. This approach does not require a prediction model for the entire conditional mean of the outcome given the history as in Q-learning; it builds only on a prediction model for the "treatment blip", thus is robust to mis-specification of the main effect of the prognostic factors in the conditional mean of the outcome. However, one potential disadvantage of this type of method is that, in practice, clinicians or the agents that assign the treatments may not be willing to use many variables to tailor the treatments, either because collecting some variables in a clinical setting is expensive or burdensome, or because of lack of interpretability when the decision rule is too complicated (e.g., involves too many variables). In these cases, they might propose a couple of variables that they are particularly willing to use to tailor the treatment assignment, i.e., to build the decision rules. The A-learning approach can only produce an estimated optimal policy in this desired form by using only this set of variables in the prediction model for the treatment blip. It is very likely that this set of variables proposed by the clinicians cannot accurately predict the treatment blip; in that case bias will be incurred in the estimation, thus affecting the value associated with the estimated optimal decision rule. Moreover, a variable that quantitatively interacts with the treatment may not be necessary to be included for the optimality of the decision rule [12]. Variable selection approach based on A-learning cannot successfully eliminate these types of variables. [88] attempts to identify important variables in the context of optimizing treatment policies in her unpublished PhD dissertation, where a two-step method is proposed to identify all the variables that have qualitative interactions with the treatment; the first step is a flexible fit of the potential outcomes under each treatment regime using state-of-the-art machine learning algorithms, and the second step utilizes a sparse classification method to do variable selection for the optimal treatment regime. This method, however, is also directly targeting variable selection for the optimal treatment regime, and is not applicable when the regimes of interest are in a pre-specified restricted class.

To avoid the problems that are encountered when taking the prediction model approach, and to be able to investigate the problem in the context of a pre-specified class of policies, we propose to build an optimization procedure based on consistent estimation of the policy values.

5.2 Problem Formulation, Challenges and Proposals

5.2.1 Problem Formulation

Consider a one-stage decision-making problem with binary treatment ($a = 0, 1$), and suppose that the class of policies/decision rules that we want to optimize among is $I\{c_0 + c_1 S_1 + ... + c_p S_p > 0\}$. This class of policies is indexed by $(p+1)$-dim parameter $c = (c_0, c_1, ..., c_p)$. Throughout we will assume that $S_1, ..., S_p$ are all standardized, i.e., each has a zero mean. Suppose our observed data is $\{(X_i, A_i, Y_i)\}_{i=1}^n$ and $(S_1, ..., S_p)$ is a subset of X; the treatment A_i is assigned according to some known randomization probability. Then the value of the policy indexed by c, $V^{(d)}(c)$, can be estimated using the IPW estimator $\hat{V}_n^{(d)}(c)$. Since the policies in the class are deterministic policies, the estimator $\hat{V}_n^{(d)}(c)$ is usually discontinuous in c, therefore optimizing $\hat{V}_n^{(d)}(c)$ over c to obtain the estimated optimal policy would be a difficult optimization problem (this issue of estimation in finite sample motivates us to consider a class of stochastic policies). Another equally important problem is the ill-posedness of the optimization of the policy value. Simply put, it is very likely that the maximizer of $V^{(d)}(c)$ is non-unique. Note that c and $K \cdot c$ specify an identical policy, thus we can restrict our attention to $\{c : \|c\| = 1\}$. This is a compact set; if $V^{(d)}(c)$ were continuous in c, then $V^{(d)}(c)$ would have a maximum on $\{c : \|c\| = 1\}$. On $\{c : \|c\| = 1\}$, there might be multiple c's that achieve the maximal value of $V^{(d)}(\cdot)$. As a result of this ill-posedness, whether a variable is useful for decision making cannot be easily detected from the optimizer(s) of the value function. We will illustrate this point later in our investigation for stochastic policies.

To resolve the discontinuity in estimated policy value in finite sample, we propose to relax the original policy search problem by searching over a larger class of stochastic policies. This class of stochastic policies uses the expit function to smooth the indicator function involved in the original deterministic policies. That is, we consider stochastic policies that assign $a = 1$ with probability $\pi_\theta(S) = \exp(\theta_0 + \theta_1 S_1 + ... + \theta_p S_p)/(1 + \exp(\theta_0 + \theta_1 S_1 + ... + \theta_p S_p))$ and $\theta = (\theta_0, \theta_1, ..., \theta_p)$. Note that this stochastic policy class in fact contains the deterministic policies as degenerate cases. From now on we consider the stochastic policy class not as a computational tool to relax the original problem, but as our new pool from which the policy search is conducted. By switching from the deterministic policy class to the stochastic policy class, we are essentially expanding the candidate policies.

5.2.2 The Policy Search Problem - At Population Level

Let $V(\theta)$ denote the policy value associated to θ. $V(\theta) = E_\theta Y = E\big[E[Y|S, A = 1] \cdot \pi_\theta(S) + E[Y|S, A = 0] \cdot (1 - \pi_\theta(S))\big]$, where E_θ denotes the expectation in the population where the stochastic policy prescribed by $\pi_\theta(S)$ is followed. It can also be written as $E\left[Y \cdot \frac{A\pi_\theta(S)+(1-A)(1-\pi_\theta(S))}{p(A|X)}\right]$, where $p(A|X)$ is the randomization distribution of A given X in the observed data. The IPW estimate $\hat{V}_n(\theta) = \mathbb{P}_n\left[Y \cdot \frac{A\pi_\theta(S)+(1-A)(1-\pi_\theta(S))}{p(A|X)}\right]$ is continuous in θ; thus the discontinuity problem is resolved. Under mild finite moment conditions, $V(\theta)$ can be shown to be bounded. We denote $\sup_\theta V(\theta)$ as V^*. The (population-level) ill-posedness issue in the context of a class of stochastic policies is similar to the issue with the class of deterministic policies. Roughly speaking, the optimizer of $V(\theta)$ is non-unique; in particular, this lack of uniqueness makes it difficult to claim whether or not a certain variable is useful for decision making.

However, now with a class of stochastic policies, we can no longer restrict the domain to a compact set. In particular, note that as the coordinates of θ go to infinity in a direction, the associated stochastic policy degenerates to a deterministic policy. For example, if $\theta = (\theta_0, \theta_1)$ in $\pi_\theta(S) = \exp(\theta_0 + \theta_1 S_1)/(1 + \exp(\theta_0 + \theta_1 S_1))$ goes to infinity in direction (c_0, c_1), then it degenerates to the deterministic policy that assigns $a = 1$ if and only if $(c_0 + c_1 S_1) > 0$. Besides, if some of the coordinates of θ stay finite and the others go to infinity in some direction, then the stochastic policy degenerates to a deterministic policy that is only determined by those infinite coordinates of θ. For example, if (θ_0, θ_1) in $\pi_\theta(S) = \exp(\theta_0 + \theta_1 S_1 + \theta_2 S_2)/(1 + \exp(\theta_0 + \theta_1 S_1 + \theta_2 S_2))$ goes to infinity in direction (c_0, c_1) and $\theta_2 = \theta_{20}$ (θ_{20} is an arbitrary finite number), then the policy degenerates to the deterministic policy that assigns $a = 1$ if and only if $(c_0 + c_1 S_1) > 0$.

For now we do not allow coordinates of θ to take values in $\{-\infty, +\infty\}$. Then the ill-posedness issue, which was mentioned in the discussion about optimizing deterministic policies, must be framed in a slightly different way. Suppose we desire to make inference about S_p. We say that optimization of $V(\theta)$ is *ill-posed with respect to* S_p, if there exists $\delta > 0$ such that for an arbitrarily small $\epsilon > 0$, we can find θ and θ' that both have a value higher than $V^* - \epsilon$, and satisfy $|\theta_p| \geq \delta$, yet $\theta'_p = 0$. We choose this definition of ill-posedness because, unlike in the case of deterministic policies, here the supremum V^* may not be attainable; thus it is more appropriate to talk about ϵ-optimizers for an arbitrarily small ϵ.

In this scenario, because of the existence of θ', we can say that S_p is *not useful* for decision making given the other variables in the policy. However, without any regularization of this (population-level) optimization problem, one might not be able to conclude that S_p

can be eliminated from the policy with no loss, because we may accidentally obtain an optimizer with nonzero θ_p instead of the $\theta'_p = 0$.

To better understand the issue of ill-posedness, consider the example below. We consider a policy class $\pi_\theta(S) = \exp(\theta_0 + \theta_1 S_1)/(1 + \exp(\theta_0 + \theta_1 S_1))$; this policy class only involves the intercept and one "tailoring variable" S_1. Consider a generative model for Y: $Y = A + \epsilon$, where $\epsilon \sim N(0, 1)$. This generative model illustrates the ill-posedness, in terms of S_1, of optimizing among this policy class. Without looking at the specified policy class, it is clear that optimality is attained if $a = 1$ is assigned to each participant; i.e., S_1 does not need to be used to tailor the treatment to achieve optimal value.

It is obvious to see that $V^* = 1$. Let S_1 be uniformly distributed in (-1,+1). Then it is easy to show that, for any small $\epsilon > 0$, if we let $\theta_0 = 2 + \log(\frac{1}{\epsilon} - 1)$, then $\theta = (\theta_0, 0)$ and $\theta' = (\theta_0, 1)$ satisfy $V(\theta) > 1 - \epsilon$ and $V(\theta') > 1 - \epsilon$. In other words, here ill-posedness is that the ϵ-optimal level optimizers (for any ϵ) include optimizers with zero coefficient and non-zero coefficient for S_1.

Note that, in general, "ill-posedness" may refer to the non-uniqueness of the maximizer with respect to any coordinate in θ. In our framework though, since the goal (aside from the main goal of optimizing the value) is to determine whether S_p has additional use in decision making given that all the other variables are already used in the decision rule, we concentrate on the ill-posedness with respect to this particular variable S_p. This particular type of ill-posedness motivates a regularization of the original policy search problem. In other words, when this type of ill-posedness arises, even when the amount of data is infinite, "inference" about the usefulness of S_p cannot be readily drawn by inspecting the p-th coordinate in the (ϵ-)optimizer of $V(\theta)$.

One approach to regularizing the problem at population level is introducing penalty to θ_p, the coefficient of S_p, in the optimization of $V(\theta)$. Using an L_2 penalty, the optimization objective can be $V_\lambda(\theta) = V(\theta) - \frac{1}{2}\lambda\|\theta_p\|_2^2$. The aim is that with the penalty, the optimization is no longer ill-posed with respect to S_p. Therefore, with a proper penalty at level λ, the regularized objective clearly indicates (through its optimizer) whether or not S_p is useful for decision making. More specifically, it is desirable that under the λ-level penalty, either all the optimizer θ's for $V_\lambda(\theta)$ have $|\theta_p| \geq \delta$ for some $\delta > 0$ (then variable S_p is in fact useful), or all the optimizer θ's for $V_\lambda(\theta)$ have $\theta_p = 0$ (then variable S_p is in fact not useful). Since later we will be motivated to also penalize the other coefficients, we denote this penalty parameter as λ_1.

5.2.2.1 Tuning Parameter λ_1

The tuning parameter λ_1 is used to resolve the ill-posedness of the optimization of the (population level) policy value, with respect to the variable, S_p, that we would like to investigate the usefulness in decision making. The quadratic penalty (i.e., $\frac{1}{2}\lambda_1\|\theta_p\|_2^2$) provides curvature when the original value function $V(\theta)$, or more accurately, the "profile value function" $\max_{\theta_0,\ldots,\theta_{p-1}} V(\theta)$ (as a function of θ_p), is flat in θ_p. Therefore it is desirable that under the penalization, the value function has a unique optimizer θ_p.

On the other hand, as the main goal is to optimize the policy value, it is desirable that any optimizer θ_λ^* of $V_\lambda(\theta)$ gives a policy value that is not much lower than V^*, the optimal value with no regularization. Consider the next two toy examples for understanding reasonable choices of λ_1.

- Toy example A: Suppose the generative model for Y is $Y = A + \epsilon$, and consider the policy class $\pi_\theta(S) = \exp(\theta_0 + \theta_1 S_1)/(1 + \exp(\theta_0 + \theta_1 S_1))$. Now the "profile value function" is exactly flat in θ_1; without the penalty, for any θ_1, $\theta_0 = +\infty$ optimizes the value. Therefore, with any positive λ_1, θ_λ^* that optimizes $V_\lambda(\theta)$ is equal to $(+\infty, 0)$; this corresponds to the policy that assigns $a = 1$ to all individuals. That is, under this generative model, λ_1 does not induce a loss in the policy value.

- Toy example B: Suppose the generative model for Y is $Y = A \cdot S_1 + \epsilon$, that is, there is zero main effect and all the treatment effect is the interaction effect with S_1. Consider the policy class $\pi_\theta(S) = \exp(\theta_0 + \theta_1 S_1)/(1 + \exp(\theta_0 + \theta_1 S_1))$. The optimal policy should assign $a = 1$ if and only if $S_1 > 0$. For ease of calculation, further assume that S_1 can only take two values {-1, 1}, each with probability 1/2. Then the optimal value is $V^* = 1/2$. The penalized value function $V_\lambda(\theta)$ is equal to:

$$\begin{aligned} V_\lambda(\theta) &= E\left[S_1 \cdot \exp(\theta_0 + \theta_1 S_1)/(1 + \exp(\theta_0 + \theta_1 S_1))\right] - \frac{1}{2}\lambda_1\|\theta_1\|_2^2 \\ &= \frac{1}{2} \cdot \frac{\exp(\theta_0 + \theta_1)}{1 + \exp(\theta_0 + \theta_1)} - \frac{1}{2} \cdot \frac{\exp(\theta_0 - \theta_1)}{1 + \exp(\theta_0 - \theta_1)} - \frac{1}{2}\lambda_1\|\theta_1\|_2^2. \end{aligned}$$

  The optimizer θ_λ^* of $V_\lambda(\theta)$ should satisfy $V_\lambda'(\theta_\lambda^*) = 0$. That is, θ_λ^* solves the following two equations:

$$\begin{aligned} \frac{1}{2} \cdot \frac{\exp(\theta_0 + \theta_1)}{(1 + \exp(\theta_0 + \theta_1))^2} - \frac{1}{2} \cdot \frac{\exp(\theta_0 - \theta_1)}{(1 + \exp(\theta_0 - \theta_1))^2} &= 0 \\ \frac{1}{2} \cdot \frac{\exp(\theta_0 + \theta_1)}{(1 + \exp(\theta_0 + \theta_1))^2} + \frac{1}{2} \cdot \frac{\exp(\theta_0 - \theta_1)}{(1 + \exp(\theta_0 - \theta_1))^2} - \lambda_1\theta_1 &= 0. \end{aligned}$$

We then obtain that for $\theta^*_\lambda = (\theta^*_{0\lambda}, \theta^*_{1\lambda})$, $\theta^*_{0\lambda} = 0$ and $\theta^*_{1\lambda}$ satisfies:

$$\frac{\exp\theta_1}{(1+\exp\theta_1)^2} = \lambda_1\theta_1. \tag{5.1}$$

Recall that, the penalization parameter λ_1 should be chosen such that the policy value corresponding to θ^*_λ is not much lower than V^*. Next we study more carefully, in the scenario of this toy example, the requirement for λ_1 if the largest amount of policy value loss that one can tolerate is ε. That is, we want to guarantee that $V(\theta^*_\lambda) \geq V^* - \varepsilon$.

$$\begin{aligned} V(\theta^*_\lambda) &= E\left[S\frac{\exp(\theta^*_{1\lambda}S)}{1+\exp(\theta^*_{1\lambda}S)}\right] = \frac{1}{2}\cdot\frac{\exp(\theta^*_{1\lambda})}{1+\exp(\theta^*_{1\lambda})} - \frac{1}{2}\cdot\frac{\exp(-\theta^*_{1\lambda})}{1+\exp(-\theta^*_{1\lambda})} \\ &= \frac{1}{2}\cdot\frac{-1+\exp(\theta^*_{1\lambda})}{1+\exp(\theta^*_{1\lambda})}. \end{aligned}$$

Thus $V(\theta^*_\lambda) \geq V^* - \varepsilon = 1/2 - \varepsilon$ is equivalent to $(1+\exp\theta^*_{1\lambda})^{-1} \leq \varepsilon$. Therefore, $\theta^*_{1\lambda}$ that satisfies (5.1) must make $(1+\exp\theta^*_{1\lambda})^{-1} \leq \varepsilon$ hold.

Note that an upper bound readily obtained from (5.1) is that $\theta^*_{1\lambda} < 1/(4\lambda_1)$, which implies a necessary condition for λ_1: $\lambda_1 \leq \left(\log(\varepsilon^{-1}-1)\right)^{-1}/4$ (a sufficient condition for λ_1 would follow from a lower bound for $\theta^*_{1\lambda}$ obtained from (5.1)). Intuitively, $\theta^*_{1\lambda}$ shrinks towards zero when λ_1 increases (implied by (5.1)); thus to satisfy $(1+\exp\theta^*_{1\lambda})^{-1} \leq \varepsilon$, λ_1 must be bounded properly from above.

To summarize, the regularization parameter λ_1 trades off higher policy value with the level of well-posedness of the policy search problem with respect to S_p, the variable we intend to make inference about. Up to now we only say that the policy search problem is or is not ill-posed with respect to S_p. Note however, that when λ_1 is small, the optimal coefficient for S_p might be unique, but the curvature of the penalized objective with respect to θ_p might still be too small for estimation purposes; in other words, this is when the level of well-posedness is low.

The "value of information" (VoI) of variable S_p mentioned in the introduction is an alternative approach to understanding the importance of a variable in decision making. Here we make some comments about the comparison between taking the regularization approach and focusing on VoI of variable S_p. First of all, VoI is not in the context of a specific class of decision rules. Furthermore, VoI provides a one-number summary of the difference in the optimal achievable values with and without knowing S_p: Vol $= 0$ indicates that S_p is not useful and Vol > 0 indicates that S_p is useful. In practice, it is unlikely that we will encounter the case Vol $= 0$. More likely, the variable S_p is somewhat useful (i.e., Vol

is a non-zero small number); in these cases, knowing how S_p should be used to construct a good decision rule, under various amounts of reduction in the optimal value that can be tolerated, can be a more interesting and practical problem. I conjecture that this problem can be addressed by the proposed regularized approach, by properly tuning the parameter λ_1.

5.2.3 The Policy Search Problem - At Finite-sample Level

We have illustrated above that, in the context of maximizing the policy value, to be able to have an identifiable estimand for the coefficient of the variable that we intend to make inference about, regularization is needed for the objective function at the population level. We proposed to use L_2 regularization. Next we investigate the policy search problem at the finite sample level.

With a finite sample of size n, the goal is to obtain $\hat{\theta}_n$ that estimates the optimal policy, and to have valid inference about the usefulness of variable S_p. More specifically, let $\theta^*_{p\lambda}$ be the coefficient for S_p in the optimizer θ^*_λ of the regularized objective at population level, we need some characterization of the distribution of $\hat{\theta}_{n,p} - \theta^*_{p\lambda}$, for the purpose of constructing a valid confidence interval for $\theta^*_{p\lambda}$.

It is natural to consider optimizing the empirical version of the regularized optimization objective, i.e., $\hat{V}_{n,\lambda}(\theta) = \hat{V}_n(\theta) - \frac{1}{2}\lambda_1\|\theta_p\|_2^2$, to obtain $\hat{\theta}_n$. However, we will discover from a toy example below that, at the finite sample level, penalizing only θ_p is problematic, which motivates another layer of penalization; in this layer the coefficients of all the variables in the specified policy form are penalized.

The intuition is that, when the signal (here the signal refers to the increase in value by including S_p in the policy) is small, if only θ_p is penalized, then optimizing the population-level objective is likely to have an effect of pushing the other coefficients to be finite (penalizing θ_p is equivalent to the optimization of the value with the constraint that $\|\theta_p\|$ is bounded from above). On the other hand, when optimizing the sample-level objective, due to the discrete nature of the sampling distribution $\mathbb{P}_n$, there might be a non-ignorable probability that, for a finite sample of size n, a deterministic policy not involving S_p happens to maximize the estimated value (i.e., coefficients other than θ_p are estimated to be infinite in some direction). Now because the other coefficients are infinite, θ_p becomes unidentifiable; with an L_2 penalty for it, $\hat{\theta}_p$ is forced to be zero and the contribution of S_p in decision making is completely buried.

5.2.3.1 Toy Example: Point Mass of $\hat{\theta}_{n,p}$ at Zero - Motivating the Penalty for All Coefficients

Suppose the generative model for Y is $Y = A \cdot (0.2 + 0.5S_1) + \epsilon$, and S_1 can only take two values $\{-1, 1\}$, each with probability 1/2. Consider the policy class $\pi_\theta(S) = \exp(\theta_0 + \theta_1 S_1)/(1 + \exp(\theta_0 + \theta_1 S_1))$. In this case the value function is $V(\theta) = E\left[Y \cdot (A\pi_\theta(S) + (1-A)(1-\pi_\theta(S))) \cdot 2\right] = E[2A \cdot (0.2 + 0.5S_1)\pi_\theta(S)] = E[(0.2 + 0.5S_1)\pi_\theta(S)]$. The optimal policy is to assign $a = 1$ if $S_1 = 1$ and assign $a = 0$ if $S_1 = -1$ (note that this rule can actually be achieved by letting (θ_0, θ_1) in the specified policy class go to infinity in many directions).

To obtain estimator $\hat{\theta}_n = (\hat{\theta}_{0n}, \hat{\theta}_{1n})$ for the coefficients in an optimal policy within the specified class, consider optimizing $\hat{V}_{n,\lambda}(\theta)$, the objective in which θ_1 has a L_2 penalty and no penalty is applied for θ_0, using a sample of size n. The structure of the data is: (S_i, A_i, Y_i) for $i = 1, ..., n$, where A_i is randomized to take values in $\{0, 1\}$ with equal probability. The objective function based on this finite sample is:

$$\begin{aligned}\hat{V}_{n,\lambda}(\theta) &= \mathbb{P}_n Y \cdot (A\pi_\theta(S) + (1-A)(1-\pi_\theta(S))) \cdot 2 - \lambda_1 \|\theta_1\|_2^2/2 \\ &= \mathbb{P}_n \left[2(A(0.2 + 0.5S_1) + \epsilon)(2A-1)\pi_\theta(S)\right] - \lambda_1 \|\theta_1\|_2^2/2 + \text{terms not involving } \theta. \end{aligned} \tag{5.2}$$

Note that the summand is equal to $(1.4A + (4A-2)\epsilon)\pi_\theta(1)$ if $S = 1$, and it is equal to $(-0.6A + (4A-2)\epsilon)\pi_\theta(-1)$ if $S = -1$. θ corresponds to the two probabilities p_1 and p_{-1}, the probabilities of assigning $a = 1$ for $S = 1$ and $S = -1$, respectively. Specifically, $p_1 = \pi_\theta(1) = expit(\theta_0 + \theta_1)$ and $p_{-1} = \pi_\theta(-1) = expit(\theta_0 - \theta_1)$. Therefore the summation in the previous display (i.e., the objective function without the penalty for θ_1) can be rewritten as

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^{n}(-0.6A_i + (4A_i - 2)\epsilon_i) I\{S_i = -1\} \cdot p_{-1} \\ +&\frac{1}{n}\sum_{i=1}^{n}(1.4A_i + (4A_i - 2)\epsilon_i) I\{S_i = 1\} \cdot p_1.\end{aligned}$$

This representation implies that, the solution to the optimization is entirely determined by the two quantities $K := \frac{1}{n}\sum_{i=1}^{n}(-0.6A_i + (4A_i - 2)\epsilon_i) I\{S_i = -1\}$ and $\tilde{K} := \frac{1}{n}\sum_{i=1}^{n}(1.4A_i + (4A_i - 2)\epsilon_i) I\{S_i = 1\}$.

The current regularization scheme only penalizes θ_1 but not θ_0. As a result, when K and $\tilde{K}$ have the same sign (i.e., either both positive or both negative), θ_1 is estimated to be zero. This is because when $K, \tilde{K} > 0$, $p_{-1} = p_1 = 1$ optimizes the objective, and this

implies $\hat{\theta}_0 = +\infty$; when $K, \tilde{K} < 0$, $p_{-1} = p_1 = 0$ optimizes the objective, and this implies $\hat{\theta}_0 = -\infty$. In addition, as a result of the L_2 penalty for θ_1, in these cases $\hat{\theta}_1 = 0$.

We run a small simulation investigating the chance that this would happen. We run an experiment with 1000 samples of size $n = 100$, and $var(\epsilon) = 1$. We find that in 163 out of the 1000 samples, $K, \tilde{K} > 0$ and in 10 out of the 1000 samples, $K, \tilde{K} < 0$. Therefore in 173 out of the 1000 samples, a deterministic policy that does not involve S_1 happens to optimize the estimated value, which implies that the contribution of S_1 is entirely buried.

The toy example above demonstrates that, for a finite sample, when only the coefficient of S_p (the variable about which we intend to make inference) is penalized, a deterministic decision rule not involving S_p may happen to optimize the estimated value. This can cause the distribution of $\hat{\theta}_p$ to have a point mass at zero when the truth is that S_p is useful; this can have an impact on the power of testing the null hypothesis of $\theta_p^* = 0$. Moreover, as the sample size becomes smaller, the sampling distribution $\mathbb{P}_n$ may deviate much from the population distribution P, such that the estimated optimal decision rule varies a lot across different samples (i.e., the coefficients other than θ_p are estimated to be infinite in directions that vary considerably). All of these discussions motivate us to consider penalizing other coefficients than θ_p; the penalization will essentially bring in randomness/stochasticity into the decision rule, i.e., prevent the estimated coefficients from going to infinity.

5.2.3.2 Tuning Parameter λ_{0n}

The discussions above motivate us to consider an estimator that optimizes the following objective:

$$\hat{V}_{n,\lambda}(\theta) = \hat{V}_n(\theta) - \frac{1}{2}\lambda_{0n}\sum_{j=0}^{p-1}\|\theta_j\|_2^2 - \frac{1}{2}\lambda_1\|\theta_p\|_2^2. \tag{5.3}$$

This objective is the empirical version of the following regularized policy value function:

$$V_{n,\lambda}(\theta) = V(\theta) - \frac{1}{2}\lambda_{0n}\sum_{j=0}^{p-1}\|\theta_j\|_2^2 - \frac{1}{2}\lambda_1\|\theta_p\|_2^2.$$

By applying penalty to the coefficients other than θ_p, the optimizer $(\theta_{0n}^*, \theta_{1n}^*)$ of $V_{n,\lambda}(\theta)$ differs from the optimizer (θ_0^*, θ_1^*) of $V_\lambda(\theta)$, the regularized policy value function in which only θ_p is penalized. Moreover, the policy value of the policy indexed by $(\theta_{0n}^*, \theta_{1n}^*)$ might be lower than the value of the policy indexed by (θ_0^*, θ_1^*). Therefore, roughly speaking, λ_{0n} should be chosen so that: (i) it is sufficiently large such that the inference about θ_p is not impaired by the other coefficients being estimated to be at/close to the boundary; and (ii) it is sufficiently small such that the bias in θ_1 (i.e., $|\theta_{1n}^* - \theta_1^*|$), or the loss in policy value, is

no greater than a pre-specified amount.

5.2.4 Summary of the Regularized Estimator for the Optimal Policy

Here we summarize the proposals and discussions above into the following policy search procedure:

- Let the clinicians/scientists propose a number of variables that they would like to use to construct the decision rule. These variables induce a class of stochastic policies, from which we aim to estimate the optimal policy. Moreover, without loss of generality, we assume the variable about which we are interested in the usefulness in decision making is S_p.
- Collect data $\{X_i, A_i, Y_i\}_{i=1}^n$ from a randomized clinical trial.
- Assume that we know the proper choices for the regularization parameters λ_1 and λ_{0n} (they should be chosen based on the amount of policy value that we can sacrifice to trade the ill-posedness of the problem as well as the ill performed distribution of the estimated coefficient; we will investigate how to tune these parameters in future work). We can optimize the estimated policy value function with penalty, shown in (5.3). This optimization yields the coefficients for each of the variables in the estimated optimal policy: $(\hat{\theta}_0, ..., \hat{\theta}_p)$.
- Construct confidence interval for θ_p based on asymptotics (future work). If the confidence interval does not contain zero, then we can conclude that in the context of the proposed policy class, including S_p in the decision rule in addition to all the other variables is useful; if the confidence interval contains zero, then we can conclude that we do not have sufficient evidence to support the usefulness of S_p.

5.2.5 Plan for Future Work

We propose a regularized estimator for the optimal policy within a parametrized class. The regularization consists of two components: one component is necessary for eliminating the ill-posedness issues that occur even with infinite amount of data; the other component is necessary for making valid inference. We have presented discussions about the intuition behind these regularization parameters. We plan to develop theoretical results and practical rules about the choices of these parameters.

We have provided an outline of the estimator for the coefficients in the optimal policy. Because one of our goals is to understand the usefulness of a particular variable S_p in

decision making in the context of the proposed policy class, we are particularly interested in constructing confidence interval for θ_p, the coefficient of S_p; the statement about S_p can then be made based on this confidence interval. We plan to derive a valid confidence interval for θ_p using asymptotics; this can be challenging because we will need to take into account the procedure of tuning the regularization parameters.

Although we discuss the problem and propose the methodology in a quite general setting, the toy examples we provide to illustrate the motivation and intuition are all oversimplified in the sense that the policy class we focus on in our toy examples only involve the intercept and one tailoring variable. When there are multiple variables in the specified policy form, and we are only interested in the usefulness of S_p, the correlation between S_p and the other variables will surely bring in challenges in developing theory of the methodology. We plan to clearly understand those challenges and investigate the solutions. Furthermore, we were motivated to consider a pre-specified parametrized class of policies, as an alternative to directly targeting the optimal policy. Therefore, we will need to investigate toy examples in which the true treatment blip involves some other variables in addition to the variables in the form of specified policy, or toy examples in which the variables in the form of specified policy and the variables that interact with the treatment in the data-generative model are two distinct sets of variables.

We frame the policy search problem in the context of a class of stochastic policies, in order to eliminate the discontinuity issue that arises when working with a class of deterministic policies. Although the deterministic policies also belong to the class of stochastic policies, the estimated optimal policy is likely to be truly stochastic (i.e., for each individual, the recommended probability of taking treatment $a = 1$ stays away from 0 and 1) as a result of the penalization scheme. On the other hand, in practice it is not reasonable for any clinical scientists to implement a stochastic policy. Therefore, we will need to suggest a practice to convert the estimated optimal stochastic policy into an implementable real-world decision rule.

Extension of the proposed methodology to two-stage or multi-stage scenarios is of great interest, and requires deeper thinking, because the earlier-stage treatment may impact the final outcome in both direct and indirect routes.

We have built our regularized estimator based on the original non-parametric IPTW estimator for a policy value (arising from a marginal mean model). It is well known that this estimator has a high variance because it does not utilize any information about the prognostic effect of the observed covariates on the outcome. A natural proposal is to consider building the regularized estimator based on a more efficient version of IPTW estimator, e.g., the augmented IPTW estimator. The hypothesis is that the efficiency in terms of esti-

mating the policy value may induce a more efficient estimator for the optimal policy, and the inference about the usefulness of S_p may also get more accurate.

Simulation studies will be conducted later to investigate the ability of the proposed methodology in eliminating a variable when it is truly not useful, and not eliminating this variable when it is useful. The simulation studies shall be conducted in various scenarios; in particular, we are interested in the performance of the estimator when the targeted variable correlates with the other variables in the policy at various levels. Moreover, we will use simulations to demonstrate the possibility that a variable, which is not useful for decision making, can easily be estimated to have significantly non-zero coefficient in Q-learning, and then be included in the optimal policy derived from Q-learning.

Up to now in terms of the aspect of making inference, we have only discussed the inference about the coefficient of S_p; this inference allows us to conclude whether there is sufficient evidence to support the usefulness of S_p in decision making. The selection of the regularization parameters is in part driven by this inference goal. Another interesting and important inference goal is the inference about the optimal value that can be achieved within the specified policy class. The hypothesis is that if one desires a well-performed confidence interval for the optimal value, a different regularization scheme, or at least a different approach to tuning the parameters, might be necessary.

BIBLIOGRAPHY

[1] Daniel Almirall, Inbal Nahum-Shani, Nancy E Sherwood, and Susan A Murphy. Introduction to smart designs for the development of adaptive interventions: with application to weight loss research. *Translational Behavioral Medicine*, pages 1–15, 2014.

[2] Daniel Almirall, Thomas Ten Have, and Susan A Murphy. Structural nested mean models for assessing time-varying effect moderation. *Biometrics*, 66(1):131–139, 2010.

[3] Oliver Bembom and Mark J van der Laan. Analyzing sequentially randomized trials based on causal effect models for realistic individualized treatment rules. *Statistics in medicine*, 27(19):3689–3716, 2008.

[4] Ronald T Brown, David O Antonuccio, George J DuPaul, Mary A Fristad, Cheryl A King, Laurel K Leslie, Gabriele S McCormick, William E Pelham Jr, John C Piacentini, and Benedetto Vitiello. *Childhood mental health disorders: Evidence base and contextual factors for psychosocial, psychopharmacological, and combined interventions.* American Psychological Association, 2008.

[5] Babette A Brumback. A note on using the estimated versus the known propensity score to estimate the average treatment effect. *Statistics & Probability Letters*, 79(4):537–542, 2009.

[6] P. Chaffee and M.J. Van der Laan. Targeted maximum likelihood estimation for dynamic treatment regimes in sequentially randomized controlled trials. *The International Journal of Biostatistics*, 8, 2012.

[7] B. Chakraborty, E.B. Laber, and Y. Zhao. Inference for optimal dynamic treatment regimes using an adaptive m-out-of-n bootstrap scheme. *Under Revision*, 2014.

[8] B. Chakraborty, S.A. Murphy, and V. Strecher. Inference for non-regular parameters in optimal dynamic treatment regimes. *Statistical Methods in Medical Research*, 19:317 – 343, 2010.

[9] R. Dawson and P.W. Lavori. Efficient design and inference for multistage randomized trials of individualized treatment policies. *Biostatistics*, 13:142 – 152, 2012.

[10] Wentao Feng and Abdus S Wahed. Supremum weighted log-rank test and sample size for comparing two-stage adaptive treatment strategies. *Biometrika*, 95(3):695–707, 2008.

[11] BA Flannery, JR Volpicelli, and HM Pettinati. Psychometric properties of the penn alcohol craving scale. *Alcoholism: Clinical and Experimental Research*, 23(8):1289–1295, 1999.

[12] L Gunter, J Zhu, and S A Murphy. Variable selection for qualitative interactions. *Statistical Methodology*, 8:42 – 55, 2011.

[13] Miguel A Hernán, Babette A Brumback, and James M Robins. Estimating the causal effect of zidovudine on cd4 count with a marginal structural model for repeated measures. *Statistics in medicine*, 21(12):1689–1709, 2002.

[14] Keisuke Hirano, Guido W Imbens, and Geert Ridder. Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica*, 71(4):1161–1189, 2003.

[15] Søren Højsgaard, Ulrich Halekoh, and Jun Yan. The r package geepack for generalized estimating equations. *Journal of Statistical Software*, 15/2:1–11, 2006.

[16] Ronald A Howard. Information value theory. *Systems Science and Cybernetics, IEEE Transactions on*, 2(1):22–26, 1966.

[17] Ronald A Howard and James E Matheson. Influence diagrams. *Decision Analysis*, 2(3):127–143, 2005.

[18] Hendrée E Jones, Kevin E O' Grady, and Michelle Tuten. Reinforcement-based treatment improves the maternal treatment and neonatal outcomes of pregnant patients enrolled in comprehensive care treatment. *The American Journal on Addictions*, 20(3):196–204, 2011.

[19] C. Kasari. Developmental and augmented intervention for facilitating expressive language (CCNIA). *National Institutes of Health, Bethesda, MD*, 2009.

[20] Connie Kasari, Ann Kaiser, Kelly Goods, Jennifer Nietfeld, Pamela Mathy, Rebecca Landa, Susan Murphy, and Daniel Almirall. Communication interventions for minimally verbal children with autism: A sequential multiple assignment randomized trial. *Journal of the American Academy of Child & Adolescent Psychiatry*, 53(6):635–646, 2014.

[21] Kelley M Kidwell and Abdus S Wahed. Weighted log-rank statistic to compare shared-path adaptive treatment strategies. *Biostatistics*, page kxs042, 2012.

[22] Amy M Kilbourne, Kristen M Abraham, David E Goodrich, Nicholas W Bowersox, Daniel Almirall, Zongshan Lai, and Kristina M Nord. Cluster randomized adaptive implementation trial comparing a standard versus enhanced implementation intervention to improve uptake of an effective re-engagement program for patients with serious mental illness. *Implementation Science*, 8(1):1–14, 2013.

[23] Eric B Laber, Min Qian, Dan J Lizotte, William E Pelham, and Susan A Murphy. Statistical inference in dynamic treatment regimes. *arXiv preprint arXiv:1006.5831*, 2010.

[24] P. W. Lavori and R. Dawson. Dynamic treatment regimes: practical design considerations. *Clinical trials*, 1(1):9–20, 2004.

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

[26] Philip W Lavori and Ree Dawson. Adaptive treatment strategies in chronic disease. *Annual Review of Medicine*, 59:443, 2008.

[27] Philip W Lavori, Ree Dawson, and A John Rush. Flexible treatment strategies in chronic disease: clinical and research implications. *Biological Psychiatry*, 48(6):605–614, 2000.

[28] H Lei, I Nahum-Shani, K Lynch, D Oslin, and S. A. Murphy. A "smart" design for building individualized treatment sequences. *Annual Review of Clinical Psychology*, 8:21–48, 2012.

[29] Z. Li and S.A. Murphy. Sampe size formulae for two-stage randomized trials with survival outcomes. *Biometrika*, 98, 2011.

[30] Kung-Yee Liang and Scott L Zeger. Longitudinal data analysis of continuous and discrete responses for pre-post designs. *Sankhyā: The Indian Journal of Statistics, Series B*, pages 134–148, 2000.

[31] Guanghan F Liu, Kaifeng Lu, Robin Mogg, Madhuja Mallick, and Devan V Mehrotra. Should baseline be a covariate or dependent variable in analyses of change from baseline in clinical trials? *Statistics in Medicine*, 28(20):2509–2530, 2009.

[32] W Lu, H H Zhang, and D Zeng. Variable selection for optimal treatment decision. *Statistical Methods in Medical Research*, 22:493–504, 2013.

[33] Jared K. Lunceford, Marie Davidian, and Anastasios A. Tsiatis. Estimation of survival distributions of treatment policies in two-stage randomization designs in clinical trials. *Biometrics*, 58(1):48–57, 2002.

[34] Lloyd A Mancl and Timothy A DeRouen. A covariance estimator for gee with improved small-sample properties. *Biometrics*, 57(1):126–134, 2001.

[35] J. R. McKay. *Treating Substance Use Disorders With Adaptive Continuing Care*. American Psychological Association, 2009.

[36] Sachiko Miyahara and Abdus S Wahed. Assessing the effect of treatment regimes on longitudinal outcome data: Application to revamp study of depression. *Journal of Statistical Research*, 46(2):233–254, 2013.

[37] E.E.M. Moodie, B. Chakraborty, and M.S. Kramer. Q-learning for estimating optimal dynamic treatment rules from observational data. *Canadian Journal of Statistics*, 40:629 – 645, 2012.

[38] E.E.M. Moodie and T.S. Richardson. Estimating optimal dynamic regimes: Correcting bias under the null. *Scandinavian Journal of Statistics*, 37:126 – 146, 2010.

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

[40] S. A. Murphy. An experimental design for the development of adaptive treatment strategies. *Statistics in Medicine*, 24(10):1455–1481, 2005.

[41] S. A. Murphy, M. J. van der Laan, and J. M. Robins. Marginal mean models for dynamic regimes. *Journal of the American Statistical Association*, 96(456):1410–1423, 2001.

[42] S.A. Murphy. A generalization error for Q-learning. *Journal of Machine Learning Research*, 6:1073 – 1097, 2005.

[43] I. Nahum-Shani, M. Qian, D. Almirall, W. E. Pelham, B. Gnagy, G. A. Fabiano, J. G. Waxmonsky, J. Yu, and S. A. Murphy. Experimental design and primary data analysis methods for comparing adaptive interventions. *Psychological methods*, 17(4):457, 2012.

[44] Inbal Nahum-Shani, Min Qian, Daniel Almirall, William E Pelham, Beth Gnagy, Gregory A Fabiano, James G Waxmonsky, Jihnhee Yu, and Susan A Murphy. Q-learning: A data analysis method for constructing adaptive interventions. *Psychological methods*, 17(4):478, 2012.

[45] Jerzy Neyman, Karolina Iwaszkiewicz, and S. Kolodziejczyk. Statistical problems in agricultural experimentation. *Supplement to the Journal of the Royal Statistical Society*, pages 107–180, 1935.

[46] 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*, 6(2):1–47, 2010.

[47] L Orellana, A Rotnitzky, and JM Robins. Generalized marginal structural models for estimating optimal treatment regimes. Technical report, Technical Report, Department of Biostatistics, Harvard School of Public Health, 2006.

[48] D. Oslin. Managing alcoholism in people who do not respond to naltrexone (ExTENd). *National Institutes of Health, Bethesda, MD*, 2005.

[49] S Pliszka and AACAP Work Group on Quality Issues. Practice parameter for the assessment and treatment of children and adolescents with attention-deficit/hyperactivity disorder. *J. Am. Acad. Child Adolesc. Psychiatry*, 46(7), 2007.

[50] J. Robins, L. Orellana, and A. Rotnitzky. Estimation and extrapolation of optimal treatment and testing strategies. *Statistics in medicine*, 27(23):4678–4721, 2008.

[51] J. M. 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*, 7(9-12):1393–1512, 1986. Mathematical models in medicine: diseases and epidemics, Part 2.

[52] J. M. Robins. Correcting for non-compliance in randomized trials using structural nested mean models. *Communications in Statistics-Theory and methods*, 23(8):2379–2412, 1994.

[53] J. M. Robins. Causal inference from complex longitudinal data. *Latent variable modeling and applications to causality*, pages 69–117, 1997.

[54] J. M. Robins. Marginal structural models. *Proceedings of the American Statistical Association. Section on Bayesian Statistics*, pages 1–10, 1997.

[55] J. M. Robins. Marginal structural models versus structural nested models as tools for causal inference. *Statistical Models in Epidemiology, the Environment and Clinical Trials*, 116:95, 1999.

[56] James M Robins. The analysis of randomized and non-randomized aids treatment trials using a new approach to causal inference in longitudinal studies. *Health service research methodology: a focus on AIDS*, 113:159, 1989.

[57] James M Robins. Information recovery and bias adjustment in proportional hazards regression analysis of randomized trials using surrogate markers. In *Proceedings of the Biopharmaceutical Section, American Statistical Association*, volume 24, page 3. American Statistical Association, 1993.

[58] James M Robins. Causal inference from complex longitudinal data. In *Latent variable modeling and applications to causality*, pages 69–117. Springer, 1997.

[59] James M Robins. Marginal structural models. *1997 Proc. Am. Stat. Assoc. Sect. Bayesian Stat. Sci.*, pages 1–10, 1998.

[60] James M Robins. Marginal structural models versus structural nested models as tools for causal inference. In *Statistical models in epidemiology, the environment, and clinical trials*, pages 95–133. Springer, 2000.

[61] James M Robins, Sander Greenland, and Fu-Chang Hu. Estimation of the causal effect of a time-varying exposure on the marginal mean of a repeated binary outcome. *Journal of the American Statistical Association*, 94(447):687–700, 1999.

[62] James M Robins, Andrea Rotnitzky, and Lue Ping Zhao. Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of the American Statistical Association*, 90(429):106–121, 1995.

[63] Jamie M. Robins. Association, causation, and marginal structural models. *Synthese*, 121:151 – 179, 1999.

[64] J.M. Robins. Optimal structural nested models for optimal sequential decisions. In D.Y. Lin and P. Heagerty, editors, *Proceedings of the Second Seattle Symposium on Biostatistics*, pages 189 – 326, New York, 2004. Springer.

[65] D. B. Rubin. Inference and missing data. *Biometrika*, 63(3):581–592, 1976.

[66] D. B. Rubin. Comment. *Journal of the American Statistical Association*, 81(396):961–962, 1986.

[67] Donald B Rubin. *Multiple imputation for nonresponse in surveys*, volume 307. John Wiley & Sons, 2009.

[68] Donald B Rubin et al. Bayesian inference for causal effects: The role of randomization. *The Annals of Statistics*, 6(1):34–58, 1978.

[69] P J Schulte, A A Tsiatis, E B Laber, and M Davidian. Q- and A-learning methods for estimating optimal dynamic treatment regimes. *arXiv*, page 1202.4177v1, 2012.

[70] Susan M Shortreed, Eric Laber, T Scott Stroup, Joelle Pineau, and Susan A Murphy. A multiple imputation strategy for sequential multiple assignment randomized trials. *Statistics in medicine*, 2014.

[71] R. Song, W. Wang, D. Zeng, and M.R. Kosorok. Penalized Q-learning for dynamic treatment regimes. *Submitted*, 2011.

[72] R.S. Sutton and A.G. Barto. *Reinforcement Learning: An introduction*. MIT Press, Cambridge, 1998.

[73] Helen Tager-Flusberg and Connie Kasari. Minimally verbal school-aged children with autism spectrum disorder: The neglected end of the spectrum. *Autism Research*, 6(6):468–478, 2013.

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

[75] Peter F Thall, Hsi-Guang Sung, and Elihu H Estey. Selecting therapeutic strategies based on efficacy and death in multicourse clinical trials. *Journal of the American Statistical Association*, 97(457):29–39, 2002.

[76] Stef van Buuren and Karin Groothuis-Oudshoorn. mice: Multivariate imputation by chained equations in r. *Journal of Statistical Software*, 45(3):1–67, 2011.

[77] M J Van der Laan. Targeted maximum likelihood based causal inference: Part I. *The International Journal of Biostatistics*, 6, 2010.

[78] MJ van der Laan. Causal effect models for intention to treat and realistic individualized treatment rules.(uc berkeley division of biostatistics working paper series, working paper 203). berkeley, ca: Division of biostatistics, school of public health. *University of California, Berkeley*, 2006.

[79] Aad W Van der Vaart. *Asymptotic statistics*, volume 3. Cambridge university press, 2000.

[80] Stijn Vansteelandt. On confounding, prediction and efficiency in the analysis of longitudinal and cross-sectional clustered data. *Scandinavian Journal of Statistics*, 34(3):478–498, 2007.

[81] Stijn Vansteelandt and Els Goetghebeur. Causal inference with generalized structural mean models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(4):817–835, 2003.

[82] Abdus S. Wahed and Anastasios A. Tsiatis. Optimal estimator for the survival distribution and related quantities for treatment policies in two-stage randomization designs in clinical trials. *Biometrics*, 60(1):124–133, 2004.

[83] Abdus S. Wahed and Anastasios A. Tsiatis. Semiparametric efficient estimation of survival distributions in two-stage randomisation designs in clinical trials with censored data. *Biometrika*, 93(1):163–177, 2006.

[84] Elizabeth J Williamson, Andrew Forbes, and Ian R White. Variance reduction in randomised trials by inverse probability weighting using the propensity score. *Statistics in medicine*, 33(5):721–737, 2014.

[85] T. Zajonc. Bayesian inference for dynamic treatment regimes: Mobility, equity, and efficiency in student tracking. *Journal of the American Statistical Association*, 107:80 – 92, 2012.

[86] Scott L Zeger and Kung-Yee Liang. Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*, pages 121–130, 1986.

[87] Baqun Zhang, Anastasios A Tsiatis, Eric B Laber, and Marie Davidian. Robust estimation of optimal dynamic treatment regimes for sequential treatment decisions. *Biometrika*, 100(3):681–694, 2013.

[88] Na Zhang. Variable selection for optimal treatment regimes, 2014.

[89] Y. Zhao, D. Zeng, A.J. Rush, and M.R. Kosorok. Estimating individual treatment rules using outcome weighted learning. *Journal of the American Statistical Association*, 107:1106 – 1118, 2012.

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