UC Berkeley
UC Berkeley Electronic Theses and Dissertations

**Title**

Targeted Learning for Adaptive Experimentation and Data-Driven Decision-Making

**Permalink**

https://escholarship.org/uc/item/0fv4p62z

**ISBN**

9798173787033

**Author**

Zhang, Wenxin

**Publication Date**

2026-08-10

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

# Targeted Learning for Adaptive Experimentation and Data-Driven Decision-Making

By

Wenxin Zhang

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Biostatistics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Mark van der Laan, Chair
Professor Alan Hubbard
Professor Peng Ding

Summer 2026

Targeted Learning for Adaptive Experimentation and Data-Driven Decision-Making

Copyright © 2026
By
Wenxin Zhang

Abstract

Targeted Learning for Adaptive Experimentation and Data-Driven Decision-Making

By

Wenxin Zhang

Doctor of Philosophy in Biostatistics

University of California, Berkeley

Professor Mark van der Laan, Chair

Human decision-making often unfolds as a feedback loop: actions generate observations, observations refine insight, and insight guides the next action. As data collection and decision-making have become more continuous and dynamic, this loop has increasingly been operationalized through sequential adaptive experiments that continuously update treatment allocation or other aspects of the experimental design as data accumulate. These designs can improve participant outcomes and enhance the efficiency of statistical estimation, with applications ranging from adaptive clinical trials that update randomization probabilities based on interim analyses, to digital health platforms that personalize recommendations from streaming data, to online experiments that adapt product features based on continuously accumulating user responses. However, their adaptive nature creates complex statistical challenges in learning from accumulating data, adapting experimental strategies to different objectives, and conducting valid statistical inference under adaptive data collection.

This dissertation brings together three lines of my doctoral work that develop general frameworks for formulating research questions and addressing statistical challenges in adaptive experimental design and estimation, building upon the literature on targeted maximum likelihood estimation (TMLE) for adaptive designs. Chapter 1 introduces a general TMLE-based Online Super Learner framework for the real-time evaluation and selection among multiple candidate adaptive designs under a benefit-driven objective, with an application to evaluating and utilizing surrogate outcomes in sequential adaptive experiments. Chapter 2 establishes a semiparametric efficient TMLE framework for estimating a broad class of target estimands under diverse adaptive designs, with additional implications for constructing efficiency-oriented designs. Chapter 3 further develops a general design-and-estimation framework for longitudinal adaptive experiments with time-varying treatments to optimize estimation efficiency while enabling valid and efficient estimation and inference. Together, these complementary frameworks provide a coherent methodological foundation for advancing adaptive experimentation and data-driven decision-making in real-world settings.

# Contents

**Contents** **i**

**1 An Online Meta-Level Adaptive Design Framework with Targeted Learning Inference: Applications to Evaluating and Utilizing Surrogate Outcomes in Adaptive Designs** **1**
1.1 Introduction . . . 1
1.2 TMLE-Based Online Super Learner Adaptive Design . . . 5
1.3 Evaluating and Utilizing Surrogates in CARA Designs . . . 10
1.4 Theoretical Analysis of TMLE . . . 13
1.5 Simulation Studies . . . 15
1.6 Application . . . 20
1.7 Conclusions . . . 21
1.8 Supplementary Material . . . 22

**2 Efficient Statistical Estimation for Sequential Adaptive Experiments with Implications for Adaptive Designs** **48**
2.1 Introduction . . . 48
2.2 Data, Model and Statistical Setup . . . 50
2.3 Targeted Maximum Likelihood Estimation . . . 51
2.4 Asymptotic Normality . . . 52
2.5 Semiparametric Efficiency . . . 54
2.6 Adaptive Designs Motivated by Efficiency Results . . . 56
2.7 Simulation Studies . . . 58
2.8 Application and Generalization . . . 62
2.9 Conclusions . . . 63
2.10 Supplementary Material . . . 64

**3 Longitudinal Adaptive Experimental Design for Learning Multiple Target Estimands with Semiparametric Efficient Inference** **67**
3.1 Introduction . . . 67
3.2 Statistical Setup . . . 69
3.3 A General Framework for Efficiency-Oriented Longitudinal Adaptive Design 70

3.4 Adaptive-Design-Likelihood-Based Longitudinal TMLE . . . . . 74
3.5 Application to Multiple Time-to-Treatment-Initiation Effects . . . . . 77
3.6 Simulation Studies . . . . . 79
3.7 Conclusions . . . . . 81
3.8 Supplementary Material . . . . . 82

**Bibliography** **94**

# Acknowledgments

First and foremost, I would like to express my deepest gratitude to my advisor, Professor Mark van der Laan. I came to Berkeley with a curiosity about causal inference, and Mark was the first to teach me to see the field through the lens of targeted learning. At first, I often found his way of thinking quite abstract and his language difficult to follow. Over time, however, I came to appreciate that this abstraction reflected an unusual sharpness and clarity—a way of cutting through complexity to the essential structure of a problem.

One of the most rewarding parts of my PhD has been the gradual process of learning to keep pace with that way of thinking. I still remember approaching him cautiously with questions in my early years, and he always responded with generosity and patience. As I began to propose research questions of my own, he continually encouraged me to think independently while never compromising on intellectual precision and rigor. In moments when I found myself uncertain or caught between possibilities, his steady and optimistic way of looking at things often helped me find the clarity and confidence to move forward.

Through all these years, Mark has been there, quietly and steadily. His thinking moves in long, decisive strides—quick, direct, and incisive, much like the way he carries himself—yet there has always been patience in the distance between his stride and mine. He let the questions retain their edge, while giving me the time and space to grow into them and find my own footing. Sometimes his guidance came through a long discussion in his office; at other times, through a few simple words over coffee that stayed with me for days. There were many moments when I could not see where the road was leading, yet somehow we kept walking. Somewhere along the way, our conversations found a shared rhythm—one in which we could think together, carry ideas forward, and shape increasingly far-reaching, intellectually engaging, and practically meaningful questions. As that rhythm deepened, the path itself began to take shape beneath our feet—question after question, conversation after conversation. Looking back, I feel extraordinarily fortunate to have grown over the course of this long journey. Whatever paths lie ahead, I will carry with me the clarity, freedom of thought, and intellectual courage I have learned from him.

I am also deeply grateful to Professor Maya Petersen for her thoughtful mentorship and steady support in my research and personal growth. Her optimism, encouragement, and genuine care for students have meant a great deal to me. I have especially valued her openness to my ideas, her patience in listening closely, and her ability to grasp the nuances of what I was trying to convey, even when the ideas were complex or still taking shape. These moments felt like looking out over the bay from her office at Berkeley Way West, as light broke through the fog and the red of the Golden Gate Bridge came into view. I have learned enormously from her scientific insight, the generous and constructive way she engages with students and collaborators, and the clarity with which she brings together statistical theory, clinical context, and practical considerations in addressing complex questions. Her integrity, rigor, and professionalism have set an enduring example that I hope to carry with me throughout the years ahead.

In addition, I sincerely thank my committee members, Professors Alan Hubbard and Peng Ding, for serving on my dissertation committee and for their time and support throughout my doctoral training. I am also grateful to the faculty members, staff, and affiliates of the Center for Targeted Machine Learning and Causal Inference (CTML), including Alejandro Schuler, Laura Balzer, Romain Pirracchio, Stathis Gennatas, Christina Da Silva, Jessica Angell, Marie Charpignon, Zeyi Wang, Andrew Mertens, and Zach Butzin-Dozier, whose generosity, encouragement, and collegiality enriched my time at Berkeley. I also cherish the friendships and intellectual exchanges with CTML alumni, especially Rachael Phillips, Ivana Malenica, Nima Hejazi, Philippe Boileau, David McCoy, Lauren Liao, Wilson Cai, Wenjing Zheng, Aurelien Bibaut, Yue You, Jeremy Coyle, Jonathan Levy, among others.

The broader academic community in the Division of Biostatistics has been instrumental to my growth. I greatly appreciate the faculty, especially Professors Jingshen Wang, Xiudi Li, Lexin Li, and John Marshall, as well as the staff, including Jordan Elmowitz, Lauren Krupa, and Sumaiya Elahi, for their care and support throughout my PhD.

I am particularly grateful to the faculty in the Department of Statistics, where my journey at Berkeley began and continued to unfold—from early coursework in causal inference and statistical modeling to later research and intellectual exchanges in Evans Hall. I am especially thankful to Professors Peng Ding, Haiyan Huang, Elizabeth Purdom, Bin Yu, Christopher Paciorek, Sam Pimentel, Avi Feller, and Cari Kaufman for their guidance and encouragement. I also greatly appreciate the support of the department staff, particularly La Shana Porlaris, and of my GSIs, especially Yun Zhou. Their warmth and generous assistance helped me navigate my early years at Berkeley.

I am sincerely thankful to the friends who brought joy and companionship to my life throughout my PhD journey, including Yunzhe Zhou, Lei Shi, Qiuran Lyu, Hao Wang, Zhongming Xie, Haodong Li, Kaiwen Hou, Xinzhu Quan, Sizhu Lu, Shuyi Yang, Yunwen Ji, Kevin Benac, Katarzyna Reluga, Waverly Wei, Feng Ji, and many others. I am also grateful for my peers from the master's program—Yifan Zheng, Kehui Zhao, Baishan Guo, and Zishan Cheng, to name a few—with whom I share many fond memories and cherished friendships.

I also want to express my sincere appreciation to the other mentors, collaborators, and colleagues, including Aaron Hudson, Ahmed Alaa, Adam Yala, Michael Nercessian, Yue Fang, Gene Pennello, Qian Cao, Nicholas Petrick, Sui Huang, Colin Dillard, Xiaochen Zhang, Tony Liu, Dongliang Lu, Jianli Shao, and Tao Huang, who have supported my work and growth at different stages of my academic and professional journey.

Finally, I am profoundly grateful to my parents, grandparents, and extended family, whose love and support have sustained me throughout every stage of my life. My gratitude also extends to all those who have shaped my path—from the friends and teachers from childhood through college to the many scholars across generations whose lives, work, and intellectual legacies continue to resonate through Evans Hall and Berkeley Way West. Some I have not seen in many years; others I have never had the chance to meet. Yet each, in a different way, has helped make this journey possible.

# Chapter 1

# An Online Meta-Level Adaptive Design Framework with Targeted Learning Inference: Applications to Evaluating and Utilizing Surrogate Outcomes in Adaptive Designs

## 1.1 Introduction

### 1.1.1 Overview

Adaptive experimental designs are gaining popularity in clinical trials and online experiments [72, 54]. Unlike traditional static experimental designs, adaptive designs allow continuous updates to treatment randomization probabilities and/or other experimental features based on accumulated data during the experiment, supporting a range of experimental objectives. A key advantage of adaptive designs is their ability to increase the probability of providing participants with their optimal treatment, particularly when this optimal treatment varies between individuals, thereby improving their outcomes.

However, in practice, experimenters often face the challenge of choosing among multiple candidate adaptive design configurations, each with its own trade-offs. A common and compelling example is an adaptive experiment with a longitudinal data structure, in which new participants are continuously enrolled and existing participants generate multiple short-term outcomes that unfold progressively over different follow-up times, while the primary outcome requires long-term follow-up. In such settings, surrogate outcomes, either the early proxies of the primary outcome or other relevant metrics requiring shorter follow-ups, have the potential to quickly identify more effective personalized treatments, inform timely adaptations of treatment randomization probabilities with sufficient data, and thereby enhance partici-

pants' welfare. However, when multiple candidate surrogates are available, selecting among them involves complex trade-offs. Early surrogates provide more abundant and timely data for rapid adaptation, but may be less accurate in detecting treatment effects relevant to the primary outcome. Later surrogates may offer more accurate guidance for optimal treatment, but their delayed availability and limited sample size reduce their impact on early adaptation and participant benefit. In addition, the sensitivity with which a candidate surrogate captures heterogeneous treatment effects is a key driver of its utility: a surrogate that provides stronger signals for distinguishing which individuals benefit from which treatments enables more timely and targeted adaptation toward personalized treatment. Despite all these important factors, in practice, only one candidate adaptation can be implemented at each time point of the experiment, which inevitably leaves the gains and losses of unchosen designs unobserved and unquantified, and further complicates the challenge of navigating these trade-offs in a real-time experiment, underscoring the need for a rigorous statistical framework to overcome these difficulties.

To address these challenges, we propose an online meta-level adaptive design framework with targeted learning inference: a general and data-driven adaptive design system that continuously evaluates and dynamically updates its choice among different candidate designs in real time during ongoing adaptive experiments. We refer to this unified framework as TMLE-OSLAD, a Targeted Maximum Likelihood Estimation-based Online Super Learner Adaptive Design, which itself constitutes a meta-level adaptive design. This framework comprises three key components, each making a distinct methodological contribution.

First, the proposed framework introduces a novel counterfactual perspective for adaptive experiments by asking: How might participants' outcomes have differed under alternative adaptive designs? To formalize this question, we define a new class of target estimands: the mean counterfactual primary outcome that would have been observed had each participant's treatment, beginning at their enrollment, been assigned according to treatment randomization functions specified by the candidate adaptive design.

Second, to enable valid inference from the non-i.i.d. data generated by adaptive experiments, we establish a Targeted Maximum Likelihood Estimation (TMLE) approach that yields a consistent and asymptotically normal estimator without parametric assumptions, building upon martingale central limit theorems and equicontinuity results via maximal inequalities for martingale processes. These developments enable robust inference on the proposed causal estimand—the expected outcome under a candidate counterfactual adaptive design. At the end of the experiment, we also provide TMLEs for estimating other causal estimands of interest, such as the average treatment effect (ATE) and the mean outcome under optimal dynamic treatment regimes (rules) [66, 74, 45].

Third, we develop a TMLE-based Online Super Learner Adaptive Design, which continuously updates estimates of the proposed estimand to evaluate and compare a library of candidate adaptive designs in real time, selects the data-driven best-performing design based on TMLE estimation and inference, dynamically adapts treatment randomization probabilities according to the chosen design, and continuously monitors and reevaluates the set of candidate designs throughout the experiment.

Returning to the motivating example of adaptive experiments involving surrogates, TMLE-OSLAD demonstrates an additional methodological advantage: it provides an adaptive design perspective for rigorously quantifying the value of information flow as it unfolds over time and is continually leveraged to guide sequential decision-making. Specifically, TMLE-OSLAD introduces, to the best of our knowledge, for the first time, a comprehensive causal estimand to evaluate the evolving surrogate utility in such dynamic settings with statistical inference. In this context, the target estimand is defined as the mean counterfactual primary outcome under a candidate surrogate-guided adaptive design, where each participant's treatment randomization probability at enrollment is determined by the estimated conditional average treatment effect function of the surrogate outcome and adaptively tilted toward the estimated optimal personalized treatment with the objective of optimizing the surrogate outcome. This metric reflects all critical factors of the utility of a surrogate in accelerating participant benefit through sequential adaptive design, including the timeliness of surrogate availability for enabling early adaptation, its sensitivity to signal heterogeneous treatment effects, and its predictive accuracy for identifying the optimal personalized treatments for the primary outcome. This is distinct from existing methods that focus on evaluating static statistical associations or causal relationships between surrogate and primary outcomes rather than adopting a dynamic, sequential decision-making perspective.

Furthermore, TMLE-OSLAD enables a principled, data-driven implementation of adaptive experiments that facilitates real-time evaluation and selection among different surrogate-guided adaptive designs to improve the primary outcome.

In summary, TMLE-OSLAD provides a comprehensive meta-level adaptive design framework for monitoring and quantifying opportunities and costs across a library of candidate adaptive experimental strategies. These strategies may differ in the quantities that guide adaptation, the statistical methods used to estimate them, or the rules governing design updates. The framework dynamically selects among these strategies as data accumulate, empowering data-driven, statistical-inference-based decision-making in diverse adaptive experimental settings.

### 1.1.2 Related Work

**Adaptive experimental designs.** Adaptive design is an active research area in statistics. In clinical trials, adaptive design methods are useful due to their flexibility and efficiency [16]. These designs enable adaptive modifications to key aspects such as eligibility criteria, randomization probabilities, and other design features based on sequentially accrued data. Among these, Response-Adaptive Randomization (RAR) updates treatment probabilities using treatment and outcome data of previously enrolled subjects to achieve estimation efficiency or enhance participant welfare by reducing assignment to inferior treatments [32, 72]. Covariate-Adjusted Response-Adaptive (CARA) designs extend RAR by incorporating baseline covariates into randomization, enabling more personalized allocation and improved efficiency [77, 32, 94, 53, 76, 1, 15]. One can also guide CARA for multiple objectives,

e.g., by integrating both efficiency and ethics considerations in adaptation procedures [34], and by both balancing prognostic covariates and facilitating adaptation procedures using predictive covariates in CARA designs [96]. The origins of RAR and CARA designs also trace back to [81] and [71] and connect to multi-armed and contextual bandit literature [10, 55, 79]. Our work differs by considering settings where multiple candidate adaptive designs are available to choose from and providing a principled way to navigate their trade-offs and select from these designs by TMLE-OSLAD framework, with applications to CARA designs with surrogate outcomes.

**Surrogate outcomes.** Surrogate outcomes are often considered as a substitute for the primary outcome when the latter requires a long follow-up time or is expensive to measure. In clinical trials, using surrogate outcomes can result in earlier detection of treatment effects, accelerated drug approvals, and reduced costs [11]. Examples include vaccine-induced immune responses as endpoints for HIV protection [26] and predictive short-term metrics to facilitate digital experiments [19, 70]. The use of surrogates when primary outcomes are delayed has also been discussed in the adaptive design literature [72]. Examples include using short-term responses to guide treatment allocation for outpatients with depressive disorder [80] and for participants in stroke trials [67], Bayesian modeling of surrogate–primary relationships to support RAR designs [38], and contextual bandits that handle delayed rewards by running pre-trained Bayesian reward models [63] or imputing delayed outcomes using surrogates constructed from external data under surrogacy assumptions [88]. Relatedly, [22] analyzes RAR designs driven by an informative surrogate and provides consistent, asymptotically normal statistical inference for the primary outcome under exponential-family models. While these approaches implement RAR or CARA procedures under a specific surrogate-guided design, a unified framework is lacking to dynamically re-evaluate and select among multiple surrogates to guide adaptive experiments.

Another related research area is surrogate evaluation. Many statistical frameworks have been developed to evaluate surrogates [69, 21, 75, 18, 58, 26] and reviewed comprehensively [11, 83, 87, 20]. However, most do not account for the potential benefits and risks of using surrogate outcomes in adaptive experiments that aim to learn and respond to treatment effect heterogeneity. Navigating these trade-offs requires both new approaches for quantifying the benefit of candidate surrogates, and novel designs and estimators that make possible online evaluation of these benefits and dynamic surrogate selection in response during the course of a study, which is addressed by our general TMLE-OSLAD framework.

**Inference based on adaptive design data.** The statistical inference procedure in TMLE-OSLAD is related to the problem of how to provide valid inference for adaptively collected data in the adaptive design literature. For example, asymptotic inference under specified parametric models has been established for data generated by CARA designs [94, 98]. Within the Targeted Maximum Likelihood Estimation framework [51, 50, 49], [53] developed TMLE tailored to adaptive designs without relying on specific parametric

assumptions; subsequent work employed TMLE to estimate risk difference and log-relative risk [12] and the mean outcome under optimal dynamic treatment rules [14] in CARA designs. Another line of work develops inference methods for adaptively collected data based on variance-stabilizing weighting techniques introduced by [59] in non-adaptive settings and extended to treatment effect estimation or policy evaluation in multi-armed and contextual bandits [28, 4, 89], as well as to M-estimation of parametric models using adaptively collected data [91]. Our proposed target estimand, the mean outcome under counterfactual adaptive designs, is distinct from these methods and related to causal inference for time series [46, 61, 62]. We develop a TMLE for this estimand, and establish asymptotic normality by martingale theories and equicontinuity results.

The rest of this chapter is organized as follows. Section 1.2 presents the statistical setup and the general TMLE-OSLAD framework. Section 1.3 specializes the framework for CARA designs involving multiple surrogates. Section 1.4 provides theoretical results. Section 1.5 presents simulation studies, followed by a real-data-based simulation in Section 1.6. Section 1.7 concludes.

## 1.2 TMLE-Based Online Super Learner Adaptive Design

This section outlines the data structure, notation, and methods for the proposed TMLE-OSLAD framework. Suppose we have baseline covariates $W \in \mathcal{W}$ and a treatment $A \in \mathcal{A} = \{0, 1\}$. Let $Y_K$ denote the primary outcome, observed $K$ follow-up periods after treatment assignment, with larger values indicating better outcomes. We also consider $K-1$ intermediate surrogate outcomes $Y_k$, each observed at $k = 1, \ldots, K-1$. We assume all outcomes $Y_k$ lie in $[0, 1]$. Let $T$ denote the total duration of enrollment in the experiment. At each time point, the experimenter uses historical data from previously enrolled participants to update the treatment assignment mechanism for new participants. These updates can aim to increase the chance that participants receive optimal personalized treatment by learning heterogeneous treatment effects, which are detailed in Section 1.3 and Section 1.8.4.

### 1.2.1 Statistical Setup

**Sequential data structure.** We use $t = 1, \ldots, T$ to index the sequential enrollment times when participants enter the study and are randomized to one of the treatment arms. Let $E(t)$ denote the number of subjects enrolled at time $t$, and define the cumulative enrollment by $N(t) = \sum_{t'=1}^{t} E(t')$. For each participant $i$, we sequentially observe $O_i = (t_i, W_i, A_i, Y_{i,1}, \ldots, Y_{i,K})$, where $t_i$ is the enrollment time, $W_i$ are baseline covariates observed at $t_i$, and $A_i$ is the treatment assigned at $t_i$. Surrogate outcomes $Y_{i,1}, \ldots, Y_{i,K-1}$ are observed at times $t_i + k$ for $k = 1, \ldots, K-1$. The final outcome of interest, $Y_{i,K}$, is observed at $t_i + K$. To emphasize the timeline when each variable becomes available, define $\tilde{W}_i(t_i) \equiv W_i$, $\tilde{A}_i(t_i) \equiv A_i$, and $\tilde{Y}_i(t_i + k) \equiv Y_{i,k}$ for $k = 1, \ldots, K$. We use these notations

interchangeably. At each time $t$, the experimenter has access to all information accrued up to $t-1$, denoted by $\bar{O}(t-1) := \{\tilde{W}_i(t'), \tilde{A}_i(t'), \tilde{Y}_i(t') : i \in [N(t-1)],\ t' \leq t-1\}$. This represents all covariates, treatments, and outcomes of previous participants that have become observable before time $t$. From $t-1$ to $t$, new data are collected from two sets: $\mathcal{I}_1(t) := \{i : t-K \leq t_i \leq t-1\}$ and $\mathcal{I}_2(t) := \{i : t_i = t\}$. $\mathcal{I}_1(t)$ contains previously enrolled participants whose outcomes are still arriving, and $\mathcal{I}_2(t)$ contains participants newly enrolled at time $t$. The experimenter first observes the outcomes revealed at time $t$ for $\mathcal{I}_1(t)$, then updates the history $\overline{\mathcal{H}}(t) := \bar{O}(t-1) \cup \{\tilde{Y}_i(t) : i \in \mathcal{I}_1(t)\}$, where $\tilde{Y}_i(t)$ is the outcome for participant $i$ that becomes observable at time $t$. Next, the experimenter enrolls $\mathcal{I}_2(t)$, observes their baseline covariates, and assigns treatments.

**Adaptive treatment randomization.** When a new participant $i$ is enrolled at time $t_i$, a treatment $A_i$ can be randomized based on the information available up to that point and the baseline covariates $W_i$. We denote the participant-specific pre-treatment history as $\mathcal{H}(i) := \big(\overline{\mathcal{H}}(t_i), W_i\big)$, where $\overline{\mathcal{H}}(t_i)$ is the observed data of previous participants enrolled before time $t_i$. We assume that treatment randomization depends on a fixed-dimensional summary context of the pre-treatment history, defined as $C_i := f_C\big(\mathcal{H}(i)\big) \in \mathcal{C}$. The treatment randomization probability is generated by a mapping $g_i : \mathcal{C} \to \mathcal{G}$, where $\mathcal{G}$ is a class of probability distributions of treatment assignment, and $g_i(a|C_i)$ is the probability of assigning $a \in \mathcal{A}$ given context $C_i$.

**Likelihood.** Let $Q_W$ denote the distribution of baseline covariates $W$ and $Q_{Y_k}$ denote the distribution of $Y_k$ given $W, A, Y_1, \ldots, Y_{k-1}$ for outcome $Y_k$ $(k = 1, \ldots, K)$. We assume for every $i$, $t_i \perp\!\!\!\perp (W_i, A_i, Y_{i,1}, \ldots, Y_{i,K})$, $W_i \sim Q_W$, and $Y_{i,k} \sim Q_{Y_k}(\cdot|W_i, A_i, Y_{i,1}, \ldots, Y_{i,k-1})$. Specifically, let $q_W$ denote the marginal density of the baseline covariates $W$, and $q_{Y_k}$ denote the conditional density of the $k$-th outcome $Y_k$ given $W, A, Y_1, \ldots, Y_{k-1}$. The likelihood of the observed data of $n$ subjects is given by $p^n(o_1, \ldots, o_n) = \prod_{i=1}^n q_W(w_i)\, g_i(a_i|C_i) \prod_{k=1}^K q_{Y_k}\big(y_{i,k} \mid w_i, a_i, y_{i,1}, \ldots, y_{i,k-1}\big)$.

**Structural Causal Model.** We formalize the data generating process through a Structural Causal Model (SCM) [68] over the experiment, which also underpins estimation and inference in our setting. At each time $t$, we first observe the history data $\overline{\mathcal{H}}(t)$, which includes, for previously enrolled participants with $t_i = t-k$ $(k = 1, \ldots, K)$, the outcomes given by $Y_{i,k} = f_{Y_k}(W_i, A_i, Y_{i,1}, \ldots, Y_{i,k-1}, U_{Y_{i,k}})$. Second, for each new unit $i$ enrolled at time $t$, we observe $t_i = t$; $W_i = f_W(U_{W_i})$; $C_i = f_C\left(\mathcal{H}(i)\right)$; and $A_i = f_A\left(C_i, U_{A_i}\right)$. Here, $U_i = (U_{W_i}, U_{A_i}, U_{Y_{i,1}}, \ldots, U_{Y_{i,K}})$ are exogenous variables independent across participants. The actual adaptive design for $n$ subjects is represented by the sequence of applied treatment-randomization functions, denoted by $\mathbf{g}_{1:n} := \{g_i : i = 1, \ldots, n\}$.

### 1.2.2 Evaluation of Counterfactual CARA Designs

Here we define a target estimand within a modified SCM under a counterfactual adaptive design. Specifically, we replace every $A_i = f_A(C_i, U_{A_i})$ in the original SCM with $A_i^* \sim g_{*,i}(\cdot|C_i)$ for every $i$, which is a treatment randomization function under a counterfactual adaptive design, based on the observed summary measure $C_i$ for each unit $i$ at their entry time. The outcomes then follow $Y_{i,k}^* = f_{Y_k}(W_i, A_i^*, Y_{i,1}^*, \ldots, Y_{i,k-1}^*, U_{Y_{i,k}})$ for $k = 1, \ldots, K$. Let $\bar{\mathcal{C}}(n) := (C_1, \cdots, C_n)$. Under this modified SCM, the proposed estimand is:

$$\Psi_{\mathbf{g}_{*,1:n}(\bar{\mathcal{C}}(n))} = \frac{1}{n}\sum_{i=1}^{n} E_{g_{*,i}(\cdot|C_i)}\left[Y_{i,K}^*\right],$$

the mean counterfactual primary outcome had all participants, at their entry, been treated according to the treatment randomization mechanism $g_{*,i}$ that could have been applied under an alternative adaptive design $\mathbf{g}_{*,1:n} = (g_{*,1}, \ldots, g_{*,n})$.

This target estimand naturally supports the counterfactual evaluation of multiple candidate adaptive designs of interest. Suppose we have $J$ different candidate adaptive designs, where each design is defined by a series of counterfactual treatment randomization functions $\{g_{j,i}^* : i = 1, \ldots, n\}$ for $j \in [J]$, and depends on the $j$-specific design configuration aiming to increase the probability that new participants receive personalized optimal treatment assignments to maximize their outcomes. The corresponding counterfactual estimand for the $j$-th candidate design is defined as: $\Psi_j := \frac{1}{n}\sum_{i=1}^{n} E_{g_{j,i}^*(\cdot|C_i)}\left[Y_{i,K}^*\right]$, reflecting the expected final outcome had the population been treated under that adaptive design. This enables a principled comparison of all candidate designs by comparing the values of $\{\Psi_j\}_{j=1}^J$.

Let $Q_0 := (Q_{0,W}, Q_{0,K}) \in \mathcal{Q}_W \times \mathcal{Q}_K$, where $Q_{0,W}$ is the true distribution of $W$, and $Q_{0,K}$ is the conditional distribution of $Y_K$ given $A$ and $W$, with the associated $\bar{Q}_{0,K}$ denoting the true conditional mean function of the final outcome $Y_K$ given $A$ and $W$. We also use $g_{0,i}$ to denote the actual treatment randomization function applied to each participant $i$ in the experiment, which is known to the experimenter. One can identify the proposed causal estimand based on the observed data under the following assumptions.

**Assumption 1.1** (Sequential Exchangeability)**.** *For any $i \in [n]$, $Y_{i,K}(a) \perp\!\!\!\perp A_i \mid C_i$.*

**Assumption 1.2** (Positivity)**.** *For any $i \in [n]$ and $a \in \mathcal{A}$, $g_{0,i}(a|C_i) > 0$.*

**Theorem 1.1** (Identification)**.** *With assumptions 1.1 and 1.2, one can identify $\Psi_j$ under $Q_0$ by*

$$\Psi_j(Q_0) = \Psi_{\mathbf{g}_{j,1:n}^*}(Q_0) = \frac{1}{n}\sum_{i=1}^{n} E_{Q_0, g_{j,i}^*}\left[Y_{i,K}^*\right] = \frac{1}{n}\sum_{i=1}^{n}\sum_{a=0}^{1} \bar{Q}_{0,K}(a, W_i) g_{j,i}^*(a|C_i).$$

We note that this estimand provides a fast and robust approximation to an ideal estimand: the expected outcome obtained by integrating the entire distribution of data that would sequentially arise from the candidate adaptive design of interest. However, the latter's

estimation is considerably more complex and computationally intensive. In contrast, our proposed context-specific estimand substantially simplifies the estimation procedure, enables doubly-robust estimation, and closely approximates the ideal estimand when the observed contexts are reflective of those that would be observed in the candidate design, especially when the adaptation procedure is based on learning from a stationary target (e.g. $Q_0$) that dominates over the experiments. This yields an efficient and effective approach to distinguishing high-performing from low-performing designs and facilitating online selection among candidate designs. We refer readers to Section 1.8.1 for further discussions and empirical results showing minimal difference between the two estimands in our simulations.

### 1.2.3 Targeted Maximum Likelihood Estimation

At each time point $t$, let $n_t := N(t-K)$ denote the total number of participants for whom the final outcome $Y_K$ has been observed at that time. We define the target estimand for the $j$-th counterfactual adaptive design at time $t$ as $\Psi_{t,j} := \frac{1}{n_t}\sum_{i=1}^{n_t} E_{Q_0,g^*_{j,i}}\left[Y^*_{i,K}\right] = \frac{1}{n_t}\sum_{i=1}^{n_t}\sum_{a=0}^{1}\bar{Q}_{0,K}(a,W_i)g^*_{j,i}(a|C_i)$. To estimate $\Psi_{t,j}$, we provide a Targeted Maximum Likelihood Estimation (TMLE) approach. Let $\bar{Q}_{n,K,t}$ be an initial estimate of $\bar{Q}_{0,K}$ based on the historical data $\overline{\mathcal{H}}(t)$ observed up to time $t$. We apply a TMLE procedure to update $\bar{Q}_{n,K,t}$ by fitting the logistic model [51, 53, 50, 62]: $\operatorname{logit}\bar{Q}_{n,K,t,\epsilon}(A,W) = \operatorname{logit}\bar{Q}_{n,K,t}(A,W) + \epsilon$, with weights $\frac{g^*_{j,i}(A_i|C_i)}{g_{0,i}(A_i|C_i)}$. The estimate of $\epsilon$ is denoted by $\hat{\epsilon}_t$, and an updated estimator of the conditional mean function $\bar{Q}^*_{n,K,t} := \bar{Q}_{n,K,t,\hat{\epsilon}_t}$ is obtained, which solves:

$$\frac{1}{n_t}\sum_{i=1}^{n_t}\frac{g^*_{j,i}(A_i|C_i)}{g_{0,i}(A_i|C_i)}\left(Y_{i,K} - \bar{Q}^*_{n,K,t}(A_i,W_i)\right) = 0.$$

The TMLE estimate of $\Psi_{t,j}$ is:

$$\hat{\Psi}_{t,j} := \frac{1}{n_t}\sum_{i=1}^{n_t}\sum_{a=0}^{1}\bar{Q}^*_{n,K,t}(a,W_i)g^*_{j,i}(a|C_i).$$

The estimated standard error is $\sqrt{\hat{\sigma}^2_{t,j}/n_t}$, where

$$\hat{\sigma}^2_{t,j} := \frac{1}{n_t}\sum_{i=1}^{n_t}\left[\frac{g^*_{j,i}(A_i|C_i)}{g_{0,i}(A_i|C_i)}\left(Y_{i,K} - \bar{Q}^*_{n,K,t}(A_i,W_i)\right)\right]^2.$$

The $100(1-\alpha)\%$ Wald confidence interval is $\left[\hat{\Psi}_{t,j} - z_{1-\alpha/2}\sqrt{\frac{1}{n_t}\hat{\sigma}^2_{t,j}}, \hat{\Psi}_{t,j} + z_{1-\alpha/2}\sqrt{\frac{1}{n_t}\hat{\sigma}^2_{t,j}}\right]$, where $z_{1-\alpha/2}$ is the $1-\alpha/2$ quantile of the standard normal distribution (e.g. $\alpha = 0.05$).

Under the conditions in Section 1.4, TMLE is consistent, asymptotically normal and doubly robust, meaning that consistency holds if the outcome regression or the treatment assignment mechanism is correctly specified. In adaptive designs, the treatment assignment

mechanism is known, so TMLE remains consistent even if the outcome regression is misspecified. That said, more accurate estimation of $\bar{Q}_{0,K}$ can reduce the variance of TMLE. When multiple estimators are available, a Super Learner [47] can be used to select the best-performing estimator by cross-validation.

### 1.2.4 Online Super Learner Adaptive Design

As the experiment progresses to a new time point $t$, there are up to $J$ candidate treatment randomization functions $g^*_{1,i}, \ldots, g^*_{J,i}$ for assigning treatment randomization probabilities for a new participant $i$ enrolled at $t$. For each $j \in [J]$, $g^*_{j,i}$ is generated based on historical data using the $j$-specific adaptive design configurations. The proposed Online Super Learner Adaptive Design begins by estimating $\Psi_{t,j}$ for each $j \in [J]$ and obtaining confidence intervals for these estimates. Then, it selects the design indexed by $j^*_t := \text{argmax}_{j\in[J]}\, \hat{\Psi}_{t,j} - z_{1-\alpha/2}\sqrt{\frac{1}{n_t}\hat{\sigma}^2_{t,j}}$, the candidate design with the highest lower confidence interval bound among all $\hat{\Psi}_{t,j}$ for $j \in [J]$, prioritizing designs with high performance with low uncertainty to mitigate the risk of selecting a suboptimal design due to unstable estimates. Finally, for all participants $i$ enrolled at time $t$, we assign treatment randomization probabilities using $g^*_{j^*_t,i}$.

### 1.2.5 Statistical Inference at the End of Adaptive Experiments

We denote by $T_{end} := T + K$ the time when the experiment concludes, i.e., when all final outcomes $Y_K$ have been observed after the last enrollment at time $T$. The corresponding sample size is $n_{end} := N(T) = \sum_{t=1}^{T} E(t)$. At time $T_{end}$, the TMLE procedure in Section 1.2.3 applies to estimate $\Psi_{T_{end},j}$, the expected final outcome under the $j$th counterfactual adaptive design. We also provide TMLEs for other causal estimands commonly studied in randomized trials, including the average treatment effect and the mean outcome under dynamic treatment rules. Their statistical inference is based on Wald confidence intervals using the corresponding TMLE estimates and their estimated standard errors.

**Average treatment effect.** We define the following average treatment effect (ATE) on the primary outcome $Y_K$ across all participants within the experiment:

$$\psi^{\text{ATE}}_{T_{end}} := \frac{1}{n_{end}} \sum_{i=1}^{n_{end}} \left[\bar{Q}_{0,K}(1, W_i) - \bar{Q}_{0,K}(0, W_i)\right].$$

For estimation, we first obtain an initial estimate $\bar{Q}_{n,K,T_{end}}$ of $\bar{Q}_{0,K}$ and then perform the TMLE update with weights $\frac{2A_i-1}{g_{0,i}(A_i|C_i)}$, yielding $\bar{Q}^*_{n,K,T_{end}}$ that solves

$$\frac{1}{n_{end}} \sum_{i=1}^{n_{end}} \frac{2A_i - 1}{g_{0,i}(A_i|C_i)} \left(Y_{i,K} - \bar{Q}^*_{n,K,T_{end}}(A_i, W_i)\right) = 0.$$

The TMLE estimate of $\psi^{\text{ATE}}_{T_{end}}$ is $\hat{\psi}^{\text{ATE}}_{T_{end}} := \frac{1}{n_{end}} \sum_{i=1}^{n_{end}} \left[\bar{Q}^*_{n,K,T_{end}}(1, W_i) - \bar{Q}^*_{n,K,T_{end}}(0, W_i)\right]$. The standard error estimate is $\sqrt{\hat{\sigma}^{2,\text{ATE}}_{T_{end}}/n_{end}}$, where

$$\hat{\sigma}^{2,\text{ATE}}_{T_{end}} = \frac{1}{n_{end}} \sum_{i=1}^{n_{end}} \left[\frac{2A_i - 1}{g_{0,i}(A_i \mid C_i)} \left(Y_{i,K} - \bar{Q}^*_{n,K,T_{end}}(A_i, W_i)\right)\right]^2.$$

**Mean outcome under a dynamic treatment rule.** Let $d : \mathcal{W} \to \mathcal{A}$ denote a fixed dynamic treatment rule mapping baseline covariates to a treatment arm. We define a target estimand $\psi^d_{T_{end}} := \frac{1}{n_{end}} \sum_{i=1}^{n_{end}} \bar{Q}_{0,K}\big(d(W_i), W_i\big)$ as the mean final outcome $Y_K$ under rule $d$. We apply TMLE with weights $\frac{I(A_i = d(W_i))}{g_{0,i}(A_i|C_i)}$ to solve

$$\frac{1}{n_{end}} \sum_{i=1}^{n_{end}} \frac{I(A_i = d(W_i))}{g_{0,i}(A_i|C_i)} (Y_{i,K} - \bar{Q}^*_{n,K,T_{end}}(A_i, W_i)) = 0,$$

and yield $\hat{\psi}^d_{T_{end}} := \frac{1}{n_{end}} \sum_{i=1}^{n_{end}} \bar{Q}^*_{n,K,T_{end}}(d(W_i), W_i)$. The standard error estimate is given by $\sqrt{\hat{\sigma}^{2,d}_{T_{end}}/n_{end}}$, where

$$\hat{\sigma}^{2,d}_{T_{end}} := \frac{1}{n_{end}} \sum_{i=1}^{n_{end}} \left[\frac{I(A_i = d(W_i))}{g_{0,i}(A_i|C_i)} (Y_{i,K} - \bar{Q}^*_{n,K,T_{end}}(A_i, W_i))\right]^2.$$

We also note their marginal versions $\psi^{\text{ATE}} := E\left[\bar{Q}_{0,K}(1, W) - \bar{Q}_{0,K}(0, W)\right]$ and $\psi^d := E\left[\bar{Q}_{0,K}\big(d(W), W\big)\right]$. The TMLE estimates are the same as above, while the standard error estimates differ, with $\sqrt{\hat{\sigma}^{2,\text{ATE}}/n_{end}}$ for $\psi^{\text{ATE}}$ and $\sqrt{\hat{\sigma}^{2,d}/n_{end}}$ for $\psi^d$, where $\hat{\sigma}^{2,\text{ATE}} := \frac{1}{n_{end}} \sum_{i=1}^{n_{end}} \left[\frac{2A_i - 1}{g_{0,i}(A_i|C_i)} \left(Y_{i,K} - \bar{Q}^*_{n,K,T_{end}}(A_i, W_i)\right) + \bar{Q}^*_{n,K,T_{end}}(1, W_i) - \bar{Q}^*_{n,K,T_{end}}(0, W_i) - \hat{\psi}^{\text{ATE}}\right]^2$ and $\hat{\sigma}^{2,d} := \frac{1}{n_{end}} \sum_{i=1}^{n_{end}} \left[\frac{I(A_i = d(W_i))}{g_{0,i}(A_i|C_i)} (Y_{i,K} - \bar{Q}^*_{n,K,T_{end}}(A_i, W_i)) + \bar{Q}^*_{n,K,T_{end}}(d(W_i), W_i) - \hat{\psi}^d\right]^2$.

## 1.3 Evaluating and Utilizing Surrogates in CARA Designs

In this section, we apply the TMLE-OSLAD framework to the motivating example of adaptive experiments involving multiple surrogate outcomes before the final outcome. Recall that there are $K - 1$ candidate surrogate outcomes, $Y_1, \ldots, Y_{K-1}$ measured before the primary outcome $Y_K$ at follow-up times $k = 1, \ldots, K - 1$. Each surrogate $Y_k$ defines a distinct candidate adaptive design, which leverages accumulating information from short-term surrogate outcomes to increase the probability for assigning the estimated personalized treatment associated with that surrogate over time. TMLE-OSLAD is implemented to continuously evaluate and dynamically select among different surrogate-specific adaptive designs over the

experiment. We first introduce a CARA procedure that adapts treatment randomization probabilities based on the estimated conditional average treatment effect (CATE) to improve participant outcomes. Building on Section 1.2, we define our target estimand that evaluates each candidate surrogate within a counterfactual adaptive design that utilizes the surrogate to adapt treatment randomization probabilities. We then demonstrate TMLE-OSLAD that uses TMLE to estimate these target estimands to evaluate surrogate-guided adaptive designs, and utilizes the most useful one to guide the ongoing adaptive experiment. Finally, we conduct post-experiment inference for other causal estimands such as ATE and the mean primary outcome under optimal dynamic treatment rules that optimize primary or surrogate outcomes.

### 1.3.1 CARA Designs Using Surrogate Outcomes

In our CARA designs, at each time point, we update the model to estimate the CATE for the outcome of interest and its associated uncertainty, and use these estimates to assign randomization probabilities to newly enrolled participants. This design aims to maximize participant outcomes by increasing the chance to allocate the estimated optimal treatment, with a principled exploration–exploitation balance ensuring sufficient exploration when the superiority of a treatment arm is uncertain and maintaining the ability to continue learning heterogeneous treatment effect from data collected at later time points. Specifically, when the estimated CATE is small relative to its uncertainty, indicating that the superiority of one treatment arm is uncertain, the treatment randomization probabilities remain close to 0.5. This avoids overcommitting to a noisy and potentially suboptimal rule and ensures sufficient exploration across treatment arms. As the estimated CATE suggests a stronger signal for the superior treatment arm, the probability of assigning that treatment increases up to a maximum exploitation probability. Meanwhile, a positive lower bound on the probability of assigning the alternative arm is maintained to ensure sufficient support for valid estimation and inference of other estimands at the end of the experiment (see Sections 1.2.5 and 1.3.3).

More details of the CARA design are provided in the Supplementary Material (Section 1.8). Specifically, CATE is estimated using a doubly robust approach based on pseudo-outcomes [43, 60] (Section 1.8.3), with point estimates and confidence intervals constructed based on higher-order Highly Adaptive Lasso (HAL) [52] (Section 1.8.2). Details of the treatment randomization function can be found in Section 1.8.4.

### 1.3.2 Surrogate-Guided Online Super Learner Adaptive Design

To enable data-adaptive selection of the most effective surrogate to guide treatment allocation, we leverage the general TMLE-OSLAD framework in Section 1.2 to both evaluate and dynamically utilize candidate surrogate outcomes within real-time CARA experiments. At each interim time $t$, we use all observed historical data to assess the utility of each surrogate-driven adaptive design and to select the most promising surrogate for adaptation.

First, at time $t$, we define the target estimand $\Psi_{t,\mathbf{g}^*_{k,1:n_t}} := \frac{1}{n_t}\sum_{i=1}^{n_t} E_{g^*_{k,i}}\left[Y^*_{i,K}\right]$ for each candidate surrogate $Y_k$, which represents the mean final outcome that would be achieved for the $n_t$ fully-followed participants if each had been assigned with the treatment randomization function $g^*_{k,i}$, at their enrollment, determined by the adaptive design using surrogate $Y_k$. We refer to this estimand as "the utility of a surrogate" in the CARA design (see Section 1.8.1 for further discussion of this estimand).

To construct $g^*_{k,i}$ from the CATE for $Y_k$, we impose the following identification assumption.

**Assumption 1.3** (Sequential Exchangeability for Surrogate Outcomes)**.** *For any* $k \in \{1,\ldots,K-1\}$ *and any subject* $i \in \{1,\ldots,n_t\}$, $Y_{i,k}(a) \perp\!\!\!\perp A_i \mid C_i$.

Under Assumptions 1.1 and 1.2, $\Psi_{t,\mathbf{g}^*_{k,1:n_t}}$ can be identified as

$$\Psi_{t,k} := \frac{1}{n_t}\sum_{i=1}^{n_t}\sum_{a=0}^{1}\bar{Q}_{0,K}(a,W_i)g^*_{k,i}(a|C_i).$$

Then, we estimate $\Psi_{t,k}$ for each surrogate $Y_k$ using TMLE, which is given by $\hat{\Psi}_{t,k} = \frac{1}{n_t}\sum_{i=1}^{n_t}\sum_{a=0}^{1}\bar{Q}^*_{n,K,t}(a,W_i)g^*_{k,i}(a|C_i)$. The estimated standard error is $\sqrt{\hat{\sigma}^2_{t,k}/n_t}$, where $\hat{\sigma}^2_{t,k} = \frac{1}{n_t}\sum_{i=1}^{n_t}\left[\frac{g^*_{k,i}(A_i|C_i)}{g_{0,i}(A_i|C_i)}\left(Y_{i,K}-\bar{Q}^*_{n,K,t}(A_i,W_i)\right)\right]^2$. Next, we compare all candidate surrogates and the final outcome by constructing confidence intervals for $\Psi_{t,k}$ $(k=1,\ldots,K)$ and identify the outcome $Y_{k^*_t}$ with the highest lower confidence bound, where $k^*_t := \text{argmax}_{k\in[K]}\,\hat{\Psi}_{t,k} - z_{1-\alpha/2}\sqrt{\hat{\sigma}^2_{t,k}/n_t}$. For example, if two surrogates have similar point estimates but one yields a much narrower confidence interval, the surrogate with the higher lower bound is prioritized, as it reflects more evidence in its utility relative to uncertainty. Accordingly, we select $Y_{k^*_t}$ as the outcome to guide the next-stage treatment assignment for newly enrolled participants at time $t$, utilizing the treatment randomization function $g^*_{k^*_t,i}$. This procedure is repeated at each subsequent time point, allowing adaptive surrogate selection and treatment allocation as new data accrue.

### 1.3.3 Inference after Adaptive Experiments Involving Surrogates

At the end of the adaptive experiment, one can follow Section 1.2.5 to conduct statistical inference for causal estimands such as the average treatment effect or the mean outcome under dynamic treatment rules. Here, we consider the mean primary outcome under optimal dynamic treatment rules (ODTR) that maximize the primary outcome [66, 74, 45] or a surrogate outcome [31, 88].

For a primary or surrogate outcome $Y_k$ $(k \in \{1,\ldots,K\})$, define the optimal dynamic treatment rule $d^*_{0,k} : \mathcal{W} \to \mathcal{A}$ by $d^*_{0,k}(W) = I(E[Y_k(1) - Y_k(0) \mid W] > 0)$. We estimate this

rule using all observed data and obtain $d^*_{n,k,T_{end}}$ (for brevity, we also write $d^*_{n,k}$). The target estimand is the mean primary outcome $Y_K$ under this rule

$$\psi^{d^*_{n,k}}_{T_{end}} := \frac{1}{n_{end}} \sum_{i=1}^{n_{end}} \bar{Q}_{0,K}(d^*_{n,k}(W_i), W_i),$$

or its marginal version $\psi^{d^*_{n,k}} := E[\bar{Q}_{0,K}(d^*_{n,k}(W), W)]$. We estimate them by TMLE solving $\frac{1}{n_{end}} \sum_{i=1}^{n_{end}} \frac{I(A_i = d^*_{n,k}(W_i))}{g_{0,i}(A_i|C_i)} (Y_{i,K} - \bar{Q}^*_{n,K,T_{end}}(A_i, W_i)) = 0$, yielding a point estimate $\hat{\psi}^{d^*_{n,k}}_{T_{end}} := \frac{1}{n_{end}} \sum_{i=1}^{n_{end}} \bar{Q}^*_{n,K,T_{end}}(d^*_{n,k}(W_i), W_i)$, with standard error estimates given by Section 1.2.5.

We note that this approach provides an additional treatment-rule-based perspective on evaluating surrogate utility by the mean primary outcome $Y_K$ achieved under ODTRs maximizing a surrogate outcome $Y_k$ ($k \neq K$), which can be compared with the value achieved under the ODTR maximizing $Y_K$ itself. Moreover, for $d^*_{n,K}$, the estimated ODTR maximizing the primary outcome $Y_K$, the standard TMLE for its evaluation may exhibit upward bias as the same data are used to both learn the rule and evaluate its value. We recommend applying cross-validated TMLE (CV-TMLE) [97, 45, 64], which uses sample splitting to avoid finite sample bias.

## 1.4 Theoretical Analysis of TMLE

This section derives consistency and asymptotic normality of the proposed TMLE. For notational continuity with the surrogate-guided adaptive design setup introduced above, we continue to index candidate designs by $k \in [K]$ in the TMLE analysis, and the target estimand of interest is $\Psi_{t,k}$. This indexing is notationally equivalent to the general index $j \in [J]$ used in Section 1.2 for estimating $\Psi_{t,j}$; the theoretical results and proofs apply identically under either notation. We first study the difference between TMLE $\hat{\Psi}_{t,k}$ and the truth $\Psi_{t,k}$. For any $\bar{Q}_K \in \bar{\mathcal{Q}}_K$, define $D_k(\bar{Q}_K)(O_i, C_i) := \frac{g^*_{k,i}(A_i|C_i)}{g_{0,i}(A_i|C_i)} \left(Y_{i,K} - \bar{Q}_K(A_i, W_i)\right)$.

**Theorem 1.2.** *For any $\bar{Q}^*_{n,K,t} \in \bar{\mathcal{Q}}_K$ and $\bar{Q}_{1,K} \in \bar{\mathcal{Q}}_K$, the difference between $\hat{\Psi}_{t,k}$ and $\Psi_{t,k}$ can be decomposed as $\hat{\Psi}_{t,k} - \Psi_{t,k} = M_{1,n_t}(\bar{Q}_{1,K}) + M_{2,n_t}(\bar{Q}^*_{n,K,t}, \bar{Q}_{1,K})$, where*

$$M_{1,n_t}\left(\bar{Q}_{1,K}\right) = \frac{1}{n_t} \sum_{i=1}^{n_t} \left[ D_k\left(\bar{Q}_{1,K}\right)(O_i, C_i) - E\left[ D_k\left(\bar{Q}_{1,K}\right)(O_i, C_i) \,|\mathcal{H}(i)\right] \right],$$

$$M_{2,n_t}\left(\bar{Q}^*_{n,K,t}, \bar{Q}_{1,K}\right) = \frac{1}{n_t} \sum_{i=1}^{n_t} \Bigg\{ \left[ D_k\left(\bar{Q}^*_{n,K,t}\right)(O_i, C_i) - E\left[ D_k\left(\bar{Q}^*_{n,K,t}\right)(O_i, C_i) \,|\mathcal{H}(i)\right] \right] - \left[ D_k\left(\bar{Q}_{1,K}\right)(O_i, C_i) - E\left[ D_k\left(\bar{Q}_{1,K}\right)(O_i, C_i) \,|\mathcal{H}(i)\right] \right] \Bigg\}.$$

The difference of TMLE $\hat{\Psi}_{t,k}$ and the true $\Psi_{t,k}$ is decomposed into $M_{1,n_t}$, an average of a martingale difference sequence, and a martingale process $M_{2,n_t}$. Under the strong

positivity and stabilized conditional variance conditions below, the asymptotic normality of $M_{1,n_t}$ follows directly from the martingale central limit theorem (e.g., Theorem 2 in [9]).

**Assumption 1.4** (Strong positivity). *For any $i$, there exists $\zeta > 0$ such that $g_{0,i} > \zeta$ and $g^*_{k,i} > \zeta$.*

**Assumption 1.5** (Stabilization of the average of conditional variances). *There exists $\sigma^2_{t,k} \in \mathbb{R}^+$ such that $\frac{1}{n_t}\sum_{i=1}^{n_t} Var\left[D_k(\bar{Q}_{1,K})(O_i, C_i) \mid \mathcal{H}(i)\right] \xrightarrow{p} \sigma^2_{t,k}$.*

**Theorem 1.3** (Asymptotic normality of the first term). *Suppose that assumptions 1.4 and 1.5 hold, then $\sqrt{n_t} M_{1,n_t}\left(\bar{Q}_{1,K}\right) \xrightarrow{d} N(0, \sigma^2_{t,k})$.*

We note that $M_{1,n_t}$ depends only on a fixed function $\bar{Q}_{1,K} \in \bar{\mathcal{Q}}_K$. For asymptotic normality of the TMLE, $\bar{Q}_{1,K}$ need not equal the true conditional mean $\bar{Q}_{0,K}$ as long as $\bar{Q}^*_{n,K,t}$ converges to $\bar{Q}_{1,K}$ satisfying the equicontinuity condition for the second term, $M_{2,n_t}$, as formalized below.

**Assumption 1.6** (Entropy integral of $\bar{\mathcal{Q}}_K$). *Let $N_\infty(\epsilon, \bar{\mathcal{Q}}_K)$ denote the $\epsilon$-covering number in supremum norm of $\bar{\mathcal{Q}}_K$. Define $J_\infty(\epsilon, \bar{\mathcal{Q}}_K) := \int_0^\epsilon \sqrt{\log\left(1 + N_\infty(u, \bar{\mathcal{Q}}_K)\right)} du$ as the $\epsilon$-entropy integral in supremum norm of $\bar{\mathcal{Q}}_K$. Assume that $\mathcal{J}_\infty\left(\epsilon, \bar{\mathcal{Q}}_K\right)$ is finite and $\mathcal{J}_\infty\left(\epsilon, \bar{\mathcal{Q}}_K\right) \to 0$ as $\epsilon \to 0$.*

**Assumption 1.7** (Convergence of $\bar{Q}^*_{n,K,t}$). *For any $\bar{Q}_K \in \bar{\mathcal{Q}}_K$, define*

$$s_{n_t}\left(\bar{Q}_K, \bar{Q}_{1,K}\right) := \sqrt{\frac{1}{n_t}\sum_{i=1}^{n_t} E\left[\phi\left(\frac{\left|D_k\left(\bar{Q}_K\right)(O_i, C_i) - D_k\left(\bar{Q}_{1,K}\right)(O_i, C_i)\right|}{\upsilon}\right) \mid \mathcal{H}(i)\right]},$$

*where $\phi(x) := e^x - x - 1$, $\upsilon := 8(1-\zeta)/\zeta$. Assume that $s_{n_t}\left(\bar{Q}^*_{n,K,t}, \bar{Q}_{1,K}\right) = O_P(\delta_{n_t})$ for some $\delta_{n_t} \to 0$.*

**Theorem 1.4** (Equicontinuity). *Under Assumptions 1.4, 1.6, and 1.7, $M_{2,n_t}(\bar{Q}^*_{n,K,t}, \bar{Q}_{1,K}) = o_P\left(n_t^{-\frac{1}{2}}\right)$.*

**Theorem 1.5** (Asymptotic normality of TMLE). *Suppose Assumptions 1.4, 1.5, 1.6, and 1.7 hold, then $\sqrt{n_t}(\hat{\Psi}_{t,k} - \Psi_{t,k}) \xrightarrow{d} N(0, \sigma^2_{t,k})$.*

The asymptotic normality of TMLE is obtained directly from Theorems 1.3 and 1.4. We estimate the standard error by $\hat{\sigma}^2_{t,k} := \frac{1}{n_t}\sum_{i=1}^{n_t}\left\{\frac{g^*_{k,i}(A_i|C_i)}{g_{0,i}(A_i|C_i)}\left[Y_{i,K} - \bar{Q}^*_{n,K,t}(A_i, W_i)\right]\right\}^2$ (which converges in probability to $\sigma^2_{t,k}$), with the confidence interval given by $\hat{\Psi}_{t,k} \pm z_{1-\alpha/2}\sqrt{\hat{\sigma}^2_{t,k}/n_t}$.

## 1.5 Simulation Studies

**Simulation setup.** We present simulation studies to evaluate the performance of our proposed adaptive design and estimation methods. [1] We consider a binary treatment $A \in \{0, 1\}$, a univariate baseline covariate $W$ following a uniform distribution $U(-4, 4)$, four surrogate outcomes $Y_k$ measured at follow-up times $k = 1, 2, 3, 4$, respectively, and a final outcome $Y_5$. Each participant's outcome $Y_k$ is generated by its conditional mean $E(Y_k|A, W)$ plus a random error following the normal distribution ($\sim N(0, 1)$). For detailed information on the simulations, we refer readers to Section 1.8.3 for CATE estimation, Section 1.8.4 for treatment randomization mechanism and Section 1.8.7 for data generating distributions. We evaluate the performance of various designs in two scenarios. In Scenario 1 (Figure 1.1a), the optimal dynamic treatment rule based on the final outcome is more aligned with that defined in terms of later versus earlier surrogates. This implies that, despite requiring some time to measure and thus delaying adaptation in randomization probabilities, later surrogates are more useful in guiding treatment assignment probabilities. In Scenario 2 (Figure 1.1b), all surrogate outcomes have the same ODTR as that of the final outcome. Notably, early surrogates have larger absolute CATE compared to later surrogates, suggesting that early surrogates are more sensitive to different treatments and potentially serve as superior indicators of the most beneficial treatment compared to the final outcome. In other words, in this scenario, earlier surrogates have the joint advantage of both timeliness and improved sensitivity for detecting treatment effect heterogeneity.

We compare seven distinct designs with varying treatment randomization mechanisms over 500 Monte Carlo runs: (a) a simple non-adaptive randomized controlled trial (RCT) design, which allocates two treatments with equal probability at each time point; (b) five adaptive designs with randomization probabilities updated in response to an *a priori* fixed choice of either a surrogate outcome $Y_k$ ($k = 1$, ..., 4) or the final outcome $Y_K$ ($K = 5$); (c) the proposed TMLE-OSLAD adaptive design that dynamically evaluates and selects which outcome (among $Y_1, \ldots, Y_5$) is used to update treatment randomization probabilities at each time point. For each design, we run experiments for $T = 50$ time points with 50 subjects enrolled per time point. The maximal treatment assignment probability is capped at 90%.

**Surrogate evaluation and TMLE performance.** We first evaluate if TMLE-OSLAD can consistently identify the oracle surrogate, which we define as the candidate with the highest true $\Psi_{t,k}$ at time $t$ among all $Y_k$ (in other words, the surrogate that if used would result in the largest expected primary outcomes for all enrolled participants in the experiment). In particular, if $\Psi_{t,K}$ is the largest, then the primary outcome $Y_K$ itself is the oracle. Table 1.1 presents the average of true $\Psi_{t,k}$ values at illustrative time points 11, 21, 31, 41, and 50 for both scenarios across Monte Carlo runs. The expected outcome of $Y_5$ under the non-adaptive randomization is also reported, denoted by $\Psi_{t,RCT}$. In Scenario 1, the oracle surrogate gradually shifts towards later surrogates as time elapses and more data from these

[1] The code can be found at https://github.com/WenxinZhang25/TMLE-OSLAD.

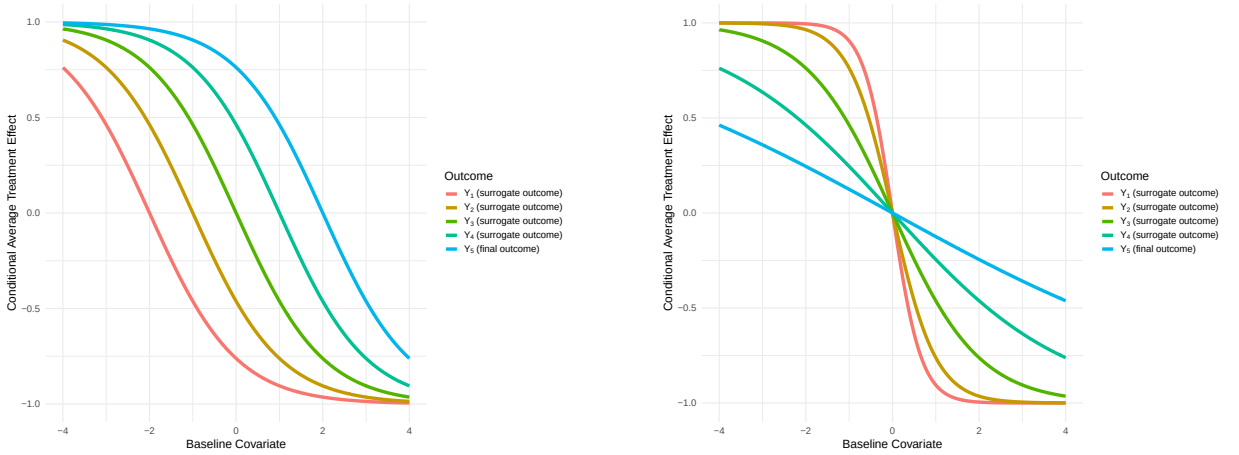



Figure 1.1: Conditional Average Treatment Effect (CATE) functions for the simulation scenarios: (a) Scenario 1, where later surrogates have higher utility; (b) Scenario 2, where earlier surrogates have higher utility. Each curve depicts the CATE function for a candidate outcome, $Y_1$ through $Y_5$.

surrogates become available to guide the adaptation of treatment randomization probabilities. Initially, surrogates $Y_3$ and $Y_4$ have the highest value at times 11 and 21, reflecting the balance they achieve between earlier availability and degree to which they reflect the true optimal individualized treatment assignment. While these surrogates provide benefit in the short term, they do not consistently predict the correct optimal treatment for the primary outcome for all individuals, leading them to be eventually outperformed by designs that utilize the true primary outcome $Y_5$ as additional follow-up time is accrued. This is expected, as the ultimate assessment of a surrogate's effectiveness as follow-up time extends hinges on its ability to predict the optimal personalized treatment for the primary outcome. In Scenario 2, as expected, surrogate $Y_1$ remains the most useful (i.e. is the oracle surrogate) throughout, due to its early availability, and the fact that it correctly identifies the true optimal individualized treatment for the long-term primary outcome. In contrast, later surrogates $Y_4$ and $Y_5$ have lower scores. This disparity can be attributed to the fact that while $Y_4$ and $Y_5$ will eventually serve equally well to identify the optimal personalized treatment as time since trial initiation goes to infinity, their utility is limited under finite trial durations because they have been observed for fewer participants. Across both scenarios, the TMLE estimates for these estimands exhibit low bias and variance, and their confidence intervals obtain nominal coverage (see Tables 1.3, 1.4, 1.5 in Section 1.8.7).

Figures 1.2a–1.2b show the frequency with which each candidate design is selected by TMLE-OSLAD over time in Scenarios 1 and 2, respectively, while Figures 1.2c–1.2d present a single-run example illustrating the selection mechanism. Without prior knowledge of relationships between surrogate and primary outcomes, the TMLE-OSLAD procedure is able to identify the oracle surrogate in each setting. In Scenario 1, where later surrogate out-

Table 1.1: True values of the target estimand $\Psi_{t,k}$ under candidate designs in Scenarios 1 and 2 ($t = 11, 21, 31, 41, 50$), where participants enrolled from $1, \ldots, t-K$ were adaptively treated according to candidate surrogate $Y_k$. Values under the oracle design are shown in bold.

(a) Scenario 1

| $t$ | $\Psi_{t,RCT}$ | $\Psi_{t,1}$ | $\Psi_{t,2}$ | $\Psi_{t,3}$ | $\Psi_{t,4}$ | $\Psi_{t,5}$ |
|---|---|---|---|---|---|---|
| 11 | 0.000 | -0.016 | 0.053 | **0.078** | 0.066 | 0.033 |
| 21 | 0.000 | -0.008 | 0.083 | 0.145 | **0.177** | 0.173 |
| 31 | 0.000 | -0.005 | 0.091 | 0.161 | 0.203 | **0.208** |
| 41 | 0.000 | -0.002 | 0.094 | 0.169 | 0.215 | **0.225** |
| 50 | 0.000 | -0.001 | 0.096 | 0.173 | 0.222 | **0.234** |

(b) Scenario 2

| $t$ | $\Psi_{t,RCT}$ | $\Psi_{t,1}$ | $\Psi_{t,2}$ | $\Psi_{t,3}$ | $\Psi_{t,4}$ | $\Psi_{t,5}$ |
|---|---|---|---|---|---|---|
| 11 | 0.000 | **0.073** | 0.057 | 0.039 | 0.019 | 0.004 |
| 21 | 0.000 | **0.087** | 0.081 | 0.073 | 0.061 | 0.038 |
| 31 | 0.000 | **0.090** | 0.086 | 0.081 | 0.073 | 0.052 |
| 41 | 0.000 | **0.091** | 0.089 | 0.085 | 0.078 | 0.060 |
| 50 | 0.000 | **0.092** | 0.090 | 0.087 | 0.081 | 0.065 |

comes serve as better proxies for the final outcome when determining optimal individualized treatments, TMLE-OSLAD progressively shifts its selections toward these later surrogates, ultimately favoring $Y_4$ and $Y_5$ most frequently. In contrast, in Scenario 2 where the earliest surrogate $Y_1$ remains the oracle surrogate throughout, TMLE-OSLAD predominantly selects early surrogates, with $Y_1$ most frequently selected, followed by $Y_2$ and $Y_3$, while $Y_4$ and $Y_5$ are rarely chosen.

**Performance of benefiting participants.** We use $d_{0,K} : \mathcal{W} \to \mathcal{A}$ to denote the true optimal individualized treatment rule for $Y_K$, which is defined as $d_{0,K}(W) := I(\bar{Q}_{0,K}(1,W) - \bar{Q}_{0,K}(0,W) > 0)$. This rule assigns to each individual the treatment that would maximize their expected final outcome, given their covariates. To evaluate the extent to which each design assigns treatments that match the underlying oracle optimal rule, we define probability of not assigning optimal individualized treatments as $p_t^{non\text{-}opt} := \frac{1}{E(t)} \sum_{i:t_i=t} I\left(A_i \neq d_{0,K}(W_i)\right)$, where $E(t)$ is the number of participants enrolled at time $t$. Similarly, $p_{t,RCT}^{non\text{-}opt}$ and $p_{t,SL}^{non\text{-}opt}$ denote this probability for the non-adaptive RCT and the proposed TMLE-OSLAD adaptive design, respectively. Table 1.6 in Section 1.8.7 compares these suboptimal treatment assignment probabilities across different designs. We also define the average regret for participants enrolled at time $t$ within an adaptive design as $R_t := \frac{1}{E(t)} \sum_{i:t_i=t} I\left(A_i \neq d_{0,K}(W_i)\right) \times |\bar{Q}_{0,K}(1,W_i) - \bar{Q}_{0,K}(0,W_i)|$. Here, each participant $i$ contributes the absolute difference in expected final outcomes between the optimal and actual treatments. A lower $R_t$ indicates a better design that brings benefit in $Y_K$ for participants enrolled at time $t$. We use $R_{t,k}$ to denote the average regret within an adaptive design responsive to $Y_k$ at time $t$, while $R_{t,RCT}$ and $R_{t,SL}$ represent the average regret for the RCT design and TMLE-OSLAD design, respectively. Table 1.7 reports average regret across simulations at $t = 11, 21, 31, 41$, and 50, with full trajectories shown in Figure 1.4.

In Scenario 1, where later surrogates are superior at identifying the optimal individualized treatment, the adaptive design responsive to $Y_5$, as expected, consistently demonstrates the lowest regret across the experiment. In contrast, in Scenario 2, the regret within the adaptive design that responds to $Y_1$ remains consistently the lowest. There is a noticeable trend that

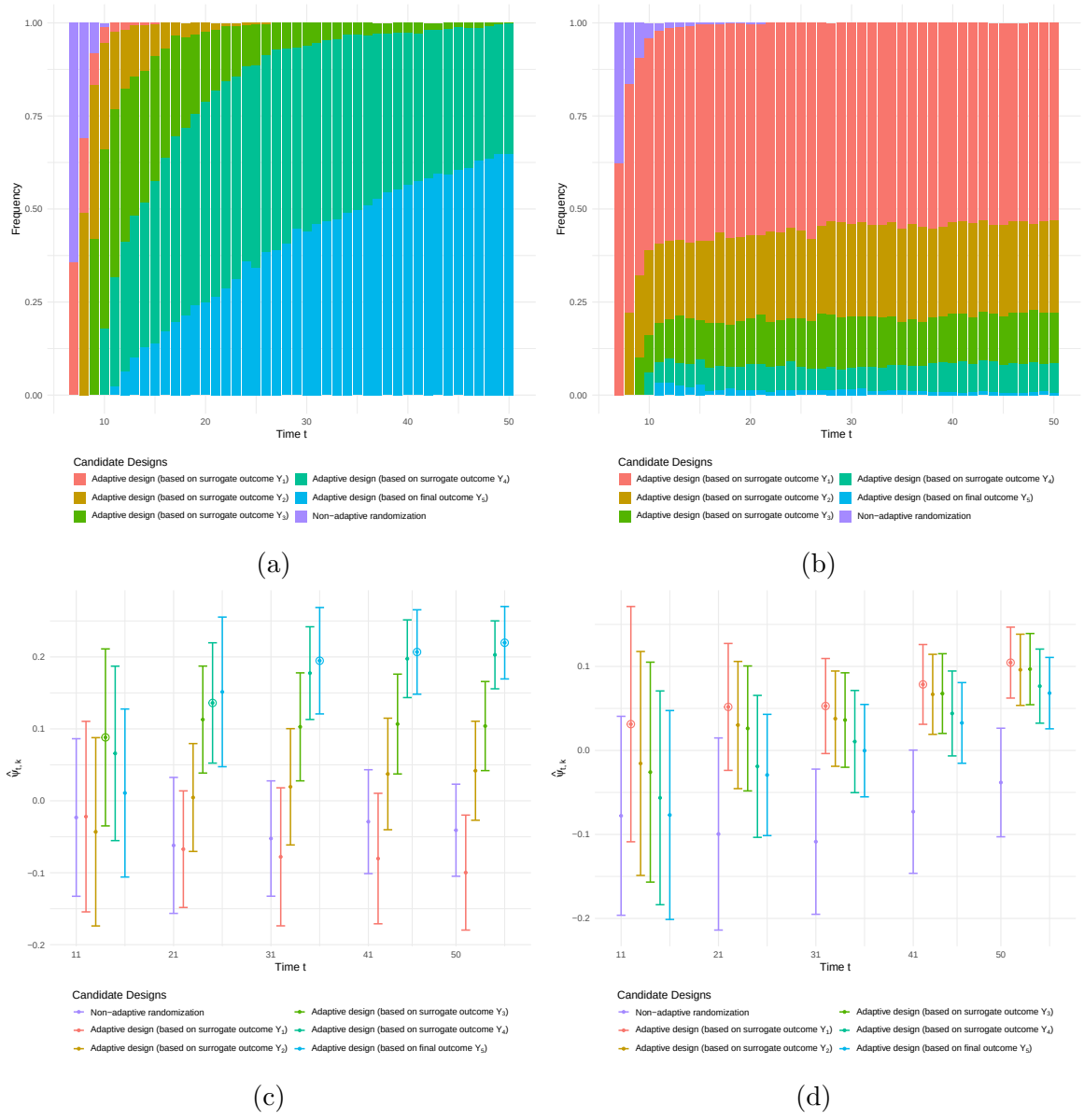



Figure 1.2: Frequency of design selection by TMLE-OSLAD at each time point. Panels (a) and (b) correspond to Scenario 1 and Scenario 2, respectively. Each color represents one candidate design, including adaptive designs based on outcomes $Y_k$ ($k \in \{1, \ldots, 5\}$) and the non-adaptive design. The height of each bar indicates the proportion of simulation runs in which that design attains the highest lower bound of the 95% confidence interval for $\hat{\Psi}_{t,k}$ among all candidate designs at time $t$, reflecting the selection mechanism of TMLE-OSLAD. Panels (c) and (d) display the selection mechanism at illustrative time points ($t = 11, 21, 31, 41, 50$) in a single simulation run under Scenarios 1 and 2, respectively. Colored points show the TMLE point estimates of $\Psi_{t,k}$ (y-axis) for evaluating each candidate design at each time point (x-axis). Vertical bars indicate the corresponding 95% Wald confidence intervals. At each time point, the selected design (defined as the one with the largest lower confidence bound among all candidate designs) is highlighted by a circular outline.

the adaptive design in response to the final outcome $Y_5$ exhibits a slower convergence toward the regret of the oracle surrogate $Y_1$, with a substantial discrepancy in regret that persists even at $t = 50$. Of course, in practice, the oracle surrogate is not known *a priori*, motivating the use of the TMLE-OSLAD procedure to evaluate and select among candidate surrogates in adaptive design. Across both scenarios, the proposed adaptive design demonstrates stable efficacy in approximating the minimal achievable regret compared to other designs. Its regret ranks second lowest starting at $t = 23$ in Scenario 1, and matches the lowest regret trajectory of $Y_1$ following $t = 21$ in Scenario 2. This robust performance reflects the advantage of the real-time evaluation and selection mechanism in TMLE-OSLAD, which continuously assesses each candidate surrogate-guided adaptive design, selects the most advantageous one to guide treatment randomization in subsequent stages.

**Post-experiment inference for other causal estimands.** Table 1.2 reports TMLE performance in estimating the additional causal estimands introduced in Section 1.3.3, including the average treatment effect and the mean final outcome under the estimated optimal dynamic treatment rules that maximize surrogate outcomes and the primary outcome. The mean final outcome $Y_5$ under the estimated optimal dynamic treatment rule for $Y_5$ is estimated using five-fold CV-TMLE. Overall, the TMLE-based estimators exhibit low bias and variance, with coverage near the nominal level across different estimands and scenarios.

Table 1.2: Performance of TMLEs for other causal estimands at the end of experiments in Scenarios 1 and 2 across Monte Carlo simulations.

(a) Scenario 1

| Estimand | Notation | Truth | Bias ($\times 10^{-3}$) | Var ($\times 10^{-3}$) | Coverage (%) |
|---|---|---|---|---|---|
| Average treatment effect on the final outcome $Y_5$ | $\psi^{\text{ATE}}_{T_{end}}$ | 0.468 | 3.94 | 3.92 | 95.8 |
| Expected final outcome $Y_5$ under the estimated ODTR based on surrogate $Y_1$ | $\psi^{d^*_{n,1}}_{T_{end}}$ | 0.011 | -1.01 | 2.85 | 94.4 |
| Expected final outcome $Y_5$ under the estimated ODTR based on surrogate $Y_2$ | $\psi^{d^*_{n,2}}_{T_{end}}$ | 0.130 | -0.87 | 2.02 | 93.8 |
| Expected final outcome $Y_5$ under the estimated ODTR based on surrogate $Y_3$ | $\psi^{d^*_{n,3}}_{T_{end}}$ | 0.235 | -0.08 | 1.62 | 93.6 |
| Expected final outcome $Y_5$ under the estimated ODTR based on surrogate $Y_4$ | $\psi^{d^*_{n,4}}_{T_{end}}$ | 0.312 | -1.01 | 0.89 | 94.6 |
| Expected final outcome $Y_5$ under the estimated ODTR based on $Y_5$ | $\psi^{d^*_{n,5}}_{T_{end}}$ | 0.340 | 0.12 | 0.73 | 95.4 |

(b) Scenario 2

| Estimand | Notation | Truth | Bias ($\times 10^{-3}$) | Var ($\times 10^{-3}$) | Coverage (%) |
|---|---|---|---|---|---|
| Average treatment effect on the final outcome $Y_5$ | $\psi^{\text{ATE}}_{T_{end}}$ | 0.000 | -1.17 | 3.99 | 94.8 |
| Expected final outcome $Y_5$ under the estimated ODTR based on surrogate $Y_1$ | $\psi^{d^*_{n,1}}_{T_{end}}$ | 0.120 | -0.25 | 0.57 | 93.8 |
| Expected final outcome $Y_5$ under the estimated ODTR based on surrogate $Y_2$ | $\psi^{d^*_{n,2}}_{T_{end}}$ | 0.120 | -0.32 | 0.59 | 94.2 |
| Expected final outcome $Y_5$ under the estimated ODTR based on surrogate $Y_3$ | $\psi^{d^*_{n,3}}_{T_{end}}$ | 0.120 | -0.57 | 0.61 | 93.6 |
| Expected final outcome $Y_5$ under the estimated ODTR based on surrogate $Y_4$ | $\psi^{d^*_{n,4}}_{T_{end}}$ | 0.119 | -0.38 | 0.60 | 94.6 |
| Expected final outcome $Y_5$ under the estimated ODTR based on $Y_5$ | $\psi^{d^*_{n,5}}_{T_{end}}$ | 0.114 | -1.37 | 0.85 | 93.0 |

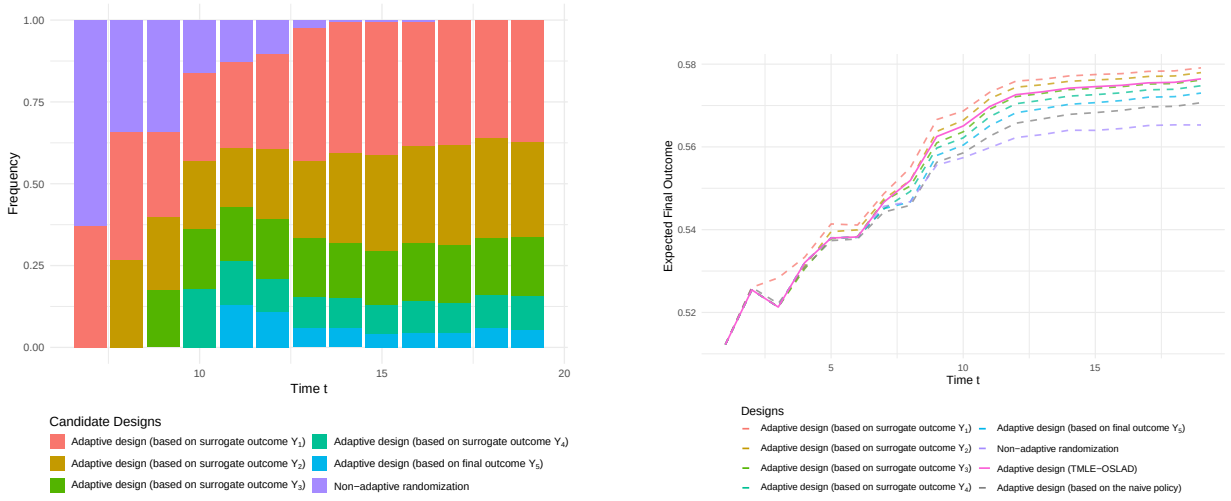
## 1.6 Application

We conduct an extended simulation study inspired by the Adaptive Strategies for Preventing and Treating Lapses of Retention in HIV Care (ADAPT-R) trial [24], a sequential multiple assignment randomized (SMART) trial designed to evaluate adaptive strategies for improving retention in care among individuals initiating HIV treatment. We demonstrate our proposed TMLE-OSLAD framework in this context, showing how it utilizes effective surrogates to benefit participants and evaluating TMLE performance in this setup.

**Simulation setup.** We focus on Phase 2 of the ADAPT-R trial, where participants who had a lapse in care in Phase 1 (defined as missing a clinic visit by 14 or more days) were re-randomized to a re-engagement strategy: either a peer navigation strategy (Navigator) or a combined Short Message Service reminder of appointments and conditional cash transfer (SMS+CCT). We discretize the study timeline into 50-day stages and define outcomes as the proportion of time in care during which a participant is engaged in across five windows, from the first day of receiving the re-engagement strategy to five specific time points: 50, 100, 150, 200, and 250 days. These outcomes are labeled as $Y_1$, $Y_2$, $Y_3$, $Y_4$, and $Y_5$, respectively, with $Y_5$ (proportion of time in care in 250 days) the primary outcome, and $Y_1$ through $Y_4$ the surrogate outcomes. We consider two covariates in the conditioning set of heterogeneous treatment effect: (a) initial Phase 1 treatment arm: SMS, CCT, or standard of care (SOC) and (b) time from Phase 1 enrollment to initial lapse in care. We used data from 215 participants who were assigned either Navigator or SMS+CCT in Phase 2 after a lapse in care in Phase 1 and fit a generalized linear model to obtain the counterfactual outcomes for simulation. The CATE functions for the surrogate outcomes and the final outcome are presented in Figure 1.13 (see Section 1.8.9 with more details). We implement a series of CARA designs as outlined in Section 1.5, including a simple sequential RCT, five adaptive designs that use a fixed choice from $Y_1$ through $Y_5$, and the proposed TMLE-OSLAD adaptive design that evaluates and selects the best surrogate in real time. We also compared their performance with a naive policy that employs the latest available outcome at each time point to adapt treatment randomization probabilities. In each simulation, participants arrive following the enrollment schedule in the original trial with their observed covariates, and treatment assignments are generated sequentially according to the type of the simulated design and the covariate, treatment and outcome data of previously enrolled participants, then the outcomes are generated in response to the assigned treatment. This construction ensures that each observation is conditionally independent given the past. We simulate each design 500 times to assess the performance of TMLEs and different adaptive designs.

**Results.** We compare regret over time obtained from implementing different design strategies (Figure 1.14 in Section 1.8.9). The adaptive design using the earliest outcome $Y_1$ outperforms those based on later outcomes or the naive policy from $t = 2$ through $t = 11$. After $t = 12$, adaptive designs using later outcomes start achieving smaller regret

for new subjects than the earliest outcome, as the underlying true CATE functions of later surrogates are closer to that for the final outcome and are estimated more precisely as data accumulate. However, this delayed improvement does not compensate the overarching benefit provided by the earliest surrogate, which delivers prompt benefits to participants by adopting surrogate-guided optimal personalized treatments learned from earlier and more abundant data. For the naive policy, although it initially matches the best performing adaptive design based on $Y_1$, its aggressive shift to the most recent outcomes leads it to behave similarly to the suboptimal later-outcome designs, resulting in higher regret. In contrast, TMLE-OSLAD predominantly selects the earliest surrogates ($Y_1$ and $Y_2$) over time (Figure 1.3a). It yields a higher expected final outcome (Figure 1.3b) than designs that adhere to later outcomes without ongoing evaluation and approaches the performance of the oracle design using the earliest surrogate. TMLE results for the proposed estimands during the experiment and for additional estimands after the experiment are reported in Section 1.8.9, showing low bias and variance across all estimands. Confidence-interval coverage is close to nominal for most estimands, with finite-sample undercoverage observed for some estimands under estimated optimal treatment rules.




(a) (b)

Figure 1.3: Panel (a) shows the frequency of candidate adaptive design selections by TMLE-OSLAD at each time point in the ADAPT-R trial simulation. Panel (b) shows the cumulative average expected final outcome over time obtained from implementing different designs.

## 1.7 Conclusions

In summary, this work provides a general TMLE-based Online Super Learner Adaptive Design Framework (TMLE-OSLAD) for constructing a real-time, data-driven ensemble system

for adaptive experimental designs and shows how this framework contributes to the statistical theory and application of surrogate outcomes in adaptive designs. We propose a novel causal estimand for surrogate utility in sequential adaptive designs: the mean primary outcome under a counterfactual CARA design where, at enrollment, each participant's randomization is shifted toward the optimal individualized rule learned from a candidate surrogate. It captures both how well the surrogate guides optimal treatment assignment and the extent to which it enables earlier benefit for more participants, with the goal of improving participants' primary outcomes. Building on this, we define an Online Super Learner adaptive design that incorporates the proposed estimand to evaluate a library of candidate surrogates, selects the most useful surrogate based on TMLE inference, and dynamically adapts randomization probabilities by the selected surrogate-guided design.

This work on surrogate outcomes is a concrete application of our highly general TMLE-OSLAD framework, which provides statistical inference on the mean counterfactual outcome under each candidate adaptive design in a user-supplied set, and adaptively evaluates and updates the best-performing design over time to adapt treatment randomization for new participants. This unified framework enables principled evaluation of the opportunities and costs of diverse adaptive designs varying in their design choices, quantifies how accumulating information benefits participants over time, and supports dynamic navigation of trade-offs among multiple designs during an ongoing experiment.

Methodologically, this framework advances non-parametric statistical inference for adaptive experiments: the proposed TMLE approach yields consistent, asymptotically normal estimators for the proposed estimand and other classical causal estimands, leveraging martingale theories and equicontinuity results. It broadly applies to estimation of various estimands in sequentially dependent settings like adaptive designs and time series.

Beyond the proposed context-specific estimand, further study of the computationally intensive "ideal estimand", the expected outcome under the full data distribution induced by a candidate design (as is discussed in Section 1.8.1), could provide more insight into the strengths and limitations of different estimands, as well as trade-offs between statistical properties and computational efficiency. We leave this to future work.

## 1.8 Supplementary Material

### 1.8.1 Additional Remarks on the Target Estimand

#### Surrogate Evaluation with Adaptive Design Perspective

Our approach adopts an adaptive design perspective to evaluate surrogates in sequential decision-making contexts. The proposed target estimand evaluates the utility of surrogates by their early availability, predictability, and sensitivity in indicating effective treatments in the following sense. We define a trajectory of treatment rules $\tilde{g}^*_{k,1:t} := (\tilde{g}^*_{k,1}, \ldots, \tilde{g}^*_{k,t})$, with each rule $\tilde{g}^*_{k,t'}$ generating a personalized treatment randomization function $g^*_{k,i}$ for each participant $i$ enrolled at time $t'$. Every $g^*_{k,i}$ is specifically tailored to maximize the surrogate

outcome $Y_k$ based on their baseline covariates $W_i$ and context. We note that at each time point $t'$, $\tilde{g}^*_{k,t'}$ depends only on historical data $\overline{\mathcal{H}}(t')$. Therefore, $\tilde{g}^*_{k,t'}$ can not deviate from a static design with equal randomization probabilities during the initial $k$ time points, as observation of $Y_k$ is not yet available. Once $Y_k$ is available at $t' = k+1$, $\tilde{g}^*_{k,t'}$ starts moving towards its optimal dynamic treatment rule $d_{0,k}(W) := I\left(\bar{Q}_{0,k}(1,W) - \bar{Q}_{0,k}(0,W) > 0\right)$ over time. If $d_{0,k}$ closely aligns with the true optimal rule for the primary outcome $d_{0,K}$, then the surrogate $Y_k$ demonstrates the advantage of its early availability in informing which treatment is better in the context of sequential decision making, compared to later outcomes. The progression of $\tilde{g}^*_{k,t'}$ from a static design with equal randomization probabilities towards the optimal dynamic treatment rule $d_{0,k}$ depends on the magnitude of treatment superiority (as measured by the point estimate of CATE) and its uncertainty (as reflected by estimated standard error of CATE). Therefore, if a surrogate $Y_k$ exhibits good alignment with the primary outcome in terms of optimal dynamic treatment rule and presents a strong signal of individualized optimal treatment over other surrogates, this surrogate will be preferred and given a higher $\Psi_{t,k}$. Of course, even if a surrogate demonstrates early availability and sensitivity, it will present a lower $\Psi_{t,k}$ value if it fails to predict optimal personalized treatments of the primary outcome. This is because the ultimate evaluation of a surrogate's performance is based on the expected final outcome, underscoring the critical role of predictability in evaluating surrogate outcomes in adaptive designs.

In summary, our proposed target estimand provides a comprehensive measure of a candidate surrogate's utility to accelerate participant benefit in sequential adaptive designs. It captures not only the timeliness with which the surrogate becomes available for early adaptation, but also its sensitivity to heterogeneous treatment effects and its predictive accuracy in identifying optimal personalized treatments for the primary outcome. This approach contrasts with existing surrogate evaluation methods, which are generally developed for static settings rather than from a dynamic adaptive design perspective.

### Approximation of an Ideal Estimand

Ideally, one could fully evaluate a candidate adaptive design by defining an ideal estimand as the expected final outcome when the entire experiment is conducted under that design, averaged across all participants. Formally, this requires integrating over the joint distribution of entry-time histories, contexts, treatments, and outcomes that would be sequentially induced if the candidate adaptive design were implemented from the start of the experiment. In practice, however, targeting this estimand is computationally intensive and entails a substantially more involved estimation procedure.

Instead, our context-specific estimand provides a fast and robust approximation to the ideal estimand by evaluating the expected outcome under the candidate design using contexts of the observed histories under the actual experiment. This estimand avoids modeling the full induced distribution under the candidate design and admits doubly robust estimation via TMLE (see Section 1.4). The approximation is most accurate when the observed contexts from the actual experiment are representative of those the candidate design would induce

- especially when the candidate design's adaptation is based on learning from a stationary target such as $Q_0$ or CATE that dominates over the experiments. This provides an effective approach to distinguishing low-utility from high-utility adaptive designs, thereby facilitating online selection among candidate designs with robust statistical support.

In simulation studies, we also compute the expected counterfactual final outcome obtained by implementing each candidate adaptive design (see Table 1.9 in Section 1.8.7), which corresponds to the ideal estimand. The results show that our context-specific estimand (Table 1.1) closely matches this ideal target with minimal differences.

### 1.8.2 The Highly Adaptive Lasso Estimator

Highly Adaptive Lasso (HAL) [42, 3] is a flexible non-parametric regression estimator that minimizes empirical risk over the class of càdlàg (right-continuous with left limits) functions with bounded sectional variation norms [27]. Instead of relying on local smoothness assumptions, HAL imposes a global smoothness constraint through the sectional variation norm and achieves a convergence rate of order $n^{-1/3}$ up to logarithmic factors when using zero-order splines [5]. Recent work by [52] studies higher-order smoothness classes of càdlàg functions and develops higher-order spline HAL, showing that the $m$-th order HAL is asymptotically normal and attains a uniform convergence rate of order $n^{-\frac{m+1}{2m+3}}$ up to a logarithmic factor ($m = 1, 2, \ldots$). In this work, we use first-order HAL to estimate conditional average treatment effect (CATE) functions (Section 1.8.3).

We first introduce a class of $d$-variate càdlàg functions defined on a unit cube $[0,1]^d$ with bounded sectional variation norm [27]. For any $x \in [0,1]^d$, denote $x(j)$ as its $j$-th coordinate. For a given subset $s \subset \{1, \ldots, d\}$, denote $x_s = (x(j) : j \in s)$. Let $\bar{Q}$ be a real-valued $d$-variate càdlàg function on a unit cube $[0,1]^d$. Define $\bar{Q}_s : (0_s, 1_s] \to \mathbb{R}$ as $\bar{Q}_s(u_s) = \bar{Q}(u_s, 0_{-s})$, i.e., the $s$-specific section of $\bar{Q}$ that sets coordinates in the complement of subset $s$ equal to 0. The sectional variation norm of $\bar{Q}$ is defined by $\|\bar{Q}\|_v^* := |\bar{Q}(0)| + \sum_{s \subset \{1,\ldots,d\}, s \neq \emptyset} \int_{(0_s,1_s]} |d\bar{Q}(u_s, 0_{-s})|$. Suppose $\|\bar{Q}\|_v^*$ is finite, then $\bar{Q}$ can be represented as [27]:

$$\bar{Q}(x) = \int_{[0,x]} d\bar{Q}(u) = \bar{Q}(0) + \sum_{s \subset \{1,\ldots,d\}, s \neq \emptyset} \int_{(0_s,1_s]} \phi_{u_s}^{(0)}(x_s) d\bar{Q}(u_s, 0_{-s}),$$

where $\phi_{u_s}^{(0)}(x_s) := I(x_s \geq u_s)$ is a zero-order basis function.

One can use discrete measure with observed support points to approximate the presentation of $\bar{Q}$ [42, 3]. Given $n$ i.i.d samples $(x_i, y_i)$ $(i = 1, \ldots, n)$ with $x_i \in [0,1]^d$, $y_i \in \mathbb{R}$, one can use the $s$-specific support points $x_{s,i} := (x_i(j) : j \in s)$ observed for each unit $i$ across different subsets $s$ to approximate $\bar{Q}$. Specifically, the approximation is given by $\bar{Q}_\beta(x) = \beta_0 + \sum_{s \subset \{1,\ldots,d\}} \sum_{i=1}^n \beta_{s,i} \phi_{x_{s,i}}^{(0)}(x(s))$.

We define $Pf := \int f(o) dP(o)$ for a function $f$ with a data generating distribution $P$ that generates data $O \sim P$. Let $P_n f := \int f(o) dP_n(o)$ under empirical distribution $P_n$. Define the class $\bar{\mathcal{Q}}_{n,C}^{(0)} = \left\{ \bar{Q}_\beta(x) : |\beta_0| + \sum_{s \subset \{1,\ldots,d\}} \sum_{i=1}^n |\beta_{s,i}| < C \right\}$ for some constant $C$, and let $L(\bar{Q})$

be a loss function. A zero-order HAL-MLE is constructed by solving the following loss-based minimization problem: $\bar{Q}_n = \text{argmin}_{\bar{Q}_\beta \in \bar{\mathcal{Q}}^{(0)}_{n,C^u_n}} P_n L(\bar{Q}_\beta)$, where $C^u_n$ is a data-adaptive cross-validation or undersmoothed selector of bounded sectional variation norm for the working model constructed on the basis functions $\phi^{(0)}_{x_{s,i}}$'s. [5] establishes the $L_2$ convergence rate of $n^{-\frac{1}{3}}$ up to logarithmic factors for the zero-order HAL-MLE.

[52] extends the HAL construction to higher-order smoothness classes. For integer $m \geq 1$, define the $m$-th order smoothness class $D^{(m)}([0,1]^d)$ $(m = 1, 2, \ldots)$ on $[0,1]^d$ as the class with functions whose $m$-th-order Lebesgue-Radon-Nikodym derivatives are càdlàg functions with bounded sectional variation norm. Like the zero-order HAL-MLE, one can construct a higher-order HAL-MLE that uses $m$-th order splines to estimate the target function in $D^{(m)}([0,1]^d)$. For the first-order $(m = 1)$, the approximation of $\bar{Q}^{(1)} \in D^{(1)}([0,1]^d)$ can be written by $\bar{Q}^{(1)}_\beta(x) = \beta_0 + \sum_{s \subset \{1,\ldots,d\}} \sum_{i=1}^n \beta_{s,i} \phi^{(1)}_{x_{s,i}}(x(s))$, where $\phi^{(1)}_u(x) = (x-u)I(x \geq u)$. Define $\bar{\mathcal{Q}}^{(1)}_{n,C} = \{\bar{Q}^{(1)}_\beta \in D^{(1)}([0,1]^d) : |\beta_0| + \sum_{s \subset \{1,\ldots,d\}} \sum_{i=1}^n |\beta_{s,i}| < C\}$. The first-order HAL-MLE is constructed by $\bar{Q}^{(1)}_n = \text{argmin}_{\bar{Q}^{(1)}_\beta \in \bar{\mathcal{Q}}^{(1)}_{n,C^u_n}} P_n L(\bar{Q}^{(1)}_\beta)$. One can also consider the second-order spline function $\phi^{(2)}_u(x) = \frac{1}{2}(x-u)^2 I(x \geq u)$ or higher-order spline basis functions to fit higher-order HAL-MLEs. Moreover, [52] establishes uniform covering number and entropy bounds for these higher-order HAL estimators, and shows that the $m$-th order spline HAL-MLE achieves asymptotic normality and a uniform convergence rate of $n^{-\frac{m+1}{2m+3}}$ up to logarithmic factors.

### 1.8.3 Estimation of Conditional Average Treatment Effect

A key procedure in our adaptive design is estimating the conditional average treatment effect (CATE). By estimating the CATE function, the experimenter can identify the optimal dynamic treatment rule that maximizes the expected outcome conditional on baseline covariates, and update treatment randomization probabilities accordingly to increase the chance of allocating the estimated optimal personalized treatment in adaptive design.

In this work, we use the Highly Adaptive Lasso (HAL) estimator to estimate the CATE function, with its advantages introduced in Section 1.8.2. Furthermore, we apply the delta-method to obtain a Wald-type confidence interval for the CATE. This confidence interval guides the exploration-exploitation trade-off in adaptive designs by adjusting treatment randomization probabilities based on the uncertainty in CATE estimates, which will be discussed in Section 1.8.4.

Recall that $\bar{Q}_{0,k}(A, W) = E[Y_k|A, W]$ is the true conditional expectation of the $k$-th surrogate outcome, given the treatment $A$ and baseline covariates $W$. For every new participant $i$ enrolled at time $t$, we use $B_{n,k,t}(W_i)$ to denote an estimate of $B_{0,k}(W_i) := E[Y_k(1) - Y_k(0)|W_i]$, the true CATE of $Y_k$ given $W_i$, the baseline covariates of participant $i$. We use $\tau_{n,k,t}(W_i)$ to denote the estimated standard error of CATE. Both estimates are based on $\mathcal{H}(i)$, the historical data observed before participant $i$ is enrolled.

For each surrogate outcome $Y_k$, we estimate CATE functions by two steps. First, we transform $Y_k$ into the pseudo-outcome [43, 60]:

$$\eta_k(\bar{Q}_k, g)(O) = \frac{2A-1}{g(A \mid W)}[Y_k - \bar{Q}_k(A, W)] + \bar{Q}_k(1, W) - \bar{Q}_k(0, W).$$

Note that the pseudo-outcome has the doubly-robust property that $E[\eta_k(\bar{Q}_k, g)(O) \mid W] = B_{0,k}(W)$ if $\bar{Q}_k = \bar{Q}_{0,k}$ or $g = g_0$. Since $g_0$ is known in adaptive design, we only need to estimate $\bar{Q}_{0,k}$ with $\bar{Q}_{n,k}$, then impute $\bar{Q}_{n,k,t}$ and $g_0$ to generate the pseudo-outcome $\eta_k(\bar{Q}_{n,k,t}, g_0)$.

Second, we estimate the CATE function by regressing the pseudo-outcome $\eta_k\left(\bar{Q}_{n,k,t}, g_0\right)$ on baseline covariates $W$ by first-order spline HAL-MLE, which minimizes the empirical risk over $D^{(1)}([0,1]^d)$, the first-order smoothness class with its first-order derivative being càdlàg function with bounded sectional variation norm. That is, at time $t$, we estimate the CATE function by solving the following minimization problem:

$$B_{n,k,t} = \underset{B_k \in \bar{\mathcal{Q}}^{(1)}_{C^u_{n,k,t}}}{\operatorname{argmin}} \; P_{n_{t,k}} L_k(B_k)$$

where $L_k(B_k)(O) := (B_k(W) - \eta_k(O))^2$, and $P_{n_{t,k}}$ denotes the empirical distribution over the $n_{t,k} := N(t-k)$ observations with observed $Y_k$ at time $t$; $C^u_{n,k,t}$ is an upper bound for the first-order sectional variation norm that is data-adaptively selected by cross-validation.

As introduced in Section 1.8.2, $B_{n,k,t}(\cdot)$ can be written as a linear combination of first-order spline basis functions with non-zero coefficients. To quantify the uncertainty of the CATE, we use the delta method under this HAL working model to obtain pointwise inference for the conditional mean of the pseudo-outcome, which corresponds to the CATE. Let $\tau_{n,k,t}(W_i)$ denote the estimate of the standard error of $B_{n,k,t}(W_i)$; a Wald-type $100(1-\alpha)\%$ confidence interval for the CATE at $W_i$ is $B_{n,k,t}(W_i) \pm z_{1-\alpha/2}\tau_{n,k,t}(W_i)$ ($\alpha = 0.05$). Both $B_{n,k,t}(W_i)$ and $\tau_{n,k,t}(W_i)$ are then incorporated into the treatment randomization function introduced in Section 1.8.4 to construct $g^*_{k,i}$ for each participant $i$ in the adaptive design guided by surrogate $Y_k$.

### 1.8.4 Details of CARA Procedures

We define a treatment randomization function $\tilde{h}_\nu : [-1,1] \times \mathbb{R}^+ \to [0,1]$ as follows:

$$\tilde{h}_\nu(x, b) = \nu I(x \le -b) + (1-\nu) I(x \ge b) + \left(-\frac{1/2-\nu}{2b^3}x^3 + \frac{1/2-\nu}{2b/3}x + \frac{1}{2}\right) I(-b < x < b),$$

where $\nu \in (0, 0.5)$ is a user-specified constant ensuring that $\nu \le \tilde{h}_\nu \le 1-\nu$.

This mapping, adapted from [14], is extended here to explicitly incorporate both the magnitude and uncertainty of the estimated CATE. In our setup, $x$ is the point estimate of CATE for each subject, while $b$ is the estimated standard error of CATE, scaled by the $z_{1-\alpha/2}$ quantile of the standard normal distribution.

For interpretation, it is helpful to re-express this function in terms of the standardized CATE $z = x/b$, which is the CATE divided by half the width of its confidence interval. The function then becomes

$$h_\nu(z) = \begin{cases} \nu, & z \le -1 \\ \nu + (1-2\nu)\left(-\frac{1}{4}z^3 + \frac{3}{4}z + \frac{1}{2}\right), & -1 < z < 1 \\ 1-\nu, & z \ge 1 \end{cases}$$

This function smoothly interpolates between balanced randomization and deterministic allocation of optimal personalized treatment based on the standardized CATE signal relative to uncertainty. When the signal is modest, a near-balanced allocation is maintained between treatment and control, avoiding over-commitment to a noisy or unstable estimated rule and ensuring continued learning of CATE by preserving exploration in both arms. As evidence accumulates and the magnitude of the CATE signal increases, the treatment randomization probability shifts further toward the estimated optimal treatment. Meanwhile, a minimal exploration probability is retained, which is necessary for estimating CATE during the experiment and a range of other causal estimands at the end of the adaptive experiment.

In our surrogate-guided adaptive designs, the treatment randomization probabilities are generated as follows. For each candidate surrogate outcome $Y_k$ and each subject $i$ with baseline covariates $W_i$, we estimate CATE $B_{n,k}(W_i)$ and its standard error $\tau_{n,k}(W_i)$ as detailed in Section 1.8.3. Consequently, the treatment randomization probability for $a = 1$ is given by $h_\nu(\frac{B_{n,k}(W_i)}{z_{1-\alpha/2}\cdot\tau_{n,k}(W_i)})$, and for $a = 0$ it is $1 - h_\nu(\frac{B_{n,k}(W_i)}{z_{1-\alpha/2}\cdot\tau_{n,k}(W_i))})$, where $\nu = 0.1$. If the surrogate outcome $Y_k$ has not yet been observed, the assignment defaults to equal randomization probability of 0.5.

### 1.8.5 Equicontinuity Condition for a General Martingale Process

In this section, we provide a theorem about the equicontinuity condition of martingale processes in a general setup that includes but is not limited to adaptive design settings.

#### Statistical Setup

We consider a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_i\}_{i\ge1}, \mathbf{P})$, where $\{O_i\}_{i\ge1}$ is a sequence of random variables taking values in $\mathcal{O}$, and $\mathcal{F}_i := \sigma(O_1, \ldots, O_i)$. Let $\bar{O}(i) := (O_1, \ldots, O_i)$. We assume that the data are generated sequentially with each $O_i$ depending on $\bar{O}(i-1)$ through a fixed-dimensional summary $C_i = f_C(\bar{O}(i-1)) \in \mathcal{C}$. Let $\Theta$ be a class of functions, and $\{f_\theta : \theta \in \Theta\}$ be measurable functions from $\mathcal{O} \times \mathcal{C} \to \mathbb{R}$. Define

$$M_N(\theta) = \frac{1}{N}\sum_{i=1}^{N} \left\{ f_\theta(O_i, C_i) - E\left[ f_\theta(O_i, C_i) \mid \bar{O}(i-1) \right] \right\}.$$

Given a fixed $\theta^* \in \Theta$ and a $\theta_N \in \Theta$ constructed from data $\bar{O}(N)$, we provide a general theorem for the equicontinuity condition with respect to the martingale process as follows:

$$M_{2,N}(\theta_N, \theta^*) := M_N(\theta_N) - M_N(\theta^*) = o_P(N^{-\frac{1}{2}}).$$

**Proof of Equicontinuity Condition**

We first define a bracketing entropy following Definition A.1 in [30] (see also Definition 8.1 in [23]).

**Definition 1.1.** *(Bracketing entropy) Consider a stochastic process* $\left\{(\xi_{\theta,i})_{i=1}^N : \theta \in \Theta\right\}$, *where* $\Theta$ *is an index set such that, for every* $\theta \in \Theta, i \in [N]$, $\xi_{\theta,i}$ *is an* $\bar{O}(i)$*-measurable random variable for all* $i$ *and* $\theta$. *We say that a collection of random variables of the form* $\mathcal{B} := \left\{\left(\Lambda_i^j, \Upsilon_i^j\right)_{i=1}^N : j \in [J]\right\}$ *is an* $(N, \Theta, Z, \epsilon)$*-bracketing if*

1. *for every* $i \in [N]$, *and* $j \in [J]$, $\left(\Lambda_i^j, \Upsilon_i^j\right)$ *is* $\bar{O}(i)$*-measurable,*
2. *for every* $\theta \in \Theta$, *there exists* $j \in [J]$, *such that, for every* $i \in [N], \Lambda_i^j \leq \xi_{\theta,i} \leq \Upsilon_i^j$,
3. *for every* $j \in [J]$,

$$\frac{2Z^2}{N} \sum_{i=1}^{N} E\left[\phi\left(\frac{\Upsilon_i^j - \Lambda_i^j}{Z}\right) \mid \bar{O}(i-1)\right] \leq \epsilon^2,$$

*where* $\phi(x) = e^x - x - 1$. *We denote* $\mathcal{N}_{[]}(N, \Theta, Z, \epsilon)$ *the minimal cardinality of an* $(N, \Theta, Z, \epsilon)$*-bracketing. We define the bracketing entropy integral as* $\mathcal{J}_{[]}(N, \Theta, Z, \epsilon) := \int_0^\epsilon \sqrt{\log(1 + \mathcal{N}_{[]}(N, \Theta, Z, u))} du$.

**Theorem 1.6** (Theorem A.4 in [30])**.** *Given* $Z > 0$, *for all* $i \geq 0$, *define*

$$M_{\theta,i} = \frac{1}{N} \sum_{l=1}^{i} \left(\xi_{\theta,l} - E\left[\xi_{\theta,l} \mid \mathcal{F}_{l-1}\right]\right),$$

$$R_{\theta,i} = \frac{1}{N} 2Z^2 \sum_{l=1}^{i} E\left[\phi\left(\frac{|\xi_{\theta,l}|}{Z}\right) \middle| \mathcal{F}_{l-1}\right].$$

*Then, for any* $N \in \mathbb{N}$, $x > 0$, *and* $0 < R < \infty$,

$$P\left(\sup_{\theta \in \Theta} I\left(R_{\theta,N} \leq R\right) \max_{i \in [N]} \sqrt{N} M_{\theta,i} \geq 16\Gamma + 32\sqrt{Rx} + \frac{16}{\sqrt{N}} Zx\right) \leq 2e^{-x},$$

*where* $\Gamma := \frac{Z}{\sqrt{N}} \log \mathcal{N}_{[]}\left(N, \Theta, Z, \sqrt{R}\right) + 4\mathcal{J}_{[]}\left(N, \Theta, Z, \sqrt{R}\right)$.

We focus on the process $\xi_{\theta,i} = f_\theta(O_i, C_i) - f_{\theta^*}(O_i, C_i)$ for $i = 1, \ldots, N$, and $\theta, \theta^* \in \Theta$. Since the bracketing entropy condition may be difficult to verify, we relate it to the supremum norm covering entropy, following [23] (Section 8.2) and [13] (Lemma 7). In our setting, this is done by working with the covering entropy of $\Theta$.

**Assumption 1.8** (Entropy Integral)**.** *Let $\mathcal{N}_\infty(\epsilon, \Theta)$ denote the $\epsilon$-covering number in supremum norm of $\Theta$ and let $\mathcal{J}_\infty(\epsilon, \Theta) := \int_0^\epsilon \sqrt{\log(1 + \mathcal{N}_\infty(u, \Theta))} du$ denote the $\epsilon$-entropy integral in supremum norm of $\Theta$. Assume that $\mathcal{J}_\infty(\epsilon, \Theta)$ is finite and $\mathcal{J}_\infty(\epsilon, \Theta) \to 0$ as $\epsilon \to 0$.*

**Assumption 1.9** (Uniform Boundedness)**.** $\sup_{\theta \in \Theta} \|\theta\|_\infty < \infty$, $\sup_{\theta \in \Theta} \|f_\theta\|_\infty < \infty$.

**Assumption 1.10.** $\sup_{\theta_1 \in \Theta, \theta_2 \in \Theta, \theta_1 \neq \theta_2} \frac{\|f_{\theta_1} - f_{\theta_2}\|_\infty}{\|\theta_1 - \theta_2\|_\infty} < \infty$.

**Lemma 1.1.** *Suppose that Assumptions 1.8, 1.9 and 1.10 hold, such that $\|\theta\|_\infty < M$ for some finite constant $M > 0$ and $\sup_{\theta_1 \in \Theta, \theta_2 \in \Theta, \theta_1 \neq \theta_2} \frac{\|f_{\theta_1} - f_{\theta_2}\|_\infty}{\|\theta_1 - \theta_2\|_\infty} \leq L$ for some finite constant $L > 0$. Then the following holds:*

$$\mathcal{N}_{[]}\left(N, \Theta, 8LM, 2\sqrt{2}L\epsilon\right) \leq \mathcal{N}_\infty(\epsilon, \Theta),$$

*and*

$$\mathcal{J}_{[]}\left(N, \Theta, 8LM, 2\sqrt{2}L\epsilon\right) \leq 2\sqrt{2}L\mathcal{J}_\infty(\epsilon, \Theta).$$

*Proof.* Suppose $\{\theta_j : j \in [J]\} \subset \Theta$ is an $\epsilon$-covering of $\Theta$ in supremum norm such that $\forall \theta \in \Theta$, there exists $j \in [J]$ such that $\theta_j \in \Theta$ and $\|\theta_j - \theta\|_\infty \leq \epsilon$. By Assumption 1.9, the diameter of $\Theta$ in supremum norm is at most $2M$, so it suffices to consider $0 < \epsilon \leq 2M$. Under Assumption 1.10, $\left\|f_{\theta_j} - f_\theta\right\|_\infty \leq L \|\theta_j - \theta\|_\infty \leq L\epsilon$. Define $\Lambda_i^j = f_{\theta_j}(O_i, C_i) - f_{\theta^*}(O_i, C_i) - L\epsilon$ and $\Upsilon_i^j = f_{\theta_j}(O_i, C_i) - f_{\theta^*}(O_i, C_i) + L\epsilon$. Then, we have that

$$\begin{aligned}
\frac{2Z^2}{N} \sum_{i=1}^N E\left[\phi\left(\frac{|\Upsilon_i^j - \Lambda_i^j|}{Z}\right) \mid \bar{O}(i-1)\right] &\leq 2Z^2 \phi\left(\frac{2L\epsilon}{Z}\right) \\
&\leq 2Z^2 \sum_{p \geq 2} \frac{1}{p!} \frac{p!}{2} \left(\frac{2L\epsilon}{Z}\right)^p \\
&= \frac{2L\epsilon}{\frac{1}{2L\epsilon} - \frac{1}{Z}}.
\end{aligned}$$

Let $Z = 8LM$. We have that

$$\frac{1}{N} 2Z^2 \sum_{i=1}^N E\left[\phi\left(\frac{|\Upsilon_i^j - \Lambda_i^j|}{Z}\right) \mid \bar{O}(i-1)\right] \leq (2\sqrt{2}L\epsilon)^2.$$

Therefore,

$$\mathcal{N}_{[]}\left(N, \Theta, 8LM, 2\sqrt{2}L\epsilon\right) \leq \mathcal{N}_\infty(\epsilon, \Theta),$$

and

$$\mathcal{J}_{[]}\left(N,\Theta,8LM,2\sqrt{2}L\epsilon\right)\leq 2\sqrt{2}L\mathcal{J}_{\infty}\left(\epsilon,\Theta\right).$$

□

For any $\theta,\theta^{*}\in\Theta$, define

$$s_{N}\left(\theta,\theta^{*}\right):=\sqrt{\frac{1}{N}\sum_{i=1}^{N}E\left[\phi\left(\frac{\left|f_{\theta}\left(O_{i},C_{i}\right)-f_{\theta^{*}}\left(O_{i},C_{i}\right)\right|}{Z}\right)\mid\bar{O}(i-1)\right]},$$

where $Z=8LM$. Let $\tilde{s}_{N}(\theta,\theta^{*}):=\sqrt{2}Z\,s_{N}(\theta,\theta^{*})$.

**Theorem 1.7** (Maximal Inequality of a Martingale Process)**.** *For any $x>0,N\in\mathbb{N}$, and $0<r<\infty$, define $a_{N}(r,x):=\frac{16Z}{\sqrt{N}}\log\mathcal{N}_{\infty}\left(\frac{r}{2\sqrt{2}L},\Theta\right)+128\sqrt{2}L\mathcal{J}_{\infty}\left(\frac{r}{2\sqrt{2}L},\Theta\right)+32r\sqrt{x}+\frac{16}{\sqrt{N}}Zx$. Suppose that Assumptions 1.8, 1.9, and 1.10 hold. Then,*

$$P\left(\sup_{\theta\in\Theta}I\left(\tilde{s}_{N}(\theta,\theta^{*})\leq r\right)\sqrt{N}M_{2,N}(\theta,\theta^{*})\geq a_{N}(r,x)\right)\leq 2e^{-x}.$$

*Proof.* Recall that

$$M_{2,N}(\theta,\theta^{*})=\frac{1}{N}\sum_{i=1}^{N}\left\{f_{\theta}\left(O_{i},C_{i}\right)-f_{\theta^{*}}\left(O_{i},C_{i}\right)-E\left[f_{\theta}(O_{i},C_{i})-f_{\theta^{*}}(O_{i},C_{i})\mid\bar{O}(i-1)\right]\right\}.$$

Let $Z=8LM$ and $b_{N}(r,x)=\frac{16Z}{\sqrt{N}}\log\mathcal{N}_{[]}\left(N,\Theta,Z,r\right)+64\mathcal{J}_{[]}\left(N,\Theta,Z,r\right)+32r\sqrt{x}+\frac{16}{\sqrt{N}}Zx$. We have that

$$\begin{aligned}&P\left(\sup_{\theta\in\Theta}I\left(\tilde{s}_{N}\left(\theta,\theta^{*}\right)\leq r\right)\sqrt{N}M_{2,N}(\theta,\theta^{*})\geq a_{N}(r,x)\right)\\&\leq P\left(\sup_{\theta\in\Theta}I\left(\tilde{s}_{N}\left(\theta,\theta^{*}\right)\leq r\right)\sqrt{N}M_{2,N}(\theta,\theta^{*})\geq b_{N}(r,x)\right)\\&\leq 2e^{-x},\end{aligned}$$

where the first inequality is given by Lemma 1.1 and the second inequality is based on the maximal inequality in Theorem 1.6 with $R_{\theta,N}=\tilde{s}_{N}^{2}(\theta,\theta^{*})$ and $R=r^{2}$. □

**Assumption 1.11** (Convergence of $\theta_{N}$)**.** *Assume that $s_{N}\left(\theta_{N},\theta^{*}\right)=O_{P}(\delta_{N})$ for some $\delta_{N}\to 0$.*

**Corollary 1.1.** *Suppose Assumptions 1.8, 1.9, 1.10 and 1.11 hold, then $M_{2,N}(\theta_{N},\theta^{*})=o_{P}(N^{-\frac{1}{2}})$.*

*Proof.* Since $s_N(\theta_N, \theta^*) = O_P(\delta_N)$, and $\tilde{s}_N(\theta, \theta^*) = \sqrt{2}Z s_N(\theta, \theta^*)$, we also have $\tilde{s}_N(\theta_N, \theta^*) = O_P(\delta_N)$. Hence, one can find a deterministic sequence $r_N \to 0$ such that $\delta_N = o(r_N)$ and $\sqrt{N} r_N^2 \to \infty$. Then, for any $x > 0$, we have

$$\begin{aligned}
a_N(r_N, x) &= \frac{16Z}{\sqrt{N}} \log \mathcal{N}_\infty\left(\frac{r_N}{2\sqrt{2}L}, \Theta\right) + 128\sqrt{2}L\mathcal{J}_\infty\left(\frac{r_N}{2\sqrt{2}L}, \Theta\right) + 32 r_N\sqrt{x} + \frac{16}{\sqrt{N}} Z x \\
&\leq \frac{16Z}{\sqrt{N}} \left(\frac{2\sqrt{2}L}{r_N}\mathcal{J}_\infty\left(\frac{r_N}{2\sqrt{2}L}, \Theta\right)\right)^2 + 128\sqrt{2}L\mathcal{J}_\infty\left(\frac{r_N}{2\sqrt{2}L}, \Theta\right) + 32 r_N\sqrt{x} + \frac{16}{\sqrt{N}} Z x \\
&\leq \left(\frac{128L^2 Z}{\sqrt{N} r_N^2}\mathcal{J}_\infty\left(\frac{r_N}{2\sqrt{2}L}, \Theta\right) + 128\sqrt{2}L\right)\mathcal{J}_\infty\left(\frac{r_N}{2\sqrt{2}L}, \Theta\right) + 32 r_N\sqrt{x} + \frac{16}{\sqrt{N}} Z x \\
&\to 0.
\end{aligned}$$

Therefore, for any $y > 0$, there exists a $N_1 > 0$ such that for $N > N_1$, $a_N(r_N, x) < y$. Hence,

$$\begin{aligned}
&P(I\left(\tilde{s}_N\left(\theta_N, \theta^*\right) \leq r_N\right)\sqrt{N} M_{2,N}(\theta_N, \theta^*) > y) \\
&\leq P\left(\sup_{\theta \in \Theta} I\left(\tilde{s}_N\left(\theta, \theta^*\right) \leq r_N\right)\sqrt{N} M_{2,N}(\theta, \theta^*) \geq a_N(r_N, x)\right) \\
&\leq 2e^{-x},
\end{aligned}$$

where the last inequality is implied by Theorem 1.7. Similarly, for any $y > 0$ and the chosen $x > 0$, there exists a $N_2 > 0$, such that for every $N > N_2$,

$$P(-\sqrt{N} M_{2,N}(\theta_N, \theta^*) I\left(\tilde{s}_N\left(\theta_N, \theta^*\right) \leq r_N\right) > y) \leq 2e^{-x}.$$

Since the bound can be made arbitrarily close to zero by choosing $x$ sufficiently large, and $P(\tilde{s}_N\left(\theta_N, \theta^*\right) \leq r_N) \to 1$, it follows that $\sqrt{N} M_{2,N}(\theta_N, \theta^*) = o_P(1)$. □

### 1.8.6 Proof of Equicontinuity Result in Adaptive Designs

Recall that in Section 1.4

$$\begin{aligned}
M_{2,n_t}\left(\bar{Q}^*_{n,K,t}, \bar{Q}_{1,K}\right) = \frac{1}{n_t}\sum_{i=1}^{n_t}\Bigg\{ &\left[D_k\left(\bar{Q}^*_{n,K,t}\right)(O_i, C_i) - E\left[D_k\left(\bar{Q}^*_{n,K,t}\right)(O_i, C_i)\,|\mathcal{H}(i)\right]\right] \\
&- \left[D_k\left(\bar{Q}_{1,K}\right)(O_i, C_i) - E\left[D_k\left(\bar{Q}_{1,K}\right)(O_i, C_i)\,|\mathcal{H}(i)\right]\right]\Bigg\},
\end{aligned}$$

where $D_k(\bar{Q}_K)(O_i, C_i) := \frac{g^*_{k,i}(A_i|C_i)}{g_{0,i}(A_i|C_i)}\left(Y_{i,K} - \bar{Q}_K\left(A_i, W_i\right)\right)$. We apply the general results in Section 1.8.5 to prove the equicontinuity condition $M_{2,n_t}(\bar{Q}^*_{n,K,t}, \bar{Q}_{1,K}) = o_P(n_t^{-\frac{1}{2}})$ in Theorem 1.4 in the context of adaptive designs.

*Proof of Theorem 1.4 .* Let $\Theta = \bar{\mathcal{Q}}_K$, $f_\theta = D_k(\bar{Q}_K)$, $N = n_t$, $\theta_N = \bar{Q}^*_{n,K,t}$ in the general setup in Section 1.8.5. Under Assumption 1.4, $\zeta < g_{0,i} < 1-\zeta$ and $\zeta < g^*_{k,i} < 1-\zeta$. Since $Y_K$ and $Q_K$ take values in the interval $[0,1]$, we have that $\sup_{\bar{Q}_K \in \bar{\mathcal{Q}}_K} \left\| D_k(\bar{Q}_K) \right\|_\infty \leq \frac{2(1-\zeta)}{\zeta} < \infty$, $\sup_{\bar{Q}_{K,1} \in \bar{\mathcal{Q}}_K, \bar{Q}_{K,2} \in \bar{\mathcal{Q}}_K, \bar{Q}_{K,1} \neq \bar{Q}_{K,2}} \frac{\left\| D_k(\bar{Q}_{K,1}) - D_k(\bar{Q}_{K,2}) \right\|_\infty}{\left\| \bar{Q}_{K,1} - \bar{Q}_{K,2} \right\|_\infty} \leq \frac{1-\zeta}{\zeta} < \infty$. Therefore, Assumptions 1.9 and 1.10 are satisfied. Since Assumption 1.7 corresponds to Assumption 1.11 in the general framework, we can utilize Corollary 1.1 to demonstrate that $M_{2,n_t}\left(\bar{Q}^*_{n,K,t}, \bar{Q}_{1,K}\right) = o_P\left(n_t^{-\frac{1}{2}}\right)$. □

### 1.8.7 Additional Results for Section 1.5

The data generating distributions of the two simulation scenarios in Section 1.5 are as follows. The baseline covariate $W$ follows a uniform distribution $U(-4,4)$. In Scenario 1, the conditional mean functions of $Y_k$ given $A$ and $W$ are $E[Y_k|A,W] = (2A-1) \times [0.5 - (1+\exp(-(3-k)-W))^{-1}]$. In Scenario 2, the conditional mean functions of $Y_k$ are $E[Y_k|A,W] = (2A-1) \times [0.5 - (1+\exp(-\gamma_k W))^{-1}]$, where $\gamma_1 = 3$,$\gamma_2 = 2$,$\gamma_3 = 1$,$\gamma_4 = 0.5$ and $\gamma_5 = 0.25$. The treatment randomization probability of binary treatment $A \in \{0,1\}$ depends on the configuration of each design.

In simulation studies, we estimate these conditional mean functions using Super Learner [47], which includes ordinary least squares, Random Forest [8], the first-order Highly Adaptive Lasso (HAL) estimator [52] and an intercept model in the library. The CATE functions and their pointwise confidence intervals are estimated using the first-order HAL, as described in Section 1.8.3.

Table 1.3: Bias ($\times 10^{-3}$) of TMLE estimators for $\Psi_{t,k}$ in Scenarios 1 and 2 ($t = 11, 21, 31, 41, 50$).

(a) Scenario 1

| $t$ | $\Psi_{t,RCT}$ | $\Psi_{t,1}$ | $\Psi_{t,2}$ | $\Psi_{t,3}$ | $\Psi_{t,4}$ | $\Psi_{t,5}$ |
|---|---|---|---|---|---|---|
| 11 | 3.90 | 3.22 | 4.20 | 5.02 | 5.09 | 4.70 |
| 21 | -0.83 | -0.55 | -0.32 | -0.78 | 1.06 | 0.98 |
| 31 | 0.03 | -0.07 | -0.16 | 1.09 | 0.18 | 1.13 |
| 41 | 0.61 | -0.02 | 1.27 | 0.25 | 0.72 | -0.51 |
| 50 | -0.51 | -0.48 | -0.91 | 0.11 | -0.94 | -0.01 |

(b) Scenario 2

| $t$ | $\Psi_{t,RCT}$ | $\Psi_{t,1}$ | $\Psi_{t,2}$ | $\Psi_{t,3}$ | $\Psi_{t,4}$ | $\Psi_{t,5}$ |
|---|---|---|---|---|---|---|
| 11 | 1.96 | 2.94 | 1.42 | 2.57 | 2.91 | 2.96 |
| 21 | -2.73 | 0.23 | -0.35 | -0.59 | 0.43 | 0.97 |
| 31 | -1.34 | -0.61 | -1.17 | -1.24 | -0.92 | -0.72 |
| 41 | -1.18 | 0.14 | -0.17 | -0.36 | -0.22 | -0.58 |
| 50 | -1.84 | -0.23 | -0.42 | -0.74 | -0.42 | -0.86 |

Table 1.4: Variance ($\times 10^{-3}$) of TMLE estimators for $\Psi_{t,k}$ in Scenarios 1 and 2 ($t = 11, 21, 31, 41, 50$).

(a) Scenario 1

| $t$ | $\Psi_{t,RCT}$ | $\Psi_{t,1}$ | $\Psi_{t,2}$ | $\Psi_{t,3}$ | $\Psi_{t,4}$ | $\Psi_{t,5}$ |
|---|---|---|---|---|---|---|
| 11 | 3.16 | 7.34 | 6.59 | 5.35 | 4.07 | 3.51 |
| 21 | 2.11 | 4.10 | 3.25 | 2.55 | 1.91 | 2.11 |
| 31 | 1.50 | 3.26 | 2.29 | 1.65 | 1.16 | 1.17 |
| 41 | 1.14 | 2.57 | 1.71 | 1.26 | 0.82 | 0.79 |
| 50 | 1.04 | 2.34 | 1.61 | 1.15 | 0.72 | 0.65 |

(b) Scenario 2

| $t$ | $\Psi_{t,RCT}$ | $\Psi_{t,1}$ | $\Psi_{t,2}$ | $\Psi_{t,3}$ | $\Psi_{t,4}$ | $\Psi_{t,5}$ |
|---|---|---|---|---|---|---|
| 11 | 3.40 | 4.74 | 4.64 | 4.24 | 3.81 | 3.76 |
| 21 | 2.46 | 1.63 | 1.62 | 1.60 | 1.64 | 1.82 |
| 31 | 1.59 | 0.96 | 0.98 | 0.97 | 1.00 | 1.13 |
| 41 | 1.30 | 0.66 | 0.67 | 0.67 | 0.71 | 0.82 |
| 50 | 1.07 | 0.54 | 0.55 | 0.54 | 0.58 | 0.68 |

Table 1.5: Coverage probability (%) of confidence intervals for $\Psi_{t,k}$ in Scenarios 1 and 2 ($t = 11, 21, 31, 41, 50$).

(a) Scenario 1

| $t$ | $\Psi_{t,RCT}$ | $\Psi_{t,1}$ | $\Psi_{t,2}$ | $\Psi_{t,3}$ | $\Psi_{t,4}$ | $\Psi_{t,5}$ |
|---|---|---|---|---|---|---|
| 11 | 96.4 | 96.2 | 95.8 | 95.6 | 95.6 | 96.2 |
| 21 | 95.0 | 93.8 | 93.4 | 94.6 | 94.2 | 95.2 |
| 31 | 96.2 | 95.2 | 94.8 | 94.2 | 95.2 | 95.0 |
| 41 | 95.6 | 95.6 | 97.2 | 94.4 | 94.6 | 95.8 |
| 50 | 94.4 | 94.2 | 95.0 | 93.6 | 94.0 | 94.8 |

(b) Scenario 2

| $t$ | $\Psi_{t,RCT}$ | $\Psi_{t,1}$ | $\Psi_{t,2}$ | $\Psi_{t,3}$ | $\Psi_{t,4}$ | $\Psi_{t,5}$ |
|---|---|---|---|---|---|---|
| 11 | 96.2 | 95.2 | 94.8 | 95.0 | 95.4 | 94.6 |
| 21 | 94.0 | 95.2 | 95.2 | 94.4 | 94.4 | 95.0 |
| 31 | 94.4 | 94.2 | 94.2 | 94.8 | 94.0 | 95.4 |
| 41 | 94.8 | 95.0 | 94.4 | 94.0 | 95.4 | 95.4 |
| 50 | 94.6 | 94.2 | 94.8 | 94.2 | 94.6 | 94.8 |

Table 1.6: Probability (%) of not assigning optimal personalized treatments in Scenarios 1 and 2 ($t = 11, 21, 31, 41, 50$). $p_t^{non\text{-}opt} := \frac{1}{E(t)} \sum_{i:t_i=t} I\left(A_i \neq d_{0,K}(W_i)\right)$, where $d_{0,K}(W) := I(\bar{Q}_{0,K}(1,W) - \bar{Q}_{0,K}(0,W) > 0)$

(a) Scenario 1

| $t$ | $p_{t,SL}^{non\text{-}opt}$ | $p_{t,1}^{non\text{-}opt}$ | $p_{t,2}^{non\text{-}opt}$ | $p_{t,3}^{non\text{-}opt}$ | $p_{t,4}^{non\text{-}opt}$ | $p_{t,5}^{non\text{-}opt}$ | $p_{t,RCT}^{non\text{-}opt}$ |
|---|---|---|---|---|---|---|---|
| 11 | 28.1 | 51.7 | 40.4 | 29.9 | 20.2 | 15.9 | 50.2 |
| 21 | 20.0 | 51.2 | 39.9 | 29.7 | 19.7 | 14.4 | 49.7 |
| 31 | 17.8 | 51.0 | 40.9 | 30.3 | 19.6 | 13.5 | 50.5 |
| 41 | 15.8 | 50.4 | 40.1 | 30.3 | 19.9 | 13.3 | 49.9 |
| 50 | 15.0 | 50.8 | 40.2 | 30.1 | 19.7 | 12.4 | 50.0 |

(b) Scenario 2

| $t$ | $p_{t,SL}^{non\text{-}opt}$ | $p_{t,1}^{non\text{-}opt}$ | $p_{t,2}^{non\text{-}opt}$ | $p_{t,3}^{non\text{-}opt}$ | $p_{t,4}^{non\text{-}opt}$ | $p_{t,5}^{non\text{-}opt}$ | $p_{t,RCT}^{non\text{-}opt}$ |
|---|---|---|---|---|---|---|---|
| 11 | 15.5 | 13.9 | 14.3 | 15.7 | 18.7 | 29.8 | 49.8 |
| 21 | 13.4 | 13.1 | 13.2 | 13.9 | 16.3 | 22.7 | 49.7 |
| 31 | 13.2 | 12.3 | 12.4 | 12.7 | 14.8 | 19.7 | 49.5 |
| 41 | 12.5 | 11.8 | 11.8 | 12.7 | 13.9 | 17.9 | 49.5 |
| 50 | 12.4 | 12.0 | 12.1 | 12.5 | 14.2 | 16.9 | 50.1 |

Table 1.7: Average regret of not selecting optimal personalized treatments in Scenarios 1 and 2 ($t = 11, 21, 31, 41, 50$).

(a) Scenario 1

| $t$ | $R_{t,SL}$ | $R_{t,1}$ | $R_{t,2}$ | $R_{t,3}$ | $R_{t,4}$ | $R_{t,5}$ | $R_{t,RCT}$ |
|---|---|---|---|---|---|---|---|
| 11 | 0.153 | 0.350 | 0.245 | 0.159 | 0.100 | 0.083 | 0.343 |
| 21 | 0.102 | 0.344 | 0.239 | 0.154 | 0.095 | 0.076 | 0.342 |
| 31 | 0.089 | 0.341 | 0.247 | 0.158 | 0.094 | 0.074 | 0.346 |
| 41 | 0.083 | 0.337 | 0.240 | 0.158 | 0.094 | 0.074 | 0.341 |
| 50 | 0.080 | 0.341 | 0.241 | 0.157 | 0.093 | 0.071 | 0.345 |

(b) Scenario 2

| $t$ | $R_{t,SL}$ | $R_{t,1}$ | $R_{t,2}$ | $R_{t,3}$ | $R_{t,4}$ | $R_{t,5}$ | $R_{t,RCT}$ |
|---|---|---|---|---|---|---|---|
| 11 | 0.030 | 0.026 | 0.026 | 0.028 | 0.033 | 0.062 | 0.120 |
| 21 | 0.026 | 0.026 | 0.026 | 0.026 | 0.030 | 0.043 | 0.120 |
| 31 | 0.025 | 0.024 | 0.025 | 0.025 | 0.027 | 0.035 | 0.119 |
| 41 | 0.024 | 0.023 | 0.023 | 0.024 | 0.025 | 0.031 | 0.118 |
| 50 | 0.025 | 0.024 | 0.024 | 0.024 | 0.025 | 0.029 | 0.120 |

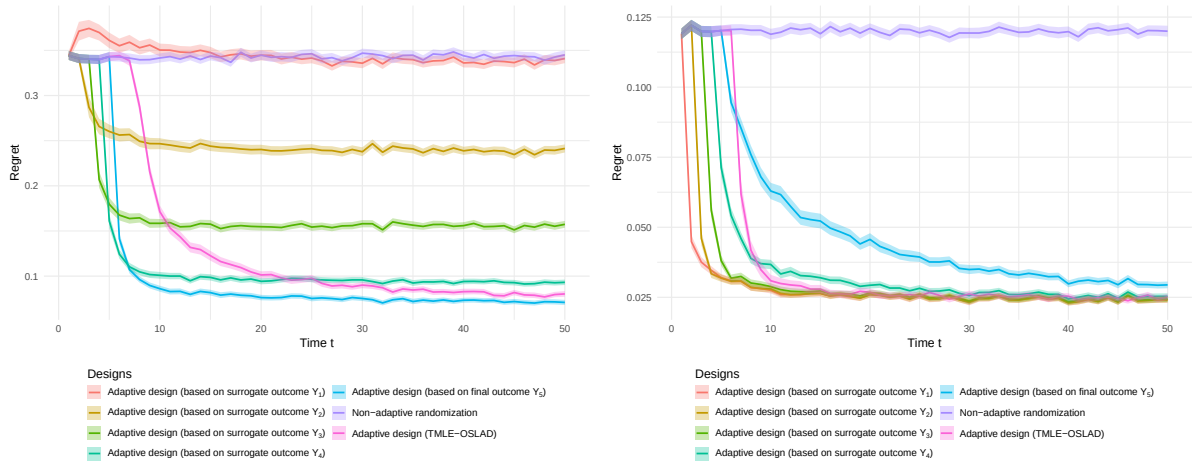



Figure 1.4: Regret at time $t$ for different design strategies in Scenario 1 (left) and Scenario 2 (right). This figure includes seven curves: five curves correspond to regrets from adaptive designs using outcomes from $Y_1$ to $Y_5$, respectively; one curve illustrates the regret for an adaptive design employing a TMLE-OSLAD to evaluate and utilize surrogates; and the last curve represents the regret of a fixed design, which assigns equal probabilities to each treatment group. The shaded regions represent twice the standard error of the regret across simulations.

Table 1.9 reports the true expected final outcome $Y_5$ obtained from implementing different adaptive designs at illustrative time points $t = 11, 21, 31, 41$, and 50, respectively. These values represent the ideal estimand as discussed in Section 1.8.1. We use $\tilde{\Psi}_{t,k}$ to denote the ideal estimand under the adaptive design guided by surrogate $Y_k$. We further use $\tilde{\Psi}_{t,RCT}$ and $\tilde{\Psi}_{t,SL}$ to denote the true expected final outcomes obtained from implementing the non-adaptive RCT and our proposed TMLE-OSLAD design, respectively. Comparing with the proposed context-specific estimands $\Psi_{t,k}$ in Table 1.1, the two estimands show minimal discrepancy, and the relative superiority between different surrogate-guided designs remains unchanged.

Table 1.8: Performance of TMLE estimators for the proposed adaptive design estimands at the end of experiments in Scenarios 1 and 2 across Monte Carlo simulations.

(a) Scenario 1

| Estimand | Notation | Truth | Bias ($\times 10^{-3}$) | Var ($\times 10^{-3}$) | Coverage (%) |
|---|---|---|---|---|---|
| Expected final outcome $Y_5$ under the adaptive design using $Y_1$ | $\Psi_{T_{end},1}$ | 0.000 | -0.88 | 2.20 | 94.0 |
| Expected final outcome $Y_5$ under the adaptive design using $Y_2$ | $\Psi_{T_{end},2}$ | 0.097 | -1.03 | 1.49 | 94.4 |
| Expected final outcome $Y_5$ under the adaptive design using $Y_3$ | $\Psi_{T_{end},3}$ | 0.175 | 0.01 | 1.06 | 93.6 |
| Expected final outcome $Y_5$ under the adaptive design using $Y_4$ | $\Psi_{T_{end},4}$ | 0.225 | 0.92 | 0.67 | 94.2 |
| Expected final outcome $Y_5$ under the adaptive design using $Y_5$ | $\Psi_{T_{end},5}$ | 0.238 | 0.06 | 0.59 | 94.6 |

(b) Scenario 2

| Estimand | Notation | Truth | Bias ($\times 10^{-3}$) | Var ($\times 10^{-3}$) | Coverage (%) |
|---|---|---|---|---|---|
| Expected Final Outcome $Y_5$ under the adaptive design using $Y_1$ | $\Psi_{T_{end},1}$ | 0.093 | -0.46 | 0.48 | 94.6 |
| Expected Final Outcome $Y_5$ under the adaptive design using $Y_2$ | $\Psi_{T_{end},2}$ | 0.091 | -0.70 | 0.50 | 94.4 |
| Expected Final Outcome $Y_5$ under the adaptive design using $Y_3$ | $\Psi_{T_{end},3}$ | 0.088 | -1.01 | 0.49 | 94.2 |
| Expected Final Outcome $Y_5$ under the adaptive design using $Y_4$ | $\Psi_{T_{end},4}$ | 0.083 | -0.62 | 0.52 | 94.6 |
| Expected Final Outcome $Y_5$ under the adaptive design using $Y_5$ | $\Psi_{T_{end},5}$ | 0.068 | -0.88 | 0.60 | 94.4 |

Table 1.9: True expected final outcome $\tilde{\Psi}_{t,k}$ of implementing different adaptive designs in Scenarios 1 and 2 ($t = 11, 21, 31, 41, 50, T_{end}$).

(a) Scenario 1

| $t$ | $\tilde{\Psi}_{t,RCT}$ | $\tilde{\Psi}_{t,1}$ | $\tilde{\Psi}_{t,2}$ | $\tilde{\Psi}_{t,3}$ | $\tilde{\Psi}_{t,4}$ | $\tilde{\Psi}_{t,5}$ | $\tilde{\Psi}_{t,SL}$ |
|---|---|---|---|---|---|---|---|
| 11 | 0.000 | -0.019 | 0.050 | 0.078 | 0.066 | 0.033 | 0.000 |
| 21 | 0.000 | -0.012 | 0.079 | 0.144 | 0.176 | 0.172 | 0.101 |
| 31 | 0.000 | -0.007 | 0.088 | 0.161 | 0.203 | 0.208 | 0.155 |
| 41 | 0.000 | -0.004 | 0.092 | 0.168 | 0.216 | 0.225 | 0.183 |
| 50 | 0.000 | -0.002 | 0.095 | 0.172 | 0.222 | 0.234 | 0.199 |
| $T_{end}$ | 0.000 | -0.001 | 0.096 | 0.174 | 0.225 | 0.238 | 0.205 |

(b) Scenario 2

| $t$ | $\tilde{\Psi}_{t,RCT}$ | $\tilde{\Psi}_{t,1}$ | $\tilde{\Psi}_{t,2}$ | $\tilde{\Psi}_{t,3}$ | $\tilde{\Psi}_{t,4}$ | $\tilde{\Psi}_{t,5}$ | $\tilde{\Psi}_{t,SL}$ |
|---|---|---|---|---|---|---|---|
| 11 | 0.000 | 0.070 | 0.057 | 0.039 | 0.019 | 0.004 | 0.000 |
| 21 | 0.000 | 0.085 | 0.079 | 0.072 | 0.060 | 0.038 | 0.054 |
| 31 | 0.000 | 0.089 | 0.085 | 0.081 | 0.072 | 0.053 | 0.069 |
| 41 | 0.000 | 0.090 | 0.088 | 0.085 | 0.078 | 0.062 | 0.076 |
| 50 | 0.000 | 0.091 | 0.090 | 0.087 | 0.081 | 0.067 | 0.080 |
| $T_{end}$ | 0.000 | 0.092 | 0.090 | 0.088 | 0.082 | 0.069 | 0.082 |

### 1.8.8 Sensitivity Analyses

We further conduct a series of sensitivity analyses to compare the performance of TMLE-OSLAD across different design configurations.

First, we perform sensitivity analyses to assess how TMLE-OSLAD's surrogate selection pattern changes with different $\alpha$'s used in (i) the $100(1-\alpha)\%$ TMLE-based Wald-type confidence interval for evaluating and choosing from candidate designs, and (ii) the $100(1-\alpha)\%$ CATE confidence intervals for generating treatment randomization probabilities. Figures 1.5 and 1.6 report the selection frequency of each surrogate-guided design at each time point when the TMLE-based selection criterion, originally based on 95% CIs, is re-evaluated using 90% and 99% confidence intervals. The selection patterns remain consistent across these choices.

Figures 1.7 and 1.8 report the selection frequency of each candidate surrogate-driven design obtained by implementing adaptive designs where treatment-randomization proba-

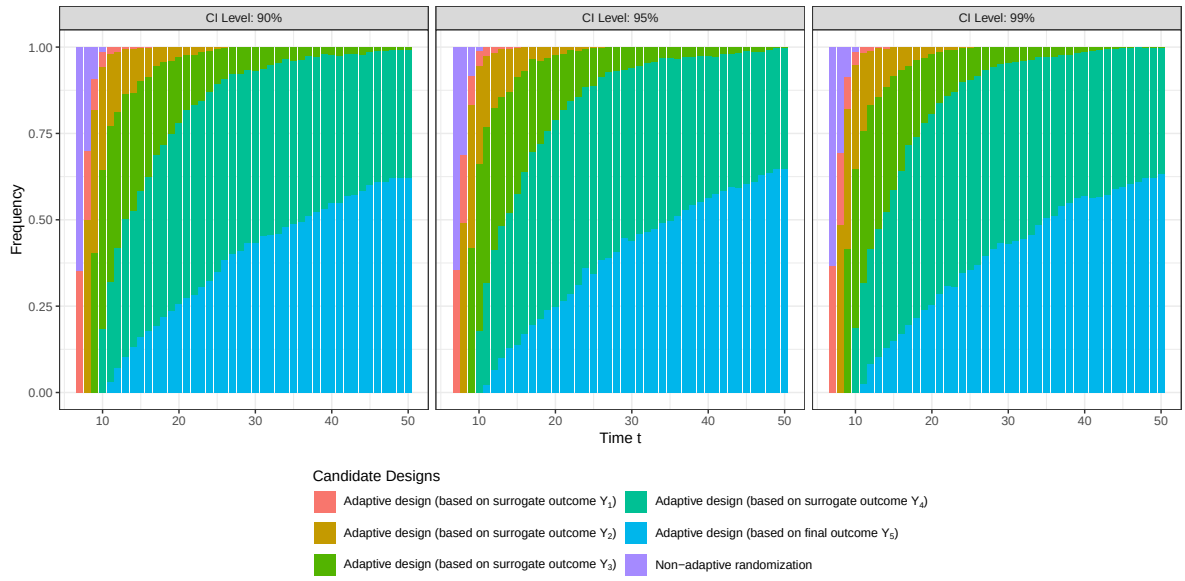



Figure 1.5: Frequency of candidate designs selected by TMLE-OSLAD using different TMLE confidence interval levels in Scenario 1.

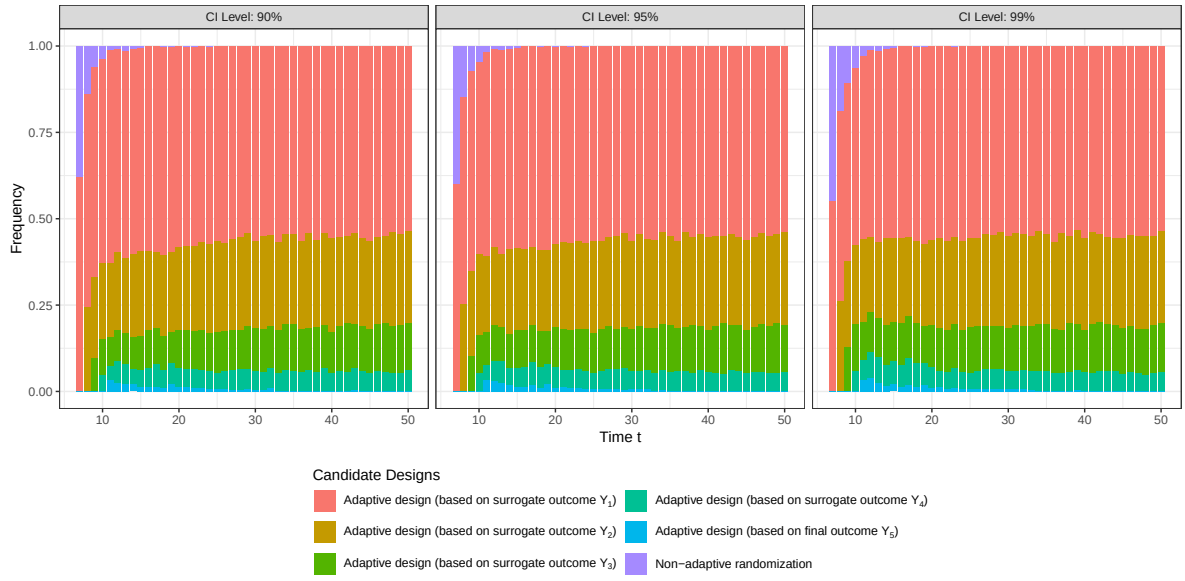



Figure 1.6: Frequency of candidate designs selected by TMLE-OSLAD using different TMLE confidence interval levels in Scenario 2.

bilities are based upon $100(1-\alpha)\%$ CATE confidence intervals with $\alpha \in \{0.01, 0.05, 0.10\}$ for Scenarios 1 and 2, respectively. The selection patterns are consistent across different CI levels.

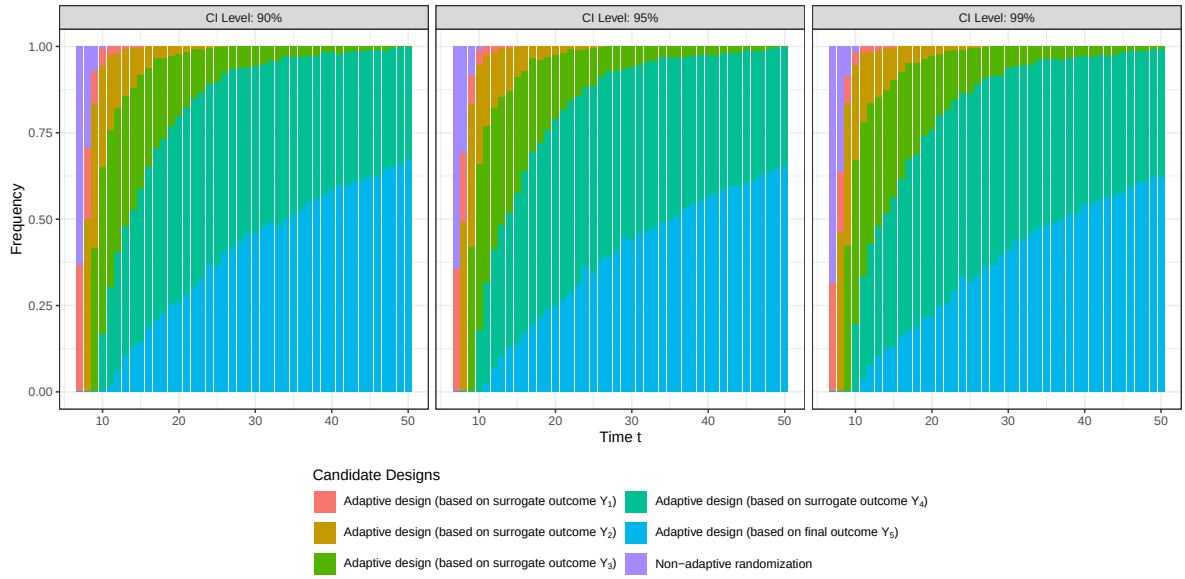



Figure 1.7: Frequency of candidate designs selected by TMLE-OSLAD using different levels of CATE confidence interval to adapt treatment randomization probabilities in Scenario 1.

Second, we conduct a sensitivity analysis on our proposed TMLE-OSLAD design with a different choice of $h_\nu$. As noted in Section 1.8.4, the treatment randomization probabilities we use are

$$h_\nu(z) = \begin{cases} \nu, & z \leq -1 \\ \nu + (1-2\nu)\left(-\frac{1}{4}z^3 + \frac{3}{4}z + \frac{1}{2}\right), & -1 < z < 1\,, \\ 1-\nu, & z \geq 1 \end{cases} \tag{1.1}$$

where the input $z$ is the ratio of the CATE estimate to half the width of its $100(1-\alpha)\%$ confidence interval. Let $f^{(0)}$ denote the function used in the second term of $h_\nu$ for $-1 < z < 1$, defined as $f^{(0)}(z) = -\frac{1}{4}z^3 + \frac{3}{4}z + \frac{1}{2}$. To examine the sensitivity of our proposed methods to different treatment randomization schemes, we consider three additional functions that vary in how aggressively the estimated CATE signal is transformed into a skewed treatment randomization probability favoring the estimated optimal personalized treatment. These functions are defined as follows: (i) $f^{(1)}(z) = 1/(1+\exp(-8z))$, (ii) $f^{(2)}(z) = \frac{1}{4}z^3 + \frac{1}{4}z + \frac{1}{2}$, and (iii) $f^{(3)}(z) = -\frac{1}{4}z^5 + \frac{3}{4}z^3 + \frac{1}{2}$. To facilitate comparison, we use $h_\nu^{(0)}$ to denote the baseline mapping function $h_\nu$ in the main text. We then define $h_\nu^{(1)}(z)$, $h_\nu^{(2)}(z)$, and $h_\nu^{(3)}(z)$

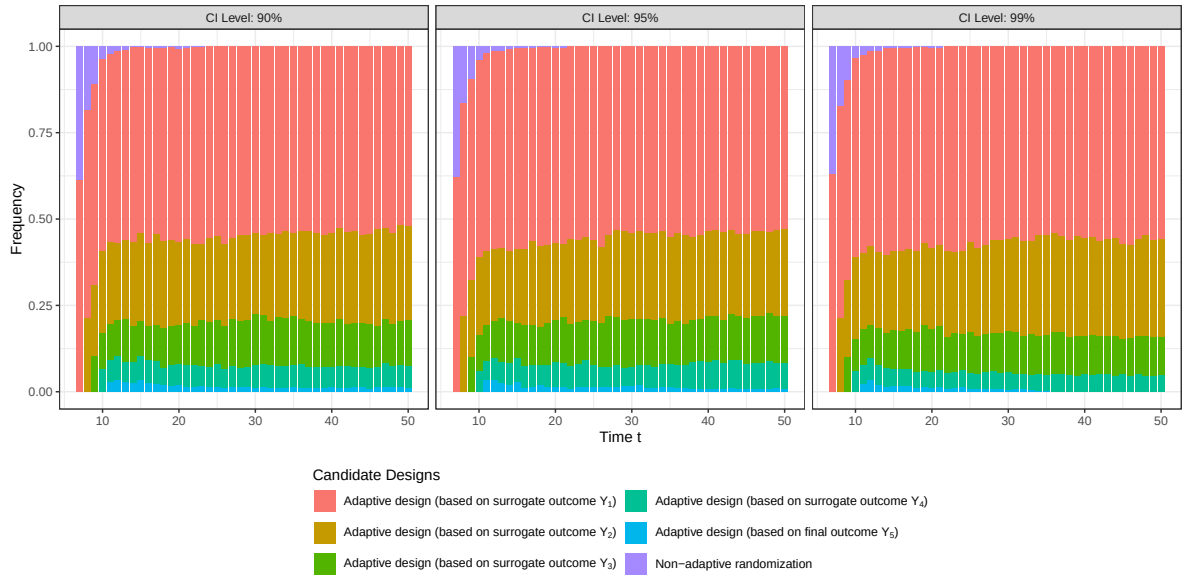



Figure 1.8: Frequency of candidate designs selected by TMLE-OSLAD using different levels of CATE confidence interval to adapt treatment randomization probabilities in Scenario 2.

as alternative mapping functions that replace $f^{(0)}(z)$ with $f^{(1)}(z)$, $f^{(2)}(z)$, and $f^{(3)}(z)$ for $-1 < z < 1$, respectively. Figure 1.9 compares the four functions. Relative to $h_\nu^{(0)}$, $h_\nu^{(1)}$ shifts the treatment randomization probability toward the estimated optimal treatment more rapidly as CATE increases, while $h_\nu^{(2)}$ adjusts more gradually, and $h_\nu^{(3)}$ exhibits the slowest transition.

Tables 1.10 and 1.11 report the true target estimand $\Psi_{t,k}$ for each surrogate-specific adaptive design under the different mapping functions $h_\nu^{(0)}$ through $h_\nu^{(3)}$. The oracle surrogate that yields the highest target estimand is consistent across different mapping functions. Figures 1.10 and 1.11 further show the selection frequency for each candidate design by TMLE-OSLAD across different mapping functions used for treatment randomization. The surrogate selection pattern is robust to the choice of the function, with TMLE-OSLAD consistently identifying and increasingly favoring the oracle surrogate as the experiment progresses. Moreover, TMLE provides robust inference for these target estimands, with confidence intervals achieving nominal coverage across simulations using different mapping functions (Tables 1.12 and 1.13).

Lastly, we extend the simulation scenarios to include three covariates to further evaluate the performance of our proposed methods. In Scenario 1, the conditional mean functions of $Y_k$ given $A$ and $W$ are

$$E[Y_k|A, W_1, W_2, W_3] = (2A-1) \times \left[0.5 - (1 + \exp(-(3-k) - 0.4W_1 - 0.4W_2 - 0.2W_3))^{-1}\right].$$

In Scenario 2, the conditional mean functions of $Y_k$ are $E[Y_k|A, W_1, W_2, W_3] = (2A-1) \times$

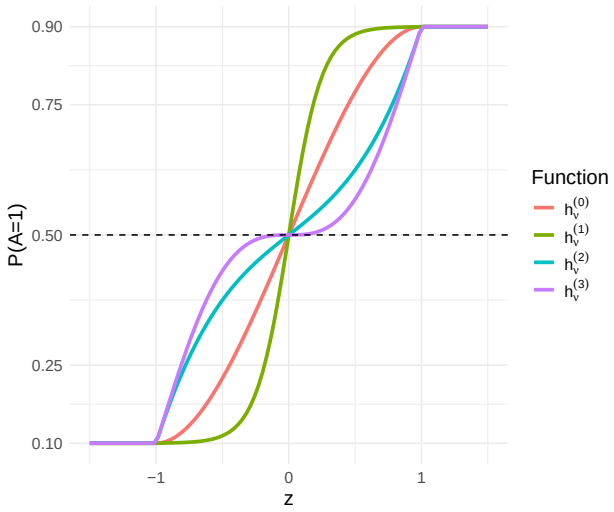



Figure 1.9: Comparison of the baseline mapping $h_\nu^{(0)}$ and three alternatives $h_\nu^{(1)}$, $h_\nu^{(2)}$, and $h_\nu^{(3)}$ ($\nu = 0.1$) for sensitivity analysis. The x-axis shows the standardized CATE ($z$), and the y-axis shows the probability of assigning treatment $a = 1$ according to each mapping function (shown as curves).

$[0.5 - (1 + \exp(-\gamma_k \times (0.4W_1 + 0.4W_2 + 0.2W_3)))^{-1}]$, where $\gamma_1 = 3$,$\gamma_2 = 2$,$\gamma_3 = 1$,$\gamma_4 = 0.5$ and $\gamma_5 = 0.25$. $W_1$, $W_2$, and $W_3$ are independent and all follow a uniform distribution $U(-4, 4)$. Let $V := 0.4W_1+0.4W_2+0.2W_3$. The CATE functions for the candidate outcomes $Y_1, \ldots, Y_5$ as functions of $V$ follow the same patterns shown in Figure 1.1.

Table 1.14 reports the true target estimand $\Psi_{t,k}$. Table 1.15 shows the coverage probabilities of TMLE-based confidence intervals for these estimands. Figure 1.12 plots the probability of each surrogate-driven design being selected over time by TMLE-OSLAD in the two scenarios. Similar to the simulation results in the main paper, TMLE-OSLAD consistently provides valid statistical inference with confidence intervals that achieve nominal coverage and selects the oracle surrogate with high frequency in both scenarios.

Table 1.10: True target estimand $\Psi_{t,k}$ if participants enrolled from $1, \cdots, t-K$ had been adaptively treated by a candidate surrogate $Y_k$ in Scenarios 1 ($t = 11, 21, 31, 41, 50$) under different mapping functions $h_\nu^{(0)}$ to $h_\nu^{(3)}$. Expected outcomes under the adaptive design using the oracle surrogate are shown in bold.

| | $h_\nu^{(0)}$ | | | | | $h_\nu^{(1)}$ | | | | | $h_\nu^{(2)}$ | | | | | $h_\nu^{(3)}$ | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $t$ | $\Psi_{t,RCT}$ | $\Psi_{t,1}$ | $\Psi_{t,2}$ | $\Psi_{t,3}$ | $\Psi_{t,4}$ | $\Psi_{t,5}$ | $\Psi_{t,RCT}$ | $\Psi_{t,1}$ | $\Psi_{t,2}$ | $\Psi_{t,3}$ | $\Psi_{t,4}$ | $\Psi_{t,5}$ | $\Psi_{t,RCT}$ | $\Psi_{t,1}$ | $\Psi_{t,2}$ | $\Psi_{t,3}$ | $\Psi_{t,4}$ | $\Psi_{t,5}$ | $\Psi_{t,RCT}$ | $\Psi_{t,1}$ | $\Psi_{t,2}$ | $\Psi_{t,3}$ | $\Psi_{t,4}$ | $\Psi_{t,5}$ |
| 11 | 0.000 | -0.016 | 0.053 | **0.078** | 0.066 | 0.033 | 0.000 | -0.013 | 0.058 | **0.083** | 0.071 | 0.036 | 0.000 | -0.015 | 0.047 | **0.070** | 0.060 | 0.030 | 0.000 | -0.015 | 0.045 | **0.068** | 0.058 | 0.029 |
| 21 | 0.000 | -0.008 | 0.083 | 0.145 | **0.177** | 0.173 | 0.000 | -0.004 | 0.086 | 0.148 | **0.181** | 0.176 | 0.000 | -0.009 | 0.079 | 0.140 | **0.171** | 0.166 | 0.000 | -0.010 | 0.078 | 0.138 | **0.169** | 0.164 |
| 31 | 0.000 | -0.005 | 0.091 | 0.161 | 0.203 | **0.208** | 0.000 | -0.002 | 0.093 | 0.163 | 0.207 | **0.211** | 0.000 | -0.007 | 0.088 | 0.158 | 0.199 | **0.202** | 0.000 | -0.008 | 0.087 | 0.156 | 0.197 | **0.201** |
| 41 | 0.000 | -0.002 | 0.094 | 0.169 | 0.215 | **0.225** | 0.000 | 0.000 | 0.096 | 0.170 | 0.218 | **0.228** | 0.000 | -0.004 | 0.092 | 0.166 | 0.212 | **0.220** | 0.000 | -0.006 | 0.091 | 0.165 | 0.211 | **0.218** |
| 50 | 0.000 | -0.001 | 0.096 | 0.173 | 0.222 | **0.234** | 0.000 | 0.001 | 0.098 | 0.174 | 0.225 | **0.236** | 0.000 | -0.003 | 0.094 | 0.170 | 0.219 | **0.229** | 0.000 | -0.004 | 0.094 | 0.170 | 0.218 | **0.228** |

Table 1.11: True target estimand $\Psi_{t,k}$ if participants enrolled from $1, \cdots, t-K$ had been adaptively treated by a candidate surrogate $Y_k$ in Scenarios 2 ($t = 11, 21, 31, 41, 50$) under different mapping functions $h_\nu^{(0)}$ to $h_\nu^{(3)}$. Expected outcomes under the adaptive design using the oracle surrogate are shown in bold.

| | $h_\nu^{(0)}$ | | | | | $h_\nu^{(1)}$ | | | | | $h_\nu^{(2)}$ | | | | | $h_\nu^{(3)}$ | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $t$ | $\Psi_{t,RCT}$ | $\Psi_{t,1}$ | $\Psi_{t,2}$ | $\Psi_{t,3}$ | $\Psi_{t,4}$ | $\Psi_{t,5}$ | $\Psi_{t,RCT}$ | $\Psi_{t,1}$ | $\Psi_{t,2}$ | $\Psi_{t,3}$ | $\Psi_{t,4}$ | $\Psi_{t,5}$ | $\Psi_{t,RCT}$ | $\Psi_{t,1}$ | $\Psi_{t,2}$ | $\Psi_{t,3}$ | $\Psi_{t,4}$ | $\Psi_{t,5}$ | $\Psi_{t,RCT}$ | $\Psi_{t,1}$ | $\Psi_{t,2}$ | $\Psi_{t,3}$ | $\Psi_{t,4}$ | $\Psi_{t,5}$ |
| 11 | 0.000 | **0.073** | 0.057 | 0.039 | 0.019 | 0.004 | 0.000 | **0.075** | 0.060 | 0.042 | 0.021 | 0.005 | 0.000 | **0.069** | 0.054 | 0.035 | 0.015 | 0.003 | 0.000 | **0.068** | 0.053 | 0.034 | 0.014 | 0.003 |
| 21 | 0.000 | **0.087** | 0.080 | 0.073 | 0.061 | 0.038 | 0.000 | **0.088** | 0.082 | 0.075 | 0.064 | 0.042 | 0.000 | **0.084** | 0.078 | 0.070 | 0.056 | 0.031 | 0.000 | **0.084** | 0.078 | 0.070 | 0.054 | 0.028 |
| 31 | 0.000 | **0.090** | 0.086 | 0.081 | 0.073 | 0.052 | 0.000 | **0.091** | 0.087 | 0.082 | 0.075 | 0.057 | 0.000 | **0.088** | 0.085 | 0.079 | 0.068 | 0.043 | 0.000 | **0.088** | 0.084 | 0.079 | 0.067 | 0.041 |
| 41 | 0.000 | **0.091** | 0.089 | 0.085 | 0.078 | 0.060 | 0.000 | **0.092** | 0.089 | 0.086 | 0.081 | 0.065 | 0.000 | **0.090** | 0.087 | 0.084 | 0.074 | 0.051 | 0.000 | **0.090** | 0.087 | 0.083 | 0.073 | 0.049 |
| 50 | 0.000 | **0.092** | 0.090 | 0.087 | 0.081 | 0.065 | 0.000 | **0.093** | 0.091 | 0.088 | 0.083 | 0.070 | 0.000 | **0.091** | 0.089 | 0.086 | 0.078 | 0.057 | 0.000 | **0.091** | 0.089 | 0.085 | 0.077 | 0.054 |

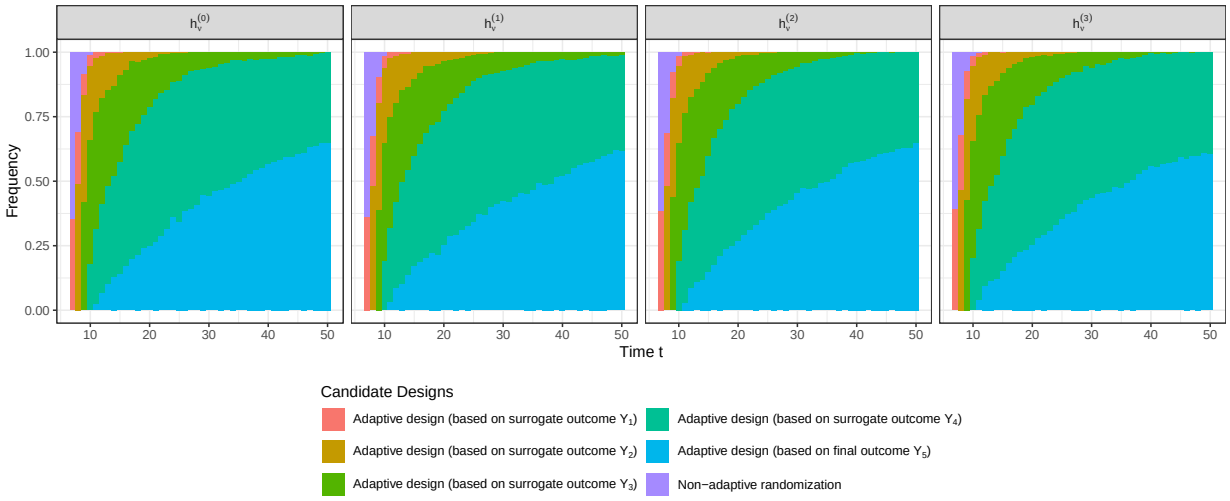



Figure 1.10: Frequency of candidate designs selected by TMLE-OSLAD under different treatment randomization functions applied in Scenario 1.

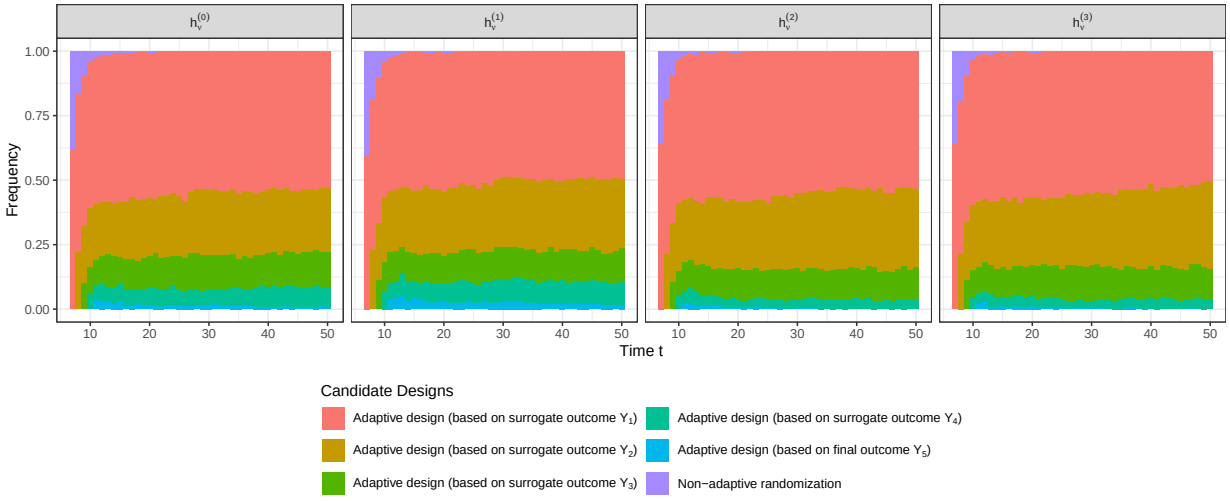



Figure 1.11: Frequency of candidate designs selected by TMLE-OSLAD under different treatment randomization functions applied in Scenario 2.

Table 1.12: Coverage probability (%) of confidence interval for target estimands $\Psi_{t,k}$ in Scenarios 1 ($t = 11, 21, 31, 41, 50$) under different mapping functions $h_\nu^{(0)}$ to $h_\nu^{(3)}$.

| | $h_\nu^{(0)}$ | | | | | $h_\nu^{(1)}$ | | | | | $h_\nu^{(2)}$ | | | | | $h_\nu^{(3)}$ | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $t$ | $\Psi_{t,RCT}$ | $\Psi_{t,1}$ | $\Psi_{t,2}$ | $\Psi_{t,3}$ | $\Psi_{t,4}$ | $\Psi_{t,5}$ | $\Psi_{t,RCT}$ | $\Psi_{t,1}$ | $\Psi_{t,2}$ | $\Psi_{t,3}$ | $\Psi_{t,4}$ | $\Psi_{t,5}$ | $\Psi_{t,RCT}$ | $\Psi_{t,1}$ | $\Psi_{t,2}$ | $\Psi_{t,3}$ | $\Psi_{t,4}$ | $\Psi_{t,5}$ | $\Psi_{t,RCT}$ | $\Psi_{t,1}$ | $\Psi_{t,2}$ | $\Psi_{t,3}$ | $\Psi_{t,4}$ | $\Psi_{t,5}$ |
| 11 | 96.4 | 96.2 | 95.8 | 95.6 | 95.6 | 96.2 | 96.8 | 96.2 | 95.0 | 95.4 | 95.8 | 96.4 | 96.8 | 95.8 | 95.8 | 95.6 | 95.8 | 96.6 | 96.8 | 95.6 | 95.8 | 95.4 | 96.0 | 96.6 |
| 21 | 95.0 | 93.8 | 93.4 | 94.6 | 94.2 | 95.2 | 94.2 | 95.0 | 93.6 | 94.4 | 92.4 | 94.8 | 93.2 | 93.6 | 92.0 | 94.2 | 92.2 | 94.8 | 93.2 | 94.0 | 92.4 | 93.8 | 93.2 | 94.4 |
| 31 | 96.2 | 95.2 | 94.8 | 94.2 | 95.2 | 95.0 | 94.4 | 94.6 | 94.2 | 93.0 | 94.0 | 95.6 | 95.4 | 94.0 | 93.2 | 94.0 | 94.4 | 94.0 | 96.6 | 94.8 | 93.4 | 94.0 | 94.6 | 94.8 |
| 41 | 95.6 | 95.6 | 97.2 | 94.4 | 94.6 | 95.8 | 96.2 | 96.8 | 96.6 | 95.4 | 95.6 | 95.0 | 95.4 | 95.4 | 95.4 | 95.4 | 95.2 | 95.6 | 95.8 | 96.2 | 96.8 | 95.4 | 94.4 | 95.8 |
| 50 | 94.4 | 94.2 | 95.0 | 93.6 | 94.0 | 94.8 | 95.8 | 95.2 | 94.2 | 93.4 | 94.4 | 94.6 | 95.8 | 93.6 | 93.8 | 95.0 | 95.0 | 95.0 | 95.8 | 95.2 | 95.8 | 95.0 | 94.4 | 95.2 |

Table 1.13: Coverage probability (%) of confidence interval for target estimands $\Psi_{t,k}$ in Scenarios 2 ($t = 11, 21, 31, 41, 50$) under different mapping functions $h_\nu^{(0)}$ to $h_\nu^{(3)}$.

| | $h_\nu^{(0)}$ | | | | | $h_\nu^{(1)}$ | | | | | $h_\nu^{(2)}$ | | | | | $h_\nu^{(3)}$ | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $t$ | $\Psi_{t,RCT}$ | $\Psi_{t,1}$ | $\Psi_{t,2}$ | $\Psi_{t,3}$ | $\Psi_{t,4}$ | $\Psi_{t,5}$ | $\Psi_{t,RCT}$ | $\Psi_{t,1}$ | $\Psi_{t,2}$ | $\Psi_{t,3}$ | $\Psi_{t,4}$ | $\Psi_{t,5}$ | $\Psi_{t,RCT}$ | $\Psi_{t,1}$ | $\Psi_{t,2}$ | $\Psi_{t,3}$ | $\Psi_{t,4}$ | $\Psi_{t,5}$ | $\Psi_{t,RCT}$ | $\Psi_{t,1}$ | $\Psi_{t,2}$ | $\Psi_{t,3}$ | $\Psi_{t,4}$ | $\Psi_{t,5}$ |
| 11 | 96.2 | 95.2 | 94.8 | 95.0 | 95.4 | 94.6 | 95.8 | 96.0 | 95.2 | 95.0 | 95.6 | 94.0 | 95.8 | 94.8 | 95.0 | 95.6 | 95.4 | 95.4 | 95.8 | 94.8 | 94.8 | 95.6 | 95.4 | 95.4 |
| 21 | 94.0 | 95.2 | 95.2 | 94.4 | 94.4 | 95.0 | 93.6 | 94.4 | 95.0 | 95.2 | 94.0 | 95.4 | 93.6 | 94.8 | 95.8 | 95.6 | 94.2 | 95.2 | 94.0 | 96.2 | 96.4 | 95.2 | 95.2 | 95.6 |
| 31 | 94.4 | 94.2 | 94.2 | 94.8 | 94.0 | 95.4 | 94.4 | 94.6 | 93.4 | 93.2 | 94.2 | 94.8 | 94.0 | 94.0 | 93.8 | 94.4 | 94.0 | 95.0 | 95.6 | 94.0 | 93.6 | 95.0 | 93.4 | 94.6 |
| 41 | 94.8 | 95.0 | 94.4 | 94.0 | 95.4 | 95.4 | 93.6 | 94.0 | 94.4 | 93.2 | 93.8 | 96.0 | 93.6 | 93.8 | 93.2 | 94.2 | 94.2 | 95.4 | 94.2 | 94.4 | 94.0 | 94.0 | 94.4 | 94.6 |
| 50 | 94.6 | 94.2 | 94.8 | 94.2 | 94.6 | 94.8 | 94.4 | 93.6 | 95.2 | 94.0 | 94.6 | 95.6 | 94.4 | 94.4 | 93.4 | 93.6 | 95.0 | 94.8 | 94.8 | 93.8 | 93.8 | 93.0 | 94.2 | 94.8 |

Table 1.14: True target estimand $\Psi_{t,k}$ if participants enrolled from $1, \cdots, t-K$ had been adaptively treated by a candidate surrogate $Y_k$ in Scenarios 1 and 2 ($t = 11, 21, 31, 41, 50$) with three covariates. Expected outcomes under the adaptive design using the oracle surrogate are shown in bold.

(a) Scenario 1

| $t$ | $\Psi_{t,RCT}$ | $\Psi_{t,1}$ | $\Psi_{t,2}$ | $\Psi_{t,3}$ | $\Psi_{t,4}$ | $\Psi_{t,5}$ |
|---|---|---|---|---|---|---|
| 11 | 0.000 | -0.174 | -0.066 | 0.035 | **0.060** | 0.036 |
| 21 | 0.000 | -0.207 | -0.078 | 0.088 | 0.162 | **0.170** |
| 31 | 0.000 | -0.214 | -0.080 | 0.103 | 0.189 | **0.204** |
| 41 | 0.000 | -0.217 | -0.081 | 0.109 | 0.202 | **0.219** |
| 50 | 0.000 | -0.218 | -0.080 | 0.112 | 0.209 | **0.228** |

(b) Scenario 2

| $t$ | $\Psi_{t,RCT}$ | $\Psi_{t,1}$ | $\Psi_{t,2}$ | $\Psi_{t,3}$ | $\Psi_{t,4}$ | $\Psi_{t,5}$ |
|---|---|---|---|---|---|---|
| 11 | 0.000 | **0.034** | 0.025 | 0.014 | 0.005 | 0.001 |
| 21 | 0.000 | **0.044** | 0.040 | 0.033 | 0.021 | 0.008 |
| 31 | 0.000 | **0.047** | 0.044 | 0.039 | 0.028 | 0.013 |
| 41 | 0.000 | **0.048** | 0.046 | 0.042 | 0.033 | 0.016 |
| 50 | 0.000 | **0.049** | 0.047 | 0.044 | 0.035 | 0.018 |

Table 1.15: Coverage probability (%) of confidence intervals for $\Psi_{t,k}$ in Scenarios 1 and 2 ($t = 11, 21, 31, 41, 50$) with three covariates.

(a) Scenario 1

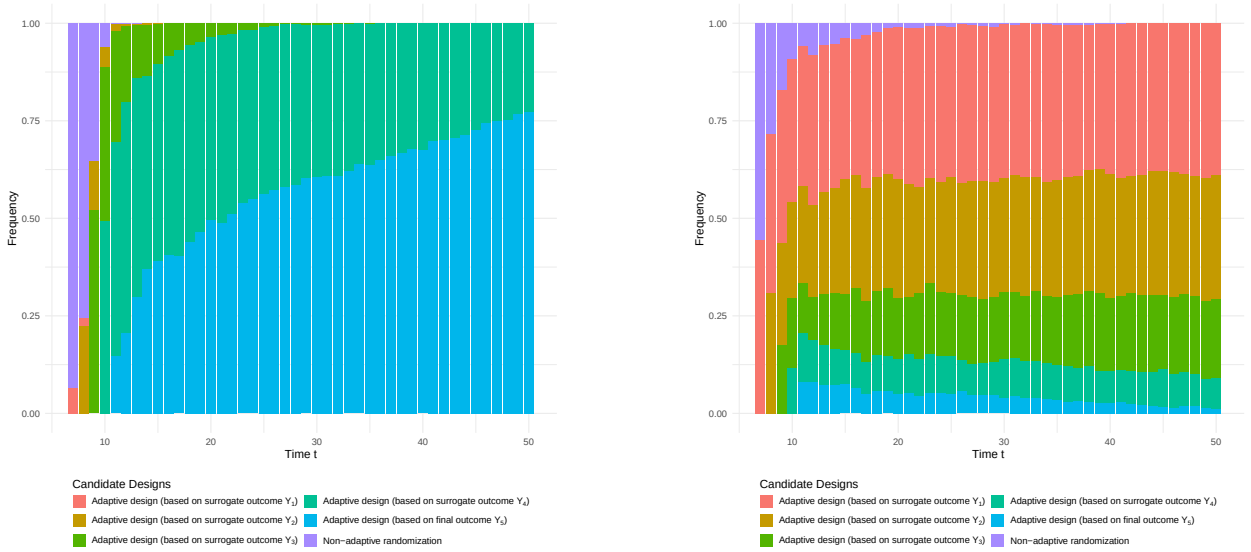
| $t$ | $\Psi_{t,RCT}$ | $\Psi_{t,1}$ | $\Psi_{t,2}$ | $\Psi_{t,3}$ | $\Psi_{t,4}$ | $\Psi_{t,5}$ |
|---|---|---|---|---|---|---|
| 11 | 93.4 | 95.0 | 93.6 | 93.6 | 94.0 | 94.0 |
| 21 | 92.4 | 93.4 | 92.8 | 94.0 | 93.0 | 93.6 |
| 31 | 94.6 | 93.2 | 94.0 | 95.0 | 95.4 | 94.0 |
| 41 | 94.2 | 95.2 | 94.8 | 94.4 | 94.2 | 94.6 |
| 50 | 94.2 | 93.6 | 94.0 | 93.0 | 94.0 | 94.4 |

(b) Scenario 2

| $t$ | $\Psi_{t,RCT}$ | $\Psi_{t,1}$ | $\Psi_{t,2}$ | $\Psi_{t,3}$ | $\Psi_{t,4}$ | $\Psi_{t,5}$ |
|---|---|---|---|---|---|---|
| 11 | 93.2 | 93.4 | 93.2 | 93.4 | 93.4 | 94.0 |
| 21 | 93.6 | 91.8 | 94.2 | 94.4 | 93.6 | 93.6 |
| 31 | 94.4 | 91.8 | 92.4 | 93.0 | 93.8 | 94.0 |
| 41 | 95.0 | 94.4 | 94.2 | 93.4 | 95.0 | 95.0 |
| 50 | 93.8 | 93.6 | 93.0 | 93.0 | 94.0 | 95.2 |

### 1.8.9 Additional Results for Section 1.6

In this section, we present additional results in ADAPT-R simulation. Figure 1.13 shows the true CATE functions of Navigator vs. SMS+CCT in ADAPT-R simulation. For participants initially assigned to SMS, the optimal personalized treatment varies with time-to-lapse after initial treatment and shows different patterns across outcomes. For earlier outcomes ($Y_1$ through $Y_3$), the CATE suggests that Navigator is preferable with shorter time-to-lapse, but the benefit wanes over time, shifting the optimal treatment to SMS+CCT. In contrast, for later outcomes ($Y_4$ and $Y_5$), SMS+CCT is preferred for short time-to-lapse, with their effectiveness decreasing over time, making Navigator the optimal choice for longer time-to-lapse. For participants whose initial treatment was SOC, the treatment effects are minimal: CATE remains close to zero over the entire range of time-to-lapse periods, indicating small differences between Navigator and SMS+CCT. For participants with CCT as their initial strategy, Navigator is preferred for participants with short time-to-lapse periods, showing a positive CATE. However, as time-to-lapse increases, SMS+CCT becomes increasingly favored, with the CATE decreasing and turning negative.




(a) (b)

Figure 1.12: Frequency of candidate adaptive design selected by TMLE-OSLAD at each time point. Panels (a) and (b) correspond to Scenario 1 and Scenario 2 with three covariates.

Figure 1.14 illustrates regret over time obtained from implementing different design strategies. Table 1.16 shows the true values of the proposed adaptive design target estimand $\Psi_{t,k}$ at the illustrative time points 11, 13, 15, 17, and 19, along with the coverage probabilities of the confidence intervals based on TMLE. Table 1.17 reports the bias and variance of TMLE at these time points. Table 1.18 presents TMLE's performance in estimating the proposed adaptive design estimands at the end of the experiment. Table 1.19 presents TMLE's performance of estimating ATE on final outcome and mean final outcome of optimal dynamic treatment rules based on surrogate outcomes and final outcomes after the experiment ends. Table 1.20 shows the true expected final outcome $\tilde{\Psi}_{t,k}$ of implementing different designs in ADAPT-R simulation.

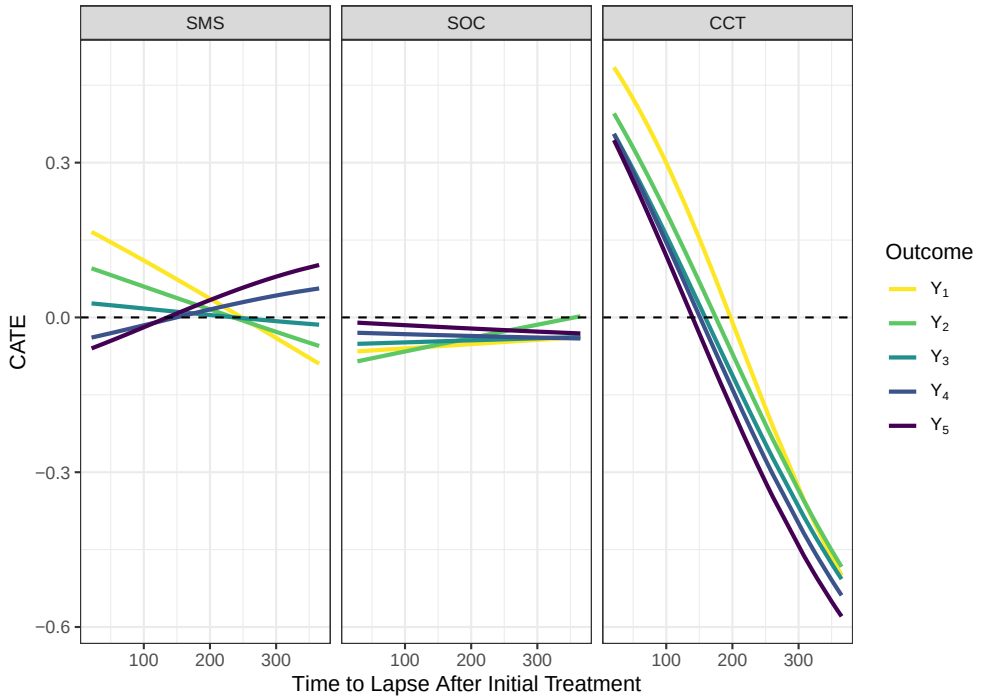



Figure 1.13: Estimated Conditional Average Treatment Effect (CATE) functions of Navigator vs. SMS+CCT in Phase 2, plotted against initial retention strategy (grids) and time to lapse after initial strategy in Phase 1 (x-axis). Each curve corresponds to the CATE function of a candidate surrogate, ranging from $Y_1$ to $Y_5$.

Table 1.16: (a) True target estimand $\Psi_{t,k}$ if participants enrolled from $1, \cdots, t-K$ had been adaptively treated by a candidate adaptive design using surrogate $Y_k$ in ADAPT-R trial simulations ($t = 11, 13, 15, 17, 19$). (b) Coverage probability (%) of confidence intervals for $\Psi_{t,k}$ in ADAPT-R trial simulations ($t = 11, 13, 15, 17, 19$).

(a) $\Psi_{t,k}$

| $t$ | $\Psi_{t,RCT}$ | $\Psi_{t,1}$ | $\Psi_{t,2}$ | $\Psi_{t,3}$ | $\Psi_{t,4}$ | $\Psi_{t,5}$ |
|---|---|---|---|---|---|---|
| 11 | 0.538 | 0.542 | 0.540 | 0.538 | 0.538 | 0.538 |
| 13 | 0.547 | 0.559 | 0.555 | 0.552 | 0.549 | 0.547 |
| 15 | 0.557 | 0.572 | 0.569 | 0.566 | 0.563 | 0.560 |
| 17 | 0.562 | 0.579 | 0.577 | 0.574 | 0.571 | 0.568 |
| 19 | 0.564 | 0.580 | 0.578 | 0.576 | 0.573 | 0.570 |

(b) Coverage probability (%)

| $t$ | $\Psi_{t,RCT}$ | $\Psi_{t,1}$ | $\Psi_{t,2}$ | $\Psi_{t,3}$ | $\Psi_{t,4}$ | $\Psi_{t,5}$ |
|---|---|---|---|---|---|---|
| 11 | 94.8 | 95.6 | 95.8 | 95.0 | 94.6 | 94.8 |
| 13 | 94.2 | 95.8 | 95.8 | 93.0 | 93.6 | 94.2 |
| 15 | 92.2 | 96.2 | 95.4 | 95.4 | 92.4 | 93.2 |
| 17 | 93.8 | 96.8 | 95.6 | 95.2 | 94.2 | 95.6 |
| 19 | 93.2 | 96.8 | 95.2 | 95.4 | 95.2 | 95.6 |

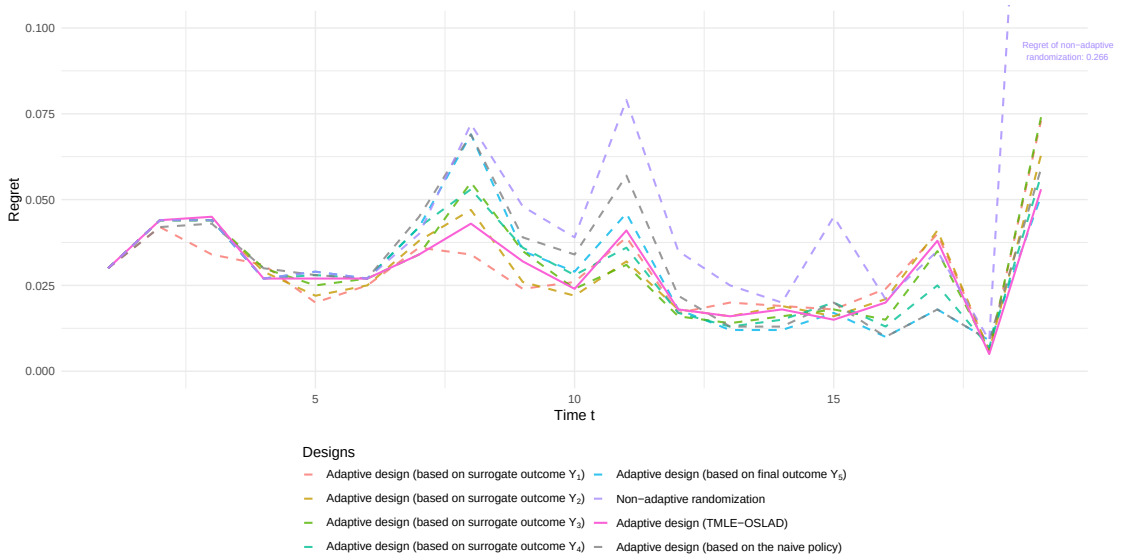



Figure 1.14: Regret at each time point $t$ obtained from implementing different design strategies in the ADAPT-R trial simulation.

Table 1.17: Bias and variance of TMLE estimators for $\Psi_{t,k}$ in ADAPT-R trial simulations ($t = 11, 13, 15, 17, 19$).

(a) Bias ($\times 10^{-3}$)

| $t$ | $\Psi_{t,RCT}$ | $\Psi_{t,1}$ | $\Psi_{t,2}$ | $\Psi_{t,3}$ | $\Psi_{t,4}$ | $\Psi_{t,5}$ |
|---|---|---|---|---|---|---|
| 11 | -0.50 | -0.40 | -0.30 | 0.22 | -0.54 | -0.31 |
| 13 | 0.45 | -0.16 | 0.18 | 0.88 | 0.36 | 0.33 |
| 15 | 0.56 | -0.11 | 0.29 | 0.52 | 0.15 | 0.10 |
| 17 | 0.29 | -0.02 | 0.09 | 0.37 | -0.16 | -0.12 |
| 19 | 0.16 | -0.08 | -0.04 | 0.23 | -0.22 | -0.12 |

(b) Variance ($\times 10^{-3}$)

| $t$ | $\Psi_{t,RCT}$ | $\Psi_{t,1}$ | $\Psi_{t,2}$ | $\Psi_{t,3}$ | $\Psi_{t,4}$ | $\Psi_{t,5}$ |
|---|---|---|---|---|---|---|
| 11 | 0.16 | 0.22 | 0.22 | 0.20 | 0.20 | 0.17 |
| 13 | 0.16 | 0.16 | 0.19 | 0.19 | 0.20 | 0.22 |
| 15 | 0.14 | 0.13 | 0.15 | 0.16 | 0.18 | 0.19 |
| 17 | 0.12 | 0.10 | 0.12 | 0.13 | 0.14 | 0.17 |
| 19 | 0.11 | 0.10 | 0.11 | 0.12 | 0.13 | 0.16 |

Table 1.18: Performance of TMLE estimators for adaptive-design estimands at the end of experiments in ADAPT-R trial simulation.

| Estimand | Notation | Truth | Bias ($\times 10^{-3}$) | Var ($\times 10^{-3}$) | Coverage (%) |
|---|---|---|---|---|---|
| Expected final outcome $Y_5$ under the adaptive design using $Y_1$ | $\Psi_{T_{end},1}$ | 0.581 | -0.11 | 0.09 | 96.2 |
| Expected final outcome $Y_5$ under the adaptive design using $Y_2$ | $\Psi_{T_{end},2}$ | 0.580 | -0.05 | 0.10 | 95.8 |
| Expected final outcome $Y_5$ under the adaptive design using $Y_3$ | $\Psi_{T_{end},3}$ | 0.578 | 0.29 | 0.11 | 95.0 |
| Expected final outcome $Y_5$ under the adaptive design using $Y_4$ | $\Psi_{T_{end},4}$ | 0.576 | -0.17 | 0.13 | 93.8 |
| Expected final outcome $Y_5$ under the adaptive design using $Y_5$ | $\Psi_{T_{end},5}$ | 0.573 | -0.15 | 0.15 | 95.0 |

Table 1.19: Performance of TMLE estimators for other causal estimands at the end of experiments in ADAPT-R trial simulation.

| Estimand | Notation | Truth | Bias ($\times 10^{-3}$) | Var ($\times 10^{-3}$) | Coverage (%) |
|---|---|---|---|---|---|
| Average treatment effect on the final outcome $Y_5$ | $\psi^{\text{ATE}}_{T_{end}}$ | -0.015 | 0.00 | 0.46 | 93.6 |
| Expected final outcome $Y_5$ under the estimated ODTR of $Y_1$ | $\psi^{d^*_{n,1}}_{T_{end}}$ | 0.590 | 0.98 | 0.12 | 95.0 |
| Expected final outcome $Y_5$ under the estimated ODTR of $Y_2$ | $\psi^{d^*_{n,2}}_{T_{end}}$ | 0.594 | 0.07 | 0.13 | 95.4 |
| Expected final outcome $Y_5$ under the estimated ODTR of $Y_3$ | $\psi^{d^*_{n,3}}_{T_{end}}$ | 0.595 | 1.87 | 0.15 | 91.2 |
| Expected final outcome $Y_5$ under the estimated ODTR of $Y_4$ | $\psi^{d^*_{n,4}}_{T_{end}}$ | 0.598 | 1.73 | 0.16 | 93.6 |
| Expected final outcome $Y_5$ under the estimated ODTR of $Y_5$ | $\psi^{d^*_{n,5}}_{T_{end}}$ | 0.597 | 0.55 | 0.19 | 92.2 |

Table 1.20: True expected final outcome $\tilde{\Psi}_{t,k}$ of implementing different designs in ADAPT-R trial simulation ($t = 11, 13, 15, 17, 19, T_{end}$).

| t | $\tilde{\Psi}_{t,RCT}$ | $\tilde{\Psi}_{t,1}$ | $\tilde{\Psi}_{t,2}$ | $\tilde{\Psi}_{t,3}$ | $\tilde{\Psi}_{t,4}$ | $\tilde{\Psi}_{t,5}$ | $\tilde{\Psi}_{t,naive}$ | $\tilde{\Psi}_{t,SL}$ |
|---|---|---|---|---|---|---|---|---|
| 11 | 0.538 | 0.541 | 0.540 | 0.538 | 0.538 | 0.538 | 0.538 | 0.538 |
| 13 | 0.547 | 0.555 | 0.552 | 0.551 | 0.549 | 0.547 | 0.546 | 0.552 |
| 15 | 0.557 | 0.569 | 0.566 | 0.564 | 0.562 | 0.560 | 0.559 | 0.565 |
| 17 | 0.562 | 0.576 | 0.574 | 0.572 | 0.570 | 0.568 | 0.566 | 0.573 |
| 19 | 0.564 | 0.577 | 0.576 | 0.574 | 0.572 | 0.570 | 0.568 | 0.574 |
| $T_{end}$ | 0.565 | 0.579 | 0.578 | 0.576 | 0.575 | 0.573 | 0.570 | 0.576 |

### 1.8.10 An Extension of the Proposed CARA Design

In some applications, when a plausible surrogate is available, experimenters may wish to adopt surrogate-guided adaptations before formally validating their utility for optimizing the final outcome. In such cases, a variant of TMLE-OSLAD can be used by designating a plausible surrogate (e.g., $Y_k$) as a primary outcome. The design then treats $Y_k$ as the working primary outcome and updates treatment randomization probabilities as early as $t = k + 1$. As the experiment progresses, later surrogates can be promoted in sequence as primary outcomes until the true final outcome $Y_K$ is observed. If prior knowledge or external data may indicate the relationship between surrogates and the primary outcome, one may also initiate an earlier adaptation based on that information. However, these approaches typically rely on generalizability assumptions or pre-specified models of surrogate and primary outcomes. Regardless of the strategy chosen, the TMLE-OSLAD framework can be invoked throughout the experiment to evaluate the opportunities and costs of each candidate adaptive design and dynamically select the best-performing one to support real-time decision making.

# Chapter 2

# Efficient Statistical Estimation for Sequential Adaptive Experiments with Implications for Adaptive Designs

## 2.1 Introduction

Adaptive experimental designs are useful for clinical trials due to their flexibility and statistical efficiency [16]. Unlike traditional, fixed trial designs, adaptive designs allow investigators to sequentially modify treatment randomization probabilities and/or other key design features based on accumulated data during the experiment. These adaptations can be targeted for a range of objectives: such as improving estimation efficiency for causal estimands and increasing participants' chances of receiving superior treatments.

Response-Adaptive Randomization (RAR), originating from the foundational work of [81], is an important adaptive design procedure in which treatment assignment probabilities are updated based on prior participants' treatments and outcomes [32, 72]. Covariate-Adjusted Response-Adaptive (CARA) designs extend RAR designs by incorporating baseline covariates into treatment assignment probabilities, enabling personalized treatment allocation for improved participant welfare or statistical efficiency [77, 32, 94, 53, 76]. The adaptive design approaches are also closely related to the literature on multi-armed bandits and contextual bandits, with broad applications in real-life decision making [10, 55, 7].

While adaptive experiments have the advantage of improving estimation efficiency and/or bringing benefits to participants compared with those with fixed, non-adaptive designs, they also present challenges for statistical inference. At the end of an adaptive experiment, investigators often seek to estimate causal estimands based on the observed data, such as the average treatment effect or mean outcome under personalized treatment rules. However, the adaptivity of such experiments induces dependence among observations, violating the

i.i.d. assumptions that standard statistical inference methods rely on. This has motivated substantial methodological work to provide statistical inference for data collected under adaptive designs. For example, [94] and [98] derived asymptotic results for inference under CARA designs within parametric modeling frameworks. [53] introduced a Targeted Maximum Likelihood Estimation (TMLE) framework which is tailored for adaptive design data and provides a non-parametric statistical inference method without relying on parametric assumptions. Subsequent developments include the development of statistical inference for causal estimands such as the risk difference and log-relative risk [12] as well as for estimating the mean outcome under optimal treatment rules [66, 60] in CARA experiments [14]. Another line of work employs variants of stabilized reweighting techniques [59] and develops asymptotically normal estimators for causal estimands in non-contextual [28] and contextual bandit settings [4, 89].

In contrast to these methods, we develop a unified framework that enables semiparametrically efficient estimation and inference for adaptive experiments and provides a new adaptive design that targets estimation efficiency.

First, we introduce a novel Adaptive-Design-Likelihood-based TMLE (ADL-TMLE) framework that enables statistical estimation and inference for a wide class of causal estimands from adaptive experiments, including the average treatment effect as our primary example. This estimator is driven by the likelihood structure of adaptive experiment and solves the score equation for components that exclusively define the causal estimand of interest. We show that ADL-TMLE is doubly robust, asymptotically normal and semiparametrically efficient.

A key insight from these results is that the minimal asymptotic variance of regular estimators under the adaptive experiment is inherently linked to that under a reference i.i.d. experiment in which treatments are assigned by the average of treatment randomization functions across the adaptive experiment (which we refer to as the "average design"). This highlights the advantage of ADL-TMLE in enabling statistical inference under relaxed assumptions for adaptive experiments. Specifically, it suffices to require positivity only for the average design, rather than for every individual treatment assignment probability. This enables greater flexibility in adaptive experiments, where local positivity violations are allowed during some segments of the experiment. The design stabilization conditions for establishing asymptotic normality and efficiency are also relaxed, as they are required only for the average design, which also leads to improved finite-sample performance of ADL-TMLE. Moreover, in scenarios where individual-level treatment randomization mechanisms are inaccessible due to operational or privacy constraints, our estimation framework remains feasible as long as the average design is accessible. These theoretical and practical advantages are further supported by simulation studies, which demonstrate the superior variance-reduction performance of ADL-TMLE across different adaptive experiments with varied objectives.

Second, while our framework provides a semiparametrically efficient estimator for a given adaptive experiment, even greater efficiency can be attained by guiding the real-time adaptive design toward an oracle design that is itself optimized for estimation efficiency. Motivated by ADL-TMLE framework, we propose a novel adaptive design framework that dynamically

steers the average design toward the evolving estimate of the oracle design as data accumulate throughout the experiment. This adaptive procedure leads to a lower variance than the standard efficiency-oriented adaptive design as shown by simulations.

Finally, our framework is directly applicable to settings with independent data structures, such as cases where certain factors induce heterogeneity in the treatment assignment mechanism but do not affect the primary outcome conditional on treatment and other baseline covariates. Furthermore, we generalize our framework to broader settings, including those with longitudinal data structures.

The remainder of this chapter is organized as follows. Section 2.2 introduces the statistical setup. Section 2.3 provides the estimation procedure of ADL-TMLE. Sections 2.4 and 2.5 establish asymptotic normality and semiparametric efficiency. Section 2.6 describes the proposed adaptive design approach. Section 2.7 reports simulation studies evaluating the performance of the proposed estimation and design methods. Section 2.8 discusses applications and generalizations of our framework. Finally, Section 2.9 concludes.

## 2.2 Data, Model and Statistical Setup

We define the full data as $X = (W, Y(1), Y(0)) \sim P^F \in \mathcal{M}^F$, where $W \in \mathcal{W}$ are baseline covariates, $Y(a) \in \mathcal{Y}$ is the counterfactual outcome under treatment level $a \in \mathcal{A}$. Without loss of generality, we consider binary treatment space $\mathcal{A} = \{0, 1\}$ and set $\mathcal{Y} = [0, 1]$. Here $P^F$ is the true distribution that generates full data $X$, living in a general model $\mathcal{M}^F$ without model prespecification. We use $E_{P^F}$ to denote the expectation under $P^F$. While our framework is generally applicable for estimating a wide class of causal estimands (see Section 2.8), here we use ATE $\Psi(P^F) := E_{P^F}[Y(1) - Y(0)]$ as the main example for illustration.

Consider an adaptive experiment that enrolls $n$ units sequentially, where $i$ is an index of each unit reflecting their enrollment order and $O_i := (W_i, A_i, Y_i)$ are the data to be observed for this unit. Upon the enrollment, the experimenter first observes baseline covariates $W_i$, then assigns treatment $A_i$ following a treatment randomization function $g_i : \mathcal{W} \times \mathcal{A} \to [0, 1]$ within a function class $\mathcal{G}$. Then, the unit's outcome $Y_i$ is observed. The adaptivity of the experiment is reflected in that treatment $A_i$ is determined by a function $g_i$ that is itself generated based on the observed data of previously enrolled units before $i$, which we denote by $\bar{O}(i-1) := (O_1, \cdots, O_{i-1})$. We use $\mathbf{g}_{1:n} := (g_1, \cdots, g_i, \cdots, g_n)$ to refer to an adaptive design that includes a collection of treatment randomization functions $g_i \in \mathcal{G}$ from $i = 1, \cdots, n$, which are known and controlled by the experimenter.

Define $\mathcal{Q} := \mathcal{Q}_W \otimes \mathcal{Q}_Y$, where $\mathcal{Q}_W$ is the set of possible distributions of $W$ and $\mathcal{Q}_Y$ is the set of possible conditional distributions of $Y$ given $A$ and $W$. For any $Q = (Q_W, Q_Y) \in \mathcal{Q}$, we denote $\bar{Q}(A, W)$ as a shorthand of $E[Y|A, W]$, the conditional mean of $Y$ given $A$ and $W$. For every $i \in [n]$, we assume that $W_i \sim Q_W$ and $Y_i \sim Q_Y(\cdot|A_i, W_i)$. We use $P_Q^n \in \mathcal{M}^n$ to denote the distribution that generates the adaptive experiment data $\bar{O}(n)$ in the adaptive

design model $\mathcal{M}^n$. The model specifies the likelihood of $\bar{o}(n) = (o_1, \cdots, o_i, \cdots, o_n)$ as

$$p^n_{q,g_1,\cdots,g_n}(\bar{o}(n)) = \prod_{i=1}^{n} q_w(w_i) \times \prod_{i=1}^{n} g_i(a_i|w_i) \times \prod_{i=1}^{n} q_y(y_i|a_i, w_i),$$

where we denote $q_w$ and $q_y$ as the probability density (or mass) functions corresponding to the distributions of $Q_W$ and $Q_Y$, and use $g_i(a|w)$ is a shorthand of $g_i(A = a|W = w)$.

One can identify the causal estimand $\Psi(P^F)$ with the following assumptions:

**Assumption 2.1** (Sequential Exchangeability)**.** $\forall i \in [n]$, $Y_i(a) \perp\!\!\!\perp A_i \mid W_i, \bar{O}(i-1)$.

**Assumption 2.2** (Theoretical Positivity)**.** $\forall w \in \mathcal{W}$, $a \in \mathcal{A}$, $\exists i \in [n]$ *such that* $g_i(a|w) > 0$.

**Theorem 2.1** (Identification of ATE)**.** *With Assumptions 2.1 and 2.2,* $\Psi(P^F)$ *is identified by*

$$\Psi(P^F) \;\; = \;\; E[\bar{Q}(1, W) - \bar{Q}(0, W)].$$

We denote this target estimand by $\psi(Q) := E[\bar{Q}(1, W) - \bar{Q}(0, W)]$.

## 2.3 Targeted Maximum Likelihood Estimation

At the end of an adaptive experiment, one observes a single series of dependent data $\bar{O}(n) = (O_1, \cdots, O_n)$. Let $\bar{Q}_n$ denote an initial estimate of the true conditional mean function $\bar{Q}_0$, estimated by a user-specified model or an ensemble learner such as Super Learner [47]. For any adaptive design $\mathbf{g}_{1:n}$, we define its average design $\bar{g}_n := \frac{1}{n}\sum_{i=1}^{n} g_i$.

Our proposed ADL-TMLE is

$$\psi(Q^*_n) := \frac{1}{n}\sum_{i=1}^{n} \left[\bar{Q}^*_n(1, W_i) - \bar{Q}^*_n(0, W_i)\right],$$

where $Q^*_n = (Q_{n,W}, Q^*_{n,Y})$ with $Q_{n,W}$ denotes the empirical distribution of baseline covariates, and $Q^*_{n,Y}$ an updated outcome distribution associated with a TMLE update $\bar{Q}^*_n$ of the initial estimate $\bar{Q}_n$. Specifically, $\bar{Q}^*_n$ is obtained by solving

$$\frac{1}{n}\sum_{i=1}^{n} \frac{2A_i - 1}{\bar{g}_n(A_i|W_i)} \left[Y_i - \bar{Q}^*_n(A_i, W_i)\right] = 0,$$

via the following regression:

$$\operatorname{logit} \bar{Q}_{n,\epsilon}(A, W) = \operatorname{logit} \bar{Q}_n(A, W) + \epsilon H_n(A, W),$$

where $H_n$ is defined by $H_n(A, W) := \frac{2A-1}{\bar{g}_n(A|W)}$ based on the average design $\bar{g}_n$. The TMLE update is $\bar{Q}^*_n = \bar{Q}_{n,\epsilon_n}$, where $\epsilon_n$ is the solution of that regression.

We note that our proposed ADL-TMLE is driven by the likelihood structure of adaptive experiment and solves the score equation for the $Q$-components that exclusively define the target parameter $\psi(Q)$. For any $Q = (Q_W, Q_Y) \in \mathcal{Q}$ and treatment randomization function $g \in \mathcal{G}$, we denote $D(Q, g)$ as the efficient influence curve (EIC) of the target estimand in an i.i.d. experiment where $W \sim Q_W, A \sim g(\cdot|W)$, and $Y \sim Q_Y(\cdot|A, W)$. For ATE, the EIC is $D(Q, g)(O) = D_1(Q, g)(O) + D_2(Q)(O)$, where $D_1(Q, g)(O) := \frac{2A-1}{g(A|W)} \left(Y - \bar{Q}(A, W)\right)$ and $D_2(Q)(O) := \bar{Q}(1, W) - \bar{Q}(0, W) - \psi(Q)$. The core procedure of ADL-TMLE is to solve $\frac{1}{n}\sum_{i=1}^{n} D(Q, \bar{g}_n)(O_i) = 0$, which is equivalent to solving the score equation as if in the i.i.d. experiment where the treatment distribution follows $A \sim \bar{g}_n(\cdot|W)$, even though the data is collected from the adaptive experiment.

For comparison, we also introduce the original TMLE tailored for adaptive design data proposed by [53], which we denote by $\psi(Q_n^\dagger)$ and refer to as "AD-TMLE" (adaptive-design TMLE). In AD-TMLE, $Q_n^\dagger = (Q_{n,W}, Q_{n,Y}^\dagger)$, and $\bar{Q}_n^\dagger$ is the TMLE update of the estimated conditional mean function $\bar{Q}_n$, which solves $\frac{1}{n}\sum_{i=1}^{n} \frac{2A_i-1}{g_i(A_i|W_i)}[Y_i - \bar{Q}_n^\dagger(A_i, W_i)] = 0$ using every $i$-specific treatment randomization function $g_i$ from the adaptive design $\mathbf{g}_{1:n}$. While next sections focus on properties of TMLEs, they also apply to the corresponding One-Step/AIPW estimators $\hat{\psi}_n(\bar{Q}_n) := \frac{1}{n}\sum_{i=1}^{n} \frac{2A_i-1}{\bar{g}_n(A_i|W_i)} \left[Y_i - \bar{Q}_n(A_i, W_i)\right] + \frac{1}{n}\sum_{i=1}^{n} \left[\bar{Q}_n(1, W_i) - \bar{Q}_n(0, W_i)\right]$, and $\hat{\psi}_n^\dagger(\bar{Q}_n) := \frac{1}{n}\sum_{i=1}^{n} \frac{2A_i-1}{g_i(A_i|W_i)} \left[Y_i - \bar{Q}_n(A_i, W_i)\right] + \frac{1}{n}\sum_{i=1}^{n} \left[\bar{Q}_n(1, W_i) - \bar{Q}_n(0, W_i)\right]$.

## 2.4 Asymptotic Normality

In this section we analyze the asymptotic normality of ADL-TMLE. We define $P_{Q,g}f := \int f(o)q(o)g(o)d\mu(o)$ for any $Q$ and $g$, and $P_{Q,g_i}f := \int f(o)q(o)g_i(o)d\mu(o)$ for every $g_i$ in adaptive design, where $\mu$ is a measure on $\mathcal{O}$ equipped with the product of the $\sigma$-fields of $\mathcal{W}$, $\mathcal{A}$ and $\mathcal{Y}$. Note that $P_{Q,\bar{g}_n}f = \frac{1}{n}\sum_{i=1}^{n} P_{Q,g_i}f$.

To establish the asymptotic normality of $\psi(Q_n^*)$, we first decompose the difference between $\psi(Q_n^*)$ and $\psi(Q_0)$ as follows.

**Theorem 2.2** (Difference between $\psi(Q_n^*)$ and $\psi(Q_0)$)**.** *The difference of $\psi(Q_n^*)$ and the true parameter $\psi(Q_0)$ can be decomposed as*

$$\psi(Q_n^*) - \psi(Q_0) = M_{1,n}(Q_\infty, \bar{g}_\infty) + M_{2,n}(Q_n^*, Q_\infty, \bar{g}_n) + M_{3,n}(Q_\infty, \bar{g}_n, \bar{g}_\infty),$$

*where*

$$\begin{aligned}
M_{1,n}(Q_\infty, \bar{g}_\infty) &:= \frac{1}{n}\sum_{i=1}^{n} \left[D(Q_\infty, \bar{g}_\infty)(O_i) - P_{Q_0,g_i}D(Q_\infty, \bar{g}_\infty)\right], \\
M_{2,n}(Q_n^*, Q_\infty, \bar{g}_n) &:= \frac{1}{n}\sum_{i=1}^{n} \Big\{ \left[D(Q_n^*, \bar{g}_n)(O_i) - P_{Q_0,g_i}D(Q_n^*, \bar{g}_n)\right] \\
&\qquad - \left[D(Q_\infty, \bar{g}_n)(O_i) - P_{Q_0,g_i}D(Q_\infty, \bar{g}_n)\right] \Big\},
\end{aligned}$$

$$M_{3,n}(Q_\infty, \bar{g}_n, \bar{g}_\infty) := \frac{1}{n}\sum_{i=1}^{n}\Big\{ [D(Q_\infty, \bar{g}_n)(O_i) - P_{Q_0,g_i} D(Q_\infty, \bar{g}_n)] - [D(Q_\infty, \bar{g}_\infty)(O_i) - P_{Q_0,g_i} D(Q_\infty, \bar{g}_\infty)] \Big\},$$

*with any fixed* $\bar{g}_\infty \in \mathcal{G}$ *and* $Q_\infty := (Q_{\infty,W}, Q_{\infty,Y}) \in \mathcal{Q}$*, where* $Q_{\infty,W} = Q_{0,W} \in \mathcal{Q}_W$ *and* $Q_{\infty,Y} \in \mathcal{Q}_Y$ *with its conditional mean function of the outcome* $\bar{Q}_\infty \in \bar{\mathcal{Q}}$*.*

The proof of this decomposition is given in Section 2.10.1.

Suppose there exist fixed $Q_\infty$ and $\bar{g}_\infty$ such that $\bar{Q}_n^*$ and $\bar{g}_n$ converge to them, respectively, under the assumptions stated below. We then establish the asymptotic normality of $M_{1,n}$, which is the average of a martingale difference sequence indexed by $Q_\infty$ and $\bar{g}_\infty$.

**Assumption 2.3** (Boundedness of the Average of Adaptive Design)**.** *There exists* $\zeta > 0$ *such that* $\inf_{a\in\mathcal{A},\, w\in\mathcal{W}} \bar{g}_n(a|w) \geq \zeta$*.*

**Assumption 2.4** (Stabilized Average of Conditional Variance)**.** *There exists* $\sigma_0^2 \in \mathbb{R}^+$ *such that*

$$\frac{1}{n}\sum_{i=1}^{n} Var_{P_{Q_0,g_i}}\left[D(Q_\infty, \bar{g}_\infty)\right] \xrightarrow{p} \sigma_0^2.$$

**Theorem 2.3** (Asymptotic Normality of $M_{1,n}$)**.** *Suppose assumptions 2.3 and 2.4 hold, then*

$$\sqrt{n} M_{1,n}(Q_\infty, \bar{g}_\infty) \xrightarrow{d} N(0, \sigma_0^2).$$

This asymptotic normality directly follows the martingale central limit theorems (e.g. Theorem 2 in [9]).

For the remaining martingale process terms $M_{2,n}$ and $M_{3,n}$, we establish their equicontinuity conditions following the equicontinuity results of the general setup in Chapter 1.

**Assumption 2.5** (Covering Number of $\bar{\mathcal{Q}}$)**.** $\bar{\mathcal{Q}}$ *is a class of functions with finite sup-norm covering number* $N_\infty(\epsilon, \bar{\mathcal{Q}})$*, with* $\mathcal{J}_\infty\left(\epsilon, \bar{\mathcal{Q}}\right) := \int_0^\epsilon \sqrt{\log\left(1 + \mathcal{N}_\infty\left(u, \bar{\mathcal{Q}}\right)\right)} du$ *the* $\epsilon$*-entropy integral in supremum norm of* $\bar{\mathcal{Q}}$*. Assume that* $\mathcal{J}_\infty\left(\epsilon, \bar{\mathcal{Q}}\right)$ *is finite and* $\mathcal{J}_\infty\left(\epsilon, \bar{\mathcal{Q}}\right) \to 0$ *as* $\epsilon \to 0$*.*

**Assumption 2.6** (Convergence of $\bar{Q}_n^*$)**.** *Define*

$$s_n\left(\bar{Q}, \bar{Q}_\infty, \bar{g}_n\right) := \sqrt{\frac{1}{n}\sum_{i=1}^{n} E\left[\rho\left(\frac{\left|D_1\left(\bar{Q}, \bar{g}_n\right)(O_i) - D_1\left(\bar{Q}_\infty, \bar{g}_n\right)(O_i)\right|}{c_1}\right)\middle| \bar{O}(i-1)\right]},$$

*where* $\rho(x) = e^x - x - 1$ *and* $c_1 = 8(1+\zeta)/\zeta^2$ *with* $\zeta$ *as assumed in Assumption 2.3. Assume that* $s_n(\bar{Q}_n^*, \bar{Q}_\infty, \bar{g}_n) = o_P(1)$*.*

**Theorem 2.4** (Equicontinuity of $M_{2,n}$)**.** *Suppose Assumptions 2.3, 2.5, 2.6 hold. Then,* $M_{2,n}(Q_n^*, Q_\infty, \bar{g}_n) = o_P(n^{-\frac{1}{2}})$.

**Assumption 2.7** (Covering Number of $\mathcal{G}$)**.** *$\mathcal{G}$ is a class of functions with finite sup-norm covering number $\mathcal{N}_\infty(\epsilon, \mathcal{G})$, with $\mathcal{J}_\infty(\epsilon, \mathcal{G}) := \int_0^\epsilon \sqrt{\log(1 + \mathcal{N}_\infty(u, \mathcal{G}))}du$ the $\epsilon$-entropy integral in supremum norm of $\mathcal{G}$. Assume that $\mathcal{J}_\infty(\epsilon, \mathcal{G})$ is finite and $\mathcal{J}_\infty(\epsilon, \mathcal{G}) \to 0$ as $\epsilon \to 0$.*

**Assumption 2.8** (Convergence of Average Design)**.** *Suppose* $\inf_{a\in\mathcal{A}, w\in\mathcal{W}} \bar{g}_\infty(a|w) \geq \eta$ *for some $\eta > 0$. Define*

$$v_n\left(\bar{Q}_\infty, \bar{g}_n, \bar{g}_\infty\right) := \sqrt{\frac{1}{n}\sum_{i=1}^n E\left[\rho\left(\frac{|D_1\left(\bar{Q}_\infty, \bar{g}_n\right)(O_i) - D_1\left(\bar{Q}_\infty, \bar{g}_\infty\right)(O_i)|}{c_2}\right)\middle| \bar{O}(i-1)\right]},$$

*where $\rho(x) = e^x - x - 1$ and $c_2 = 8/\eta^2$. Assume that $v_n(\bar{Q}_\infty, \bar{g}_n, \bar{g}_\infty) = o_P(1)$.*

**Theorem 2.5** (Equicontinuity of $M_{3,n}$)**.** *Suppose Assumptions 2.3, 2.7, 2.8 hold. Then,* $M_{3,n}(Q_\infty, \bar{g}_n, \bar{g}_\infty) = o_P(n^{-\frac{1}{2}})$.

Combining the asymptotic normality and equicontinuity conditions, we establish the asymptotic normality of the proposed ADL-TMLE $\psi(Q_n^*)$.

**Theorem 2.6** (Asymptotic Normality of ADL-TMLE)**.** *Suppose Assumptions 2.3,2.4, 2.5, 2.6, 2.7, and 2.8 hold, then $\psi(Q_n^*)$ is asymptotically normal:*

$$\sqrt{n}\left[\psi(Q_n^*) - \psi(Q_0)\right] \xrightarrow{d} N(0, \sigma_0^2).$$

Note that our assumptions do not require that $\bar{Q}_n^*$ converges to a $\bar{Q}_\infty$ that equals the true conditional mean function $\bar{Q}_0$. Thus ADL-TMLE has the double robustness property despite using a misspecified estimate of $\bar{Q}_0$, since the average design $\bar{g}_n$ is known.

Let $\sigma_n^2$ be a consistent estimator of $\sigma_0^2$. Then the standard error of $\psi(Q_n^*)$ can be given by $\sigma_n/\sqrt{n}$. The $100(1-\alpha)\%$ confidence interval is $[\psi(Q_n^*) - z_{1-\alpha/2}\sigma_n/\sqrt{n}, \psi(Q_n^*) + z_{1-\alpha/2}\sigma_n/\sqrt{n}]$, where $z_{1-\alpha/2}$ is the $1-\alpha/2$ quantile of standard normal distribution (e.g. $\alpha = 0.05$).

## 2.5 Semiparametric Efficiency

In this section, we provide semiparametric efficiency results of ADL-TMLE for adaptive experiments. We start by defining a Hilbert space: $\mathcal{H}(Q, g) := \{h \in L_2^0(P_{Q,g}) : h(o) = h_w(w) + h_y(w, a, y), \int h_w(w)dQ_w(w) = 0, \int h_y(w, a, y)dQ_y(y|a, w) = 0\}$ for any $Q = (Q_W, Q_Y) \in \mathcal{Q}$, with a reference treatment randomization distribution $g \in \mathcal{G}$ and the corresponding density $p_{Q,g}(o) = q_w(w)g(a|w)q_y(y|a, w)$ under $P_{Q,g}$. For any $h_1, h_2 \in \mathcal{H}(Q, g)$, define their inner product as $< h_1, h_2 >_{\mathcal{H}(Q,g)} := \int h_1(o)h_2(o)dP_{Q,g}(o)$. Define a class of paths:

$$\{Q_\epsilon^h = (Q_{w,\epsilon}^h, Q_{y,\epsilon}^h) : dQ_{w,\epsilon}^h = (1 + \epsilon h_w)dQ_w, dQ_{y,\epsilon}^h = (1 + \epsilon h_y)dQ_y, h \in \mathcal{H}(Q, \bar{g}_\infty)\}$$

for any $Q \in \mathcal{Q}$. We denote by $P^n_{Q^h_{\epsilon_n}}$ the fluctuated distribution of $P^n_Q$ with density $p^n_{Q^h_{\epsilon_n}, g_1, \cdots, g_n}$, where $\epsilon_n = 1/\sqrt{n}$.

In the following, we establish the semiparametric efficiency of ADL-TMLE in adaptive design settings by showing that it achieves the minimal asymptotic variance among regular estimators in the associated asymptotically normal experiment, as implied by the convolution theorem (Theorem 3.11.2 in [82]). The proofs are in Section 2.10.1.

### 2.5.1 Asymptotically Normal Experiment

We first analyze the sequence of statistical experiments $(\bar{\mathcal{O}}(n), \mathcal{F}_n, P^n_{Q^h_{\epsilon_n}} : h \in \mathcal{H}(Q, \bar{g}_\infty))$ and show that it is asymptotically normal. Let $\Delta_{n,h} = \frac{1}{\sqrt{n}} \sum_{i=1}^n h(O_i) + o_{P^n_Q}(1)$, and $\|h\|^2_{\mathcal{H}(Q,\bar{g}_\infty)} = \int h^2(o) dP_{Q,\bar{g}_\infty}(o)$. Let $\{\Delta_h : h \in \mathcal{H}(Q, \bar{g}_\infty)\}$ be an iso-Gaussian process with zero mean and covariance $P_{Q,\bar{g}_\infty} \Delta_{h_1} \Delta_{h_2} =< h_1, h_2 >_{\mathcal{H}(Q,\bar{g}_\infty)}$.

**Assumption 2.9** (Hilbert Space Norm Stability under Adaptive Design)**.** *For any $h \in \mathcal{H}(Q, \bar{g}_\infty)$, $\|h\|^2_{\mathcal{H}(Q,\bar{g}_n)} = \|h\|^2_{\mathcal{H}(Q,\bar{g}_\infty)} + o_{P^n_Q}(1)$.*

**Theorem 2.7** (Asymptotically Normal Experiment)**.** *Suppose Assumption 2.9 holds. Then, the experiment is asymptotically (shift) normal, i.e., for each $h \in \mathcal{H}(Q, \bar{g}_\infty)$,*

$$\log \frac{dP^n_{Q^h_{\epsilon_n}}(O_1, \cdots, O_n)}{dP^n_Q(O_1, \cdots, O_n)} = \Delta_{n,h} - \frac{1}{2}\|h\|^2_{\mathcal{H}(Q,\bar{g}_\infty)}$$

*such that $\Delta_{n,h} \xrightarrow{d} \Delta_h$ under $P^n_Q$, and $\Delta_h \sim N(0, \|h\|^2_{\mathcal{H}(Q,\bar{g}_\infty)})$.*

### 2.5.2 Regularity and Efficiency

Now we introduce the definition of regular estimator and present the limit distribution for regular estimators of regular parameters based on the convolution theorem. Define a sequence of parameters $\psi(Q^h_{\epsilon_n}) := E_{Q^h_{w,\epsilon_n}}[\bar{Q}^h_{y,\epsilon_n}(1, W) - \bar{Q}^h_{y,\epsilon_n}(0, W)]$ for every $h \in \mathcal{H}(Q, \bar{g}_\infty)$, where $Q^h_{w,\epsilon_n}$ denotes the perturbed distribution of $Q_W$ and $\bar{Q}^h_{y,\epsilon_n}$ is the conditional mean function of the outcome under the perturbed distribution of $Q_Y$.

For every $h \in \mathcal{H}(Q, \bar{g}_\infty)$, the sequence of parameters $\psi(Q^h_{\epsilon_n})$ is regular in that

$$\sqrt{n}\left(\psi(Q^h_{\epsilon_n}) - \psi(Q)\right) \to \dot{\psi}(Q)(h),$$

for a continuous, linear map $\dot{\psi} : \mathcal{H}(Q, \bar{g}_\infty) \to \mathbb{R}$ and a certain norming operator $\sqrt{n}$. Here, $\dot{\psi}(Q)(h)$ is the pathwise derivative of $\psi(Q^h_\epsilon)$ at $\epsilon = 0$ and equals $< D(Q, \bar{g}_\infty), h >_{\mathcal{H}(Q,\bar{g}_\infty)}$ by Riesz representation.

**Definition 2.1** (Regular Estimator)**.** *If a sequence of estimators $T_n(O_1, \cdots, O_n)$ satisfies*

$$\sqrt{n}\left(T_n - \psi(Q^h_{\epsilon_n})\right) \xrightarrow{d} L$$

*under* $P^n_{Q^h_{\epsilon_n}}$ *for every* $h \in \mathcal{H}(Q, \bar{g}_\infty)$ *for a fixed, tight Borel probability measure* $L$ *on* $\mathbb{R}$*, then* $T_n$ *is regular with respect to the norming operators* $\sqrt{n}$*.*

**Theorem 2.8** (Limit Distribution of Regular Estimators via the Convolution Theorem)**.** *Suppose Assumption 2.9 holds such that the sequence of statistical experiments* $(\bar{\mathcal{O}}(n), \mathcal{F}_n, P^n_{Q^h_{\epsilon_n}} : h \in \mathcal{H}(Q, \bar{g}_\infty))$ *is asymptotically normal. Then, for regular estimators* $T_n$ *of* $\psi(Q)$*, the limit distribution of* $\sqrt{n}(T_n - \psi(Q))$ *equals the distribution of a sum* $Z + B$ *of independent and tight random variables such that*

$$Z \sim N(0, \|\dot{\psi}^*(Q)\|^2_{\mathcal{H}(Q,\bar{g}_\infty)}),$$

*where* $\dot{\psi}^*(Q)$ *is the adjoint of* $\dot{\psi}(Q)$ *such that* $< \dot{\psi}^*(Q), h >_{\mathcal{H}(Q,\bar{g}_\infty)} = \dot{\psi}(Q)(h)$*.*

We establish the regularity and efficiency of ADL-TMLE as follows.

**Theorem 2.9** (Regularity and Efficiency)**.** *Suppose Assumptions 2.3 -2.9 hold with* $Q_\infty = Q_0$*, then* $\psi(Q^*_n)$ *is regular and efficient, i.e., the asymptotic variance of* $\sqrt{n}(\psi(Q^*_n) - \psi(Q_0))$ *achieves the minimal asymptotic variance* $\|\dot{\psi}^*(Q_0)\|^2_{\mathcal{H}(Q_0,\bar{g}_\infty)}$ *among all regular estimators.*

**Remark 2.1** (Comparison of two TMLEs)**.** *We compare ADL-TMLE* $\psi(Q^*_n)$ *with AD-TMLE* $\psi(Q^\dagger_n)$ *based on the results of Sections 2.4 and 2.5. First, AD-TMLE generally requires* $\inf_{a\in\mathcal{A},w\in\mathcal{W}} g_i(a \mid w) \geq \zeta$ *for every* $i$*, whereas ADL-TMLE only requires positivity of the average design* $\bar{g}_n$*. Thus, ADL-TMLE accommodates greater variation in treatment randomization probabilities over the course of the experiment, including local violations of positivity, provided that the average design remains bounded away from zero. Second, the asymptotic normality of both estimators requires stabilization of the average conditional variance of their respective martingale differences. For AD-TMLE, the relevant variance is* $\frac{1}{n}\sum_{i=1}^n \mathrm{Var}_{P_{Q_0,g_i}}\{D(Q_\infty, g_i)(O_i)\}$*, whereas for ADL-TMLE it is* $\frac{1}{n}\sum_{i=1}^n \mathrm{Var}_{P_{Q_0,g_i}}\{D(Q_\infty, \bar{g}_\infty)(O_i)\}$*. Suppose that both martingale variances stabilize and that the two estimators have the same limiting* $Q_\infty$*. For the finite-sample reference design* $\bar{g}_n$*, comparison of the corresponding variances and application of Jensen's inequality yields* $\frac{1}{n}\sum_{i=1}^n \mathrm{Var}_{P_{Q_0,g_i}}\{D(Q_\infty, \bar{g}_n)(O_i)\} \leq \frac{1}{n}\sum_{i=1}^n \mathrm{Var}_{P_{Q_0,g_i}}\{D(Q_\infty, g_i)(O_i)\}$*. The inequality is strict unless the designs* $g_1, \ldots, g_n$ *are identical across subjects. If both variance sequences converge and replacing* $\bar{g}_n$ *by* $\bar{g}_\infty$ *is asymptotically negligible, then the limiting variance of ADL-TMLE is no larger than that of AD-TMLE.*

## 2.6 Adaptive Designs Motivated by Efficiency Results

Beyond efficient estimation after the completion of a given adaptive experiment, we investigate how to design real-time adaptive experiments to further reduce the variance of the estimator for a specific causal estimand of interest. Ideally, one can characterize an oracle experimental design that minimizes semiparametric efficiency bound for estimating that estimand

using the data collected through that oracle design, which we denote by $\pi^* : \mathcal{W} \times \mathcal{A} \to [0,1]$. Using ATE as an example, it is known that the oracle design $\pi^*$ follows the conditional Neyman allocation: $\pi^*(a|w) := \frac{\sigma(a,w)}{\sigma(1,w)+\sigma(0,w)}$, where $\sigma^2(a,w) := Var[Y|A=a,W=w]$.

However, the oracle design is unknown *a priori* in practice. This necessitates adaptive experimental designs that sequentially learn the oracle design from the accumulated data and determine treatment randomization probabilities for newly enrolled units according to the evolving estimate of the oracle design, with the aim of improving estimation efficiency. Motivated by efficiency results of ADL-TMLE, in this section we propose a novel adaptive design that guides the average of adaptive design toward the oracle design, which we refer to "$\bar{g}$-driven adaptive design".

This design is constructed as follows. At the beginning of the experiment, we enroll $n_0$ units and randomly assign treatments with a baseline probability (e.g. 0.5). Here, $n_0$ is a user-specified "burn-in" sample size to estimate the oracle design for adapting treatment randomization probabilities. After that, suppose we have enrolled $i$ units and define $\bar{g}_i = \frac{1}{i}\sum_{j=1}^{i} g_j$ as the current average of adaptive design. Upon the enrollment of each new unit $i+1$, we use the accumulated data $\bar{O}(i) = (O_1, \ldots, O_i)$ to construct $\hat{\pi}^*_{i+1}$ as an estimate of the oracle design $\pi^*$ for the causal estimand of interest. In the example of ATE, $\hat{\pi}^*_{i+1}$ is constructed by first estimating the unknown conditional variance $\sigma^2(a,w)$ by $\hat{\sigma}^2_{1:i}(a,w)$ based on $\bar{O}(i)$, and then plugging that into the Neyman allocation formula: $\hat{\pi}^*_{i+1}(a|w) = \frac{\hat{\sigma}_{1:i}(a,w)}{\hat{\sigma}_{1:i}(1,w)+\hat{\sigma}_{1:i}(0,w)}$. We then integrate this estimate into $\overline{\hat{\pi}}^*_{i+1} := \frac{1}{i-n_0+1}\sum_{j=n_0+1}^{i+1} \hat{\pi}^*_j$ as the target for the next-step average design, which mitigates the variability inherent in adapting solely to the latest estimate and produces a smoother, more stable design trajectory in finite samples. Finally, by solving $\bar{g}_{i+1} = \overline{\hat{\pi}}^*_{i+1}$ for the next average design, the treatment randomization function $g_{i+1}$ is given by

$$g_{i+1} := (i+1)\overline{\hat{\pi}}^*_{i+1} - i\,\bar{g}_i,$$

clipped to $[0,1]$. This adaptation procedure is repeated upon enrollment of the new unit.

**Remark 2.2** (Comparison with standard Neyman-allocation adaptation)**.** *Compared with the adaptive designs with standard Neyman allocation that directly sets $g_{i+1} = \hat{\pi}^*_{i+1}$, our $\bar{g}$-driven adaptive design adopts a "self-calibrating" approach: if the average design in earlier stages presents higher probability to assign treatment (control) relative to the estimated optimal design, the $\bar{g}$-driven adaptive design compensates in subsequent stages by reducing the randomization probability for treatment (control) accordingly. This design strategy offers global control over the adaptive design as a whole, steering the overall average design toward the oracle design in a stable manner, rather than optimizing $g_{i+1}$ alone in isolation without considering previous treatment randomization in the adaptive experiment. Together, the $\bar{g}$-driven design framework and efficient estimation framework of ADL-TMLE work jointly to improve estimation efficiency in answering various causal questions of interest.*

## 2.7 Simulation Studies

### 2.7.1 Setup

This section presents simulation studies to evaluate the performance of the proposed estimator and adaptive design to minimize variance in estimating ATE as a representative example. Specifically, we examine the following aspects: **(A)** Estimation: We evaluate the performance of ADL-TMLE in their efficiency in estimating ATE and compared that with AD-TMLE across adaptive designs that either aim to improve estimation efficiency for the causal estimand or target a different objective. **(B)** Design: With the goal of minimizing variance in ATE estimation, we evaluate the performance of the proposed $\bar{g}$-driven design and compare that with other alternative designs.

To answer these questions, we consider five types of experimental designs that span 10 time points, starting with $n_0 = 1000$ units enrolled initially, and 250 units added at each subsequent time point. Each design differs in its mechanism to determine treatment randomization probabilities given the accumulated data at each time point. One design is non-adaptive, assigning treatments and controls with equal probability ($g_i \equiv 0.5$ for every $i$) throughout the experiment. Another three designs are based on Neyman allocation: one is a standard adaptive design that directly applies the current estimate of Neyman allocation for treatment randomization for the new enrollee that sets $g_i = \hat{\pi}_i^*$; another is our proposed $\bar{g}$-driven Neyman allocation design; and the third is an ideal oracle design, which applies the oracle Neyman allocation from the outset ($g_i = \pi^*$ for every $i$), as if it were known. We also implement a "benefit-driven" design that estimates conditional average treatment effect (CATE) function $E[Y(1) - Y(0)|W]$ from the accumulated data and tilts the treatment randomization probabilities toward the estimate of optimal personalized treatment $I(E[Y(1) - Y(0)|W] > 0)$, thereby increasing the expected outcome of new enrollees compared with that in the fixed non-adaptive design. Each design is repeated across 5000 Monte Carlo runs. For each run, the TMLE approach is implemented after rescaling the outcome to $[0, 1]$ using the sample minimum and maximum. More details are deferred to Section 2.10.2.

Figure 2.1 displays the oracle Neyman allocation and CATE functions across baseline covariates. The oracle Neyman allocation function varies across covariates, resulting in heterogeneous treatment randomization probabilities, while CATE function is also heterogeneous and follows a different pattern. Consequently, the benefit-driven design is expected to yield higher variance in ATE estimation than the adaptive designs using Neyman allocation.

### 2.7.2 Results

**Estimation.** Table 2.1 presents the bias, variance, and MSE of the proposed ADL-TMLE in adaptive experiments with benefit-driven design and standard Neyman-allocation design, respectively. We also evaluate the coverage rate of the 95% Wald-type confidence interval using the empirical EIC variance $\sigma_n^2 := \frac{1}{n}\sum_{i=1}^{n} D(Q_n^*, \bar{g}_n)(O_i)^2$. It also reports the coverage of the 95% Wald-type "oracle confidence interval" based on the estimator's Monte Carlo stan-

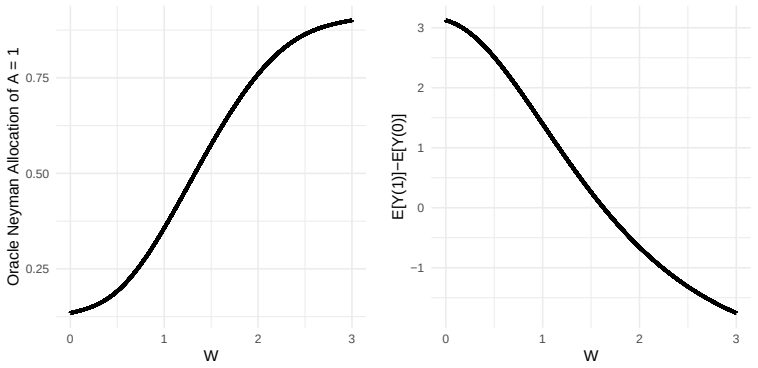



Figure 2.1: Left: Probability of assigning treatment assignment $a = 1$ (y-axis) under the oracle Neyman allocation across baseline covariates (x-axis). Right: Conditional average treatment effect (CATE) (y-axis) across baseline covariates (x-axis).

dard deviation of the point estimates, with its nominal coverage supporting the estimator's asymptotic normality.

| Design | Time | Bias ($10^{-2}$) | Var. ($10^{-2}$) | MSE ($10^{-2}$) | Cov. (%) | Oracle Cov. (%) |
|---|---|---|---|---|---|---|
| Benefit Driven | 2 | −0.55 | 6.22 | 6.23 | 93.7 | 94.6 |
| Benefit Driven | 4 | −1.56 | 5.37 | 5.40 | 94.4 | 94.5 |
| Benefit Driven | 6 | −2.40 | 5.00 | 5.05 | 95.0 | 94.6 |
| Benefit Driven | 8 | −3.27 | 4.82 | 4.93 | 94.6 | 94.4 |
| Benefit Driven | 10 | −2.94 | 4.91 | 5.00 | 94.2 | 94.4 |
| Standard Neyman | 2 | 1.90 | 5.49 | 5.53 | 92.9 | 94.9 |
| Standard Neyman | 4 | 2.93 | 3.72 | 3.81 | 92.3 | 94.4 |
| Standard Neyman | 6 | 1.16 | 2.94 | 2.95 | 92.0 | 94.7 |
| Standard Neyman | 8 | 1.65 | 2.30 | 2.33 | 92.6 | 94.7 |
| Standard Neyman | 10 | 1.09 | 2.03 | 2.04 | 92.6 | 95.1 |

Table 2.1: Performance of ADL-TMLE in adaptive experiments with benefit-driven design and standard Neyman-allocation design at time points 2, 4, 6, 8, and 10. The table includes bias, variance, mean squared error (MSE), coverage probability of 95% Wald-type confidence intervals based on estimated standard errors, along with coverage probability of the 95% Wald-type "oracle confidence intervals" constructed by the Monte Carlo standard deviation of the point estimates.

Figure 2.2 compares the relative variance of the proposed TMLE estimator ADL-TMLE $\psi(Q_n^*)$ with that of AD-TMLE $\psi(Q_n^\dagger)$. The results show that ADL-TMLE $\psi(Q_n^*)$ significantly reduces variance compared with AD-TMLE in adaptive experiments, under both benefit-driven design and standard Neyman allocation, demonstrating ADL-TMLE's efficiency across different adaptive experiments with varying design objectives.

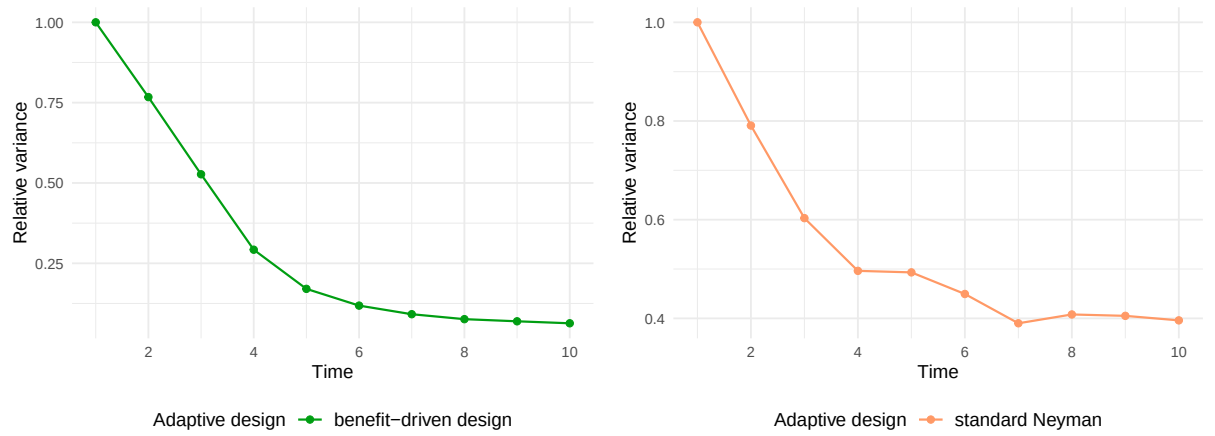



Figure 2.2: Relative variance of ADL-TMLE $\psi(Q_n^*)$ versus AD-TMLE $\psi(Q_n^\dagger)$ over the course of adaptive experiments with benefit-driven design (left) and with standard Neyman allocation (right). In each plot, the x-axis is the time point of the adaptive experiment, and the y-axis is the relative variance of the two TMLEs, each constructed by the cumulative data up to that time point.

**Design.** Figure 2.3 presents the relative variance of ADL-TMLE under different designs compared with that in the non-adaptive design with fixed treatment randomization probability 0.5. The results show that both adaptive designs with standard Neyman allocation and the proposed $\bar{g}$-driven Neyman allocation yield lower variance than the non-adaptive design. Moreover, the $\bar{g}$-driven adaptive design consistently results in a smaller variance than the adaptive design with standard Neyman allocation and approaches the oracle design more closely.

Additionally, we evaluate the effects of different designs on semiparametric efficiency for ATE estimation. Specifically, we construct oracle versions of Neyman-allocation-based adaptive designs, in which the treatment randomization probabilities are based on the oracle Neyman allocation functions using the true conditional variances rather than the estimated ones for units $i \geq n_0$. Figure 2.4 plots the trajectory of the relative value of $D(Q_0, \bar{g}_n)^2$ over time for these designs (including the non-adaptive one) against the ideal oracle design that applies the oracle Neyman allocation from the outset. This illustrates how quickly the variance of ADL-TMLE under an adaptive design approaches the semiparametric efficiency bound minimized by the oracle Neyman allocation. The results show that this quantity under the $\bar{g}$-driven adaptive design converges to the oracle one more rapidly than the standard Neyman allocation, highlighting the advantage of guiding the global average design toward the optimal target.

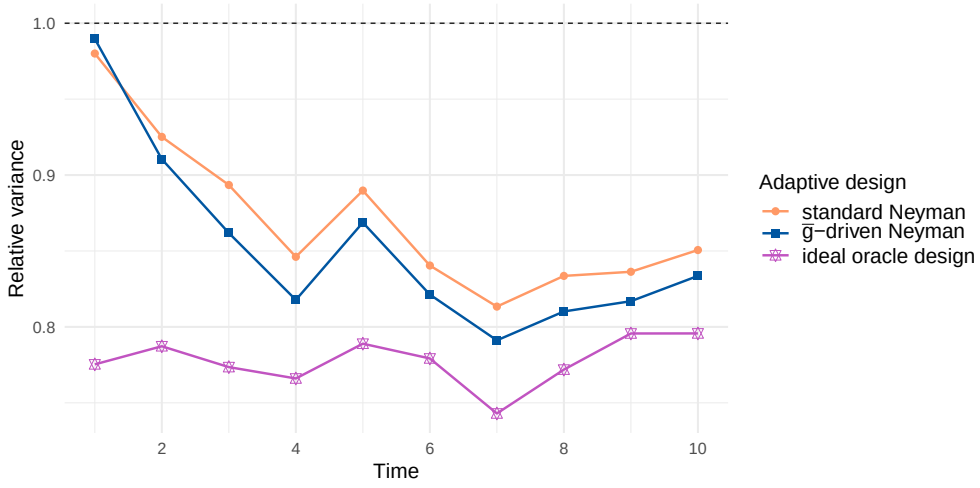



Figure 2.3: Relative variance of ADL-TMLE in experiments with different designs versus that in the non-adaptive design with a fixed treatment probability of 0.5. Each estimate is constructed using cumulative data up to each time point. The x-axis is the time point of the experiment, and the y-axis is the relative variance.

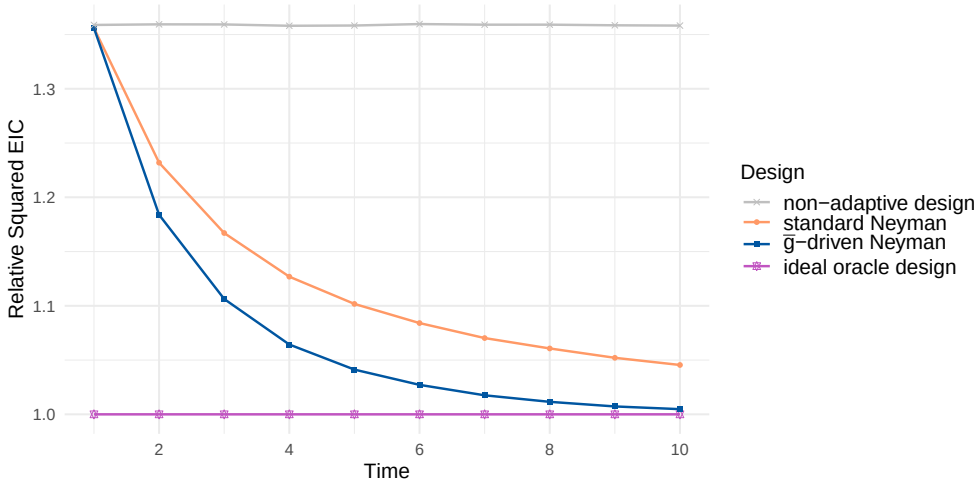



Figure 2.4: Relative value of $D(Q_0, \bar{g}_n)^2$ under various designs versus that under the ideal design with oracle Neyman allocation (y-axis) across different time points (x-axis). Different lines show the trajectories of this quantity over time under non-adaptive design, the oracle version of the adaptive design using standard Neyman allocation and the oracle version of $\bar{g}$-driven Neyman allocation, respectively.

## 2.8 Application and Generalization

### 2.8.1 Application

Our proposed framework is directly applicable for cases with independent data structure. For example, consider estimating causal estimands from data collected across multiple sites where the site index does not affect the primary outcome but introduces heterogeneity in treatment distribution across sites. Specifically, define the full data of interest as $X = (W, Y(a)) \sim P^F \in \mathcal{M}^F$, where $W$ denotes the baseline covariates; $Y(a)$ is the counterfactual outcome with treatment level $a \in \mathcal{A}$. Let the average treatment effect $\Psi^F(P^F) = E_{P^F}[Y(1) - Y(0)]$ be the causal estimand of interest. Suppose one observes $n$ independent observations $O_1, \cdots, O_n$ with $O_i = (S_i, W_i, A_i, Y_i)$, where $W_i \sim Q_W$ and $Y_i \sim Q_Y(\cdot|A_i, W_i)$ for $Q_W \in \mathcal{Q}_W$ and $Q_Y \in \mathcal{Q}_Y$. For every unit $i$, $S_i \in \{1, \ldots, J\}$ is the site index that determines the treatment distribution of $A_i \sim g_i(\cdot|W_i)$ with $g_i = g_{s_i}$. Each site $j \in [J]$ has its own treatment assignment mechanism $g_j : \mathcal{W} \times \mathcal{A} \to [0, 1]$. Note that each site does not impact the outcome given treatment and baseline covariates, and $\Psi^F(P^F)$ can be identified by the target estimand $\psi(Q) = E[\bar{Q}(1, W) - \bar{Q}(0, W)]$ under ignorability and positivity assumptions, where $\bar{Q}(A, W) := E[Y|A, W]$.

Define $\bar{g}_n := \frac{1}{n}\sum_{i=1}^n g_i$. With the convergence of $\bar{g}_n$ to a fixed $\bar{g}_\infty$ (e.g. $\bar{g}_\infty = \sum_{j \in J} p_s(j) g_j$ for a probability mass function $p_s$ of the site index), one can construct an ADL-TMLE by solving the score equation $\frac{1}{n}\sum_{i=1}^n \frac{2A_i - 1}{\bar{g}_n(A_i|W_i)}[Y_i - \bar{Q}^*_n(A_i, W_i)] = 0$. This yields an asymptotically normal and efficient estimator that inherits the advantages of ADL-TMLE established for adaptive design settings. This approach also applies to other settings where the site index $S$ is replaced by other variables that do not directly affect the outcome given treatment and other baseline covariates, but only influence the treatment assignment probabilities.

### 2.8.2 The General Framework

We present the general estimation and design framework for various causal estimands in broader settings, including those with longitudinal data structure.

**Data, Model and Target Estimand.** Consider a longitudinal data structure $O = (L(0), A(0), L(1), A(1), \cdots, L(K), A(K), L(K+1) \equiv Y)$. Let

$$\mathcal{M}(g) = \{p_{q,g} = (q,\, g) : q \in \mathcal{Q}\}$$

be a model with a known $g$, where $q(o) = \prod_{t=0}^{K+1} q_t(o)$ and $g(o) = \prod_{t=0}^{K} g_t(o)$, which corresponds to an i.i.d. likelihood with each $L(t) \sim q_t(\cdot|Pa(L(t)))$ and $A(t) \sim g_t(\cdot|Pa(A(t)))$, where $Pa(L(t))$ and $Pa(A(t))$ denote the set of nodes that appear before $L(t)$ and $A(t)$ in the longitudinal structure of $O$, respectively. Let $g^* \in \mathcal{G}^*$ denote a fixed function that defines the causal estimand of interest as $\Psi^F(Q) = EY_{g^*} = \int_o y\, q(o)\, g^*(o)\, d\mu(o)$. Suppose that we

have $n$ observations $\bar{O}(n) = (O_i : i = 1, \ldots, n)$ with likelihood

$$p^n_{q,\mathbf{g}_{1:n}}(\bar{o}(n)) = \prod_{i=1}^{n} q(o_i)\, g_i(o_i) = \prod_{i=1}^{n} \prod_{t=0}^{K+1} q_t(o_i) \prod_{t=0}^{K} g_{i,t}(o_i).$$

This defines the model $M^n = \{p^n_{q,\mathbf{g}_{1:n}} : q\}$, where $g_i$ may be specified as known, externally controlled, or partly unknown but estimable from the data. Note that the statistical estimand of interest only depends on the $Q$-component of the likelihood.

**Semiparametrically Efficient Estimation.** Define $\bar{g}_n = \frac{1}{n}\sum_{i=1}^{n} g_i$ as the average of the whole product $g_i = \prod_t g_{i,t}$. Let $D(Q, g)(O_i)$ denote the efficient influence curve in the i.i.d. model $\mathcal{M}(g)$. To estimate $EY_{g^*}$ from data generated by the model $\mathcal{M}^n$, it is useful to consider a reference i.i.d. model $\mathcal{M}(\bar{g}_n)$ with the average design $\bar{g}_n$. In that i.i.d. model, the efficient influence curve for the parameter $\Psi(Q)$ is given by $D(Q, \bar{g}_n)(O)$. Importantly, any TMLE constructed by solving $n^{-1}\sum_{i=1}^{n} D(Q, \bar{g}_n)(O_i) = 0$ under the i.i.d. model with $\bar{g}_n$ is also an efficient estimator for the original model $\mathcal{M}^n$ for the observed data, with its asymptotic normality supported by Martingale theories and its semiparametric efficiency provided by the convolution theorem, as shown in Sections 2.4 and 2.5. This unified estimation framework applies not only to adaptive design settings discussed above, but also to other contexts with longitudinal data structure with dynamic interventions [73, 44] and time-series structures [46, 61]. For the design problem, the construction of the self-calibrating $\bar{g}$-driven adaptive design implied by this efficient estimation approach directly mirrors that in Section 2.6.

## 2.9 Conclusions

In this work, we propose a novel ADL-TMLE framework for estimating a wide class of causal estimands from sequential adaptive experiments. We establish both asymptotic normality and semiparametric efficiency of ADL-TMLE, and demonstrate superior variance-reduction performance in finite samples. Notably, our framework substantially relaxes the requirements for adaptive experiments that enable valid statistical inference, as it only requires positivity and stabilization for the average design. Motivated by these results, we further propose a novel $\bar{g}$-driven adaptive design that dynamically steers the average design toward the estimated optimal design in a stable self-calibrating manner. This adaptive design achieves faster convergence to the oracle design and yields improved estimation efficiency compared with standard approach. Finally, we provide a general estimation and design framework applicable to broader settings with longitudinal structures. Future research could explore extensions to other related settings and broader areas of applications.

## 2.10 Supplementary Material

### 2.10.1 Proofs for Sections 2.4 and 2.5

*Proof of Theorem 2.2.* By applying von Mises expansions with respect to $P_{Q,\bar{g}_n}$ and $P_{Q_0,g_i}$ across $g_i \in \mathbf{g}_{1:n}$, we have that

$$\begin{aligned}
& \psi(Q) - \psi(Q_0) \\
= \; & \frac{1}{n}\sum_{i=1}^{n} \int D(Q,\bar{g}_n)(o) d\left(P_{Q,\bar{g}_n} - P_{Q_0,g_i}\right)(o) \\
& + \frac{1}{n}\sum_{i=1}^{n} \int \sum_{a=0}^{1} (2a-1) \left( \frac{g_i(a|w) - \bar{g}_n(a|w)}{\bar{g}_n(a|w)} \right) \left( \bar{Q}(a,w) - \bar{Q}_0(a,w) \right) dQ_W(w) \\
= \; & -\frac{1}{n}\sum_{i=1}^{n} P_{Q_0,g_i} D(Q,\bar{g}_n) \\
= \; & -P_{Q_0,\bar{g}_n} D(Q,\bar{g}_n).
\end{aligned}$$

Since ADL-TMLE solves $\frac{1}{n}\sum_{i=1}^{n} D(Q_n^*,\bar{g}_n)(O_i) = 0$, we have

$$\psi(Q_n^*) - \psi(Q_0) = \frac{1}{n}\sum_{i=1}^{n} [D(Q_n^*,\bar{g}_n)(O_i) - P_{Q_0,\bar{g}_n} D(Q_n^*,\bar{g}_n)].$$

By adding and subtracting $D(Q_\infty,\bar{g}_n)(O_i)$, $D(Q_\infty,\bar{g}_\infty)(O_i)$, and their conditional expectations given $\bar{O}(i-1)$, we obtain the desired decomposition equation. □

*Proof of Theorem 2.7.* The log likelihood ratio of the perturbed distribution and the original distribution follows:

$$\begin{aligned}
& \log \frac{dP^n_{Q^h_{\epsilon_n}}(O_1,\cdots,O_n)}{dP^n_Q(O_1,\cdots,O_n)} \\
= \; & \frac{1}{\sqrt{n}}\sum_{i=1}^{n} h(O_i) - \frac{1}{2n}\sum_{i=1}^{n} h^2(O_i) + o_{P^n_Q}(1) \\
= \; & \frac{1}{\sqrt{n}}\sum_{i=1}^{n} h(O_i) - \frac{1}{2n}\sum_{i=1}^{n} E_{P^n_Q}\left[h^2(O_i) \mid \bar{O}(i-1)\right] + o_{P^n_Q}(1) \\
= \; & \frac{1}{\sqrt{n}}\sum_{i=1}^{n} h(O_i) - \frac{1}{2}\|h\|^2_{\mathcal{H}(Q,\bar{g}_n)} + o_{P^n_Q}(1) \\
= \; & \Delta_{n,h} - \frac{1}{2}\|h\|^2_{\mathcal{H}(Q,\bar{g}_\infty)},
\end{aligned}$$

where we denote $\Delta_{n,h} := \frac{1}{\sqrt{n}}\sum_{i=1}^{n} h(O_i) + o_{P^n_Q}(1)$, with the first term a zero-mean martingale. The second equation is based on the convergence of conditional variance of the zero-mean

martingale according to Theorem 2.23 of [29], as $E[h^2(O_i)|\bar{O}(i-1)]$ is bounded. The third equation is based on the dependence of $g_i$ on the past $\bar{O}(i-1)$ in our adaptive design setup and the definition of $\|h\|^2_{\mathcal{H}(Q,\bar{g}_n)}$. The last equation relies on the convergence of the adaptive design described in Assumption 2.9 and results in $\Delta_{n,h} \xrightarrow{d} \Delta_h \sim N(0, \|h\|^2_{\mathcal{H}(Q,\bar{g}_\infty)})$ by Martingale central limit theorem. □

*Proof of Theorem 2.9.* By Assumptions 2.3, 2.4, 2.5, 2.6, 2.7, and 2.8, one can apply Theorem 2.6 to establish asymptotic normality of $\psi(Q_n^*)$ under $P^n_{Q_0}$:

$$\sqrt{n}[\psi(Q_n^*) - \psi(Q_0)] \xrightarrow{d} N\left(0, P_{Q_0,\bar{g}_\infty} D(Q_0, \bar{g}_\infty)^2\right).$$

With Assumption 2.9, the log likelihood ratio for the perturbed distribution of $P^n_{Q_0}$ is obtained by $log \frac{dP^n_{Q^h_{\epsilon_n}}(O_1,\cdots,O_n)}{dP^n_{Q_0}(O_1,\cdots,O_n)} = \Delta_{n,h} - \frac{1}{2}\|h\|^2_{\mathcal{H}(Q_0,\bar{g}_\infty)}$, where $h \in \mathcal{H}(Q_0, \bar{g}_\infty)$ and $\Delta_{n,h} \xrightarrow{d} \Delta_h \sim N(0, \|h\|^2_{\mathcal{H}(Q_0,\bar{g}_\infty)})$ under $P^n_{Q_0}$ via Theorem 2.7. Then, by Le Cam's third lemma (Theorem 3.10.7 in [82]), we have that

$$\sqrt{n}[\psi(Q_n^*) - \psi(Q_0)] \xrightarrow{d} N(< D(Q_0, \bar{g}_\infty), h >, P_{Q_0,\bar{g}_\infty} D(Q_0, \bar{g}_\infty)^2)$$

under $P^n_{Q^h_{\epsilon_n}}$. Combining with the regularity of the parameter

$$\sqrt{n}\left(\psi(Q^h_{\epsilon_n}) - \psi(Q_0)\right) \to \dot{\psi}(Q_0)(h),$$

where $\dot{\psi}(Q_0)(h) =< D(Q_0, \bar{g}_\infty), h >$, we have that

$$\sqrt{n}\left(\psi(Q_n^*) - \psi(Q^h_{\epsilon_n})\right) \xrightarrow{d} N(0, P_{Q_0,\bar{g}_\infty} D(Q_0, \bar{g}_\infty)^2)$$

under $P^n_{Q^h_{\epsilon_n}}$. Hence $\psi(Q_n^*)$ is regular.

By the convolution theorem (Theorem 2.8), the asymptotic variance of $\sqrt{n}(T_n - \psi(Q_0))$ for regular estimators $T_n$ is no less than $\|\dot{\psi}^*(Q_0)\|^2_{\mathcal{H}(Q_0,\bar{g}_\infty)}$, which equals $P_{Q_0,\bar{g}_\infty} D(Q_0, \bar{g}_\infty)^2$ and is achieved by the asymptotic variance of $\psi(Q_n^*)$. Therefore, $\psi(Q_n^*)$ is efficient. □

### 2.10.2 Simulation Details

The data generating distribution is as follows: $W \sim U(0,3)$; $\bar{Q}_0(1,W) = 25 + 10/\left[1 + exp(-(2W+1))\right]$; $\bar{Q}_0(0,W) = 20 + 17.5/\left[1 + exp(-(W+0.1))\right]$; $Var_{Q_0}[Y|A = 1, W] = 1 + 3W^3$; $Var_{Q_0}[Y|A = 0, W] = 1 + 1.5 \times (3 - W)^3$. Figure 2.5 shows the true conditional expectation and variance of the outcome given different treatment levels and baseline covariates.

For benefit-driven design, we guide treatment randomization probabilities toward optimal personalized treatment using baseline covariates through estimating the conditional average treatment effect (CATE). Define the true CATE function as $B_0 : \mathcal{W} \to \mathbb{R}$, with $B_0(W) =$

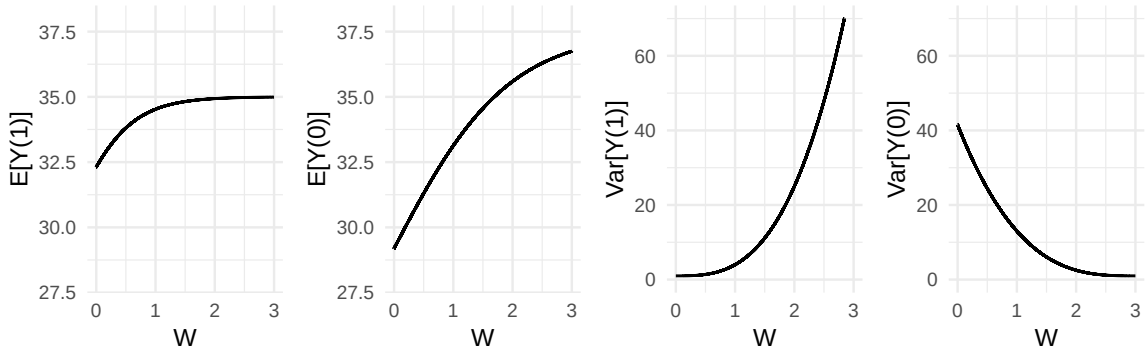



Figure 2.5: Summary plots of the data generating distribution. The first two plots show the conditional mean of the outcome (y-axis) for treatment ($a = 1$) and control ($a = 0$) against covariates $W$ (x-axis). The last two plots present the corresponding conditional variances.

$\bar{Q}_0(1, W) - \bar{Q}_0(0, W)$. For a newly enrolled unit $i$ with baseline covariates $w_i$, we use $B_n(w_i)$ to denote the estimate of $B_0(w_i)$ based on the current data $\bar{O}(i-1)$. We estimate CATE functions by first transforming the outcomes into doubly robust pseudo-outcomes [78, 43, 39], then regressing them on baseline covariates using Super Learner [47]. This estimated CATE guides the treatment randomization probabilities toward the estimated optimal personalized treatment $I(B_n(w_i) > 0)$ through a function $\Gamma_{\nu,b}$ introduced by [14]:

$$\Gamma_{\nu,b}(x) = \nu I(x \leq -b) + (1-\nu) I(x \geq b) + \left( -\frac{1/2 - \nu}{2b^3} x^3 + \frac{1/2 - \nu}{2b/3} x + \frac{1}{2} \right) I\left(-b \leq x \leq b\right).$$

Here, $B_n(w_i)$ is used as input to $\Gamma_{\nu,b}$ to determine the probability of assigning treatment $a = 1$; $\nu \in [0, 0.5]$ sets a lower bound on the treatment randomization probability, ensuring that each arm is chosen with probability at least $\nu$; $b > 0$ controls the range within which these probabilities respond smoothly to the estimated CATE $B_n(w_i)$. When $B_n(w_i) > b$, indicating a substantially higher expected outcome under treatment $a = 1$ compared with $a = 0$, the treatment allocation probability of $a = 1$ reaches the maximum $1 - \nu$; when $B_n(w_i) < -b$, the treatment allocation probability of $a = 1$ drops to its minimum value $\nu$. When $B_n(w_i) \in [-b, b]$, that probability varies smoothly with the value of CATE. As the estimated CATE approaches 0, the probability gets closer to 0.5, balancing exploration and exploitation. In simulation, we set $b = 1$ and $\nu_i = exp(-t_i)$ for every unit $i$, where $t_i = \left\lceil \frac{i - n_0}{250} \right\rceil$ with $n_0 = 1000$, thereby increasing the chance of assigning substantially better treatments as the experiment progresses.

# Chapter 3

# Longitudinal Adaptive Experimental Design for Learning Multiple Target Estimands with Semiparametric Efficient Inference

## 3.1 Introduction

Adaptive designs are increasingly used in clinical trials and digital experiments [16, 72, 54], allowing treatment randomization probabilities to adapt in response to accumulating data. Dating back to [81], a central class of such methods is Response-Adaptive Randomization (RAR), in which treatment assignment probabilities are updated using treatment and outcome information from previously enrolled participants, with objectives ranging from improving efficiency to reducing assignment to inferior treatments [86, 33, 32, 72]. Covariate-Adjusted Response-Adaptive (CARA) designs generalize RAR by incorporating baseline covariates into the randomization mechanism [77, 32, 94, 53, 76, 1, 34]. These adaptive designs are also related to developments in the multi-armed and contextual bandit literature, which studies sequential decision-making strategies with the objective of regret minimization [10, 55].

However, relatively few studies have investigated adaptive experimental designs in longitudinal settings. Existing work in this area primarily considers settings with longitudinally observed responses or covariates but a single treatment decision, where intermediate information is used to adapt the randomization for that decision [36, 37, 6, 38, 88, 93, 22]. Nevertheless, these approaches do not consider adaptive designs in more complex longitudinal settings involving multi-stage, time-varying treatments. Such settings are canonical in the causal inference literature [73, 66, 2, 44, 49] and arise commonly in clinical decision-making, precision medicine, and sequential multiple assignment randomized trials [56, 65, 57, 40, 41]. In these scenarios, participants are followed over multiple stages, intermediate

covariates or responses are revealed sequentially, and treatment at each stage is assigned based on a participant's evolving history until the final outcome is observed. This longitudinal structure fundamentally differs from experiments with a single treatment stage and calls for adaptive randomization strategies that explicitly account for the sequential nature of treatment assignment and information accrual. While prior work has considered adaptive experiments or bandit algorithms with time-varying treatments to improve participants' outcomes [53, 35, 84, 25], principled approaches to optimizing the estimation efficiency of statistical estimands remain relatively underdeveloped in these longitudinal settings.

In this chapter, we develop a semiparametric efficiency framework for constructing longitudinal adaptive experiments with stage-wise treatment randomization mechanisms that efficiently learn one or multiple target estimands. We further propose a general framework for semiparametric efficient estimation of statistical estimands from dependent data collected under longitudinal adaptive experiments, which applies to both the efficiency-oriented designs in this work and broader classes of adaptive experiments.

The contributions of this work are as follows. (i) We propose a general efficiency-oriented design criterion for longitudinal experiments that targets the estimation efficiency of one or multiple estimands within a unified formulation. (ii) For a broad class of longitudinal target estimands, we characterize the oracle design that optimizes the proposed criterion and uncover its forward-looking structure: optimal randomization at earlier stages depends on optimal allocations at later stages. This structure yields a backward-recursive strategy for deriving explicit stage-specific oracle randomization probabilities. (iii) We develop a covariate-adjusted response-adaptive (CARA) design that sequentially updates stage-wise randomization probabilities by learning and targeting the oracle design using accumulating data. (iv) Finally, we propose an adaptive-design-likelihood longitudinal targeted maximum likelihood estimator (ADL-LTMLE) that enables asymptotically normal and efficient estimation of target estimands in the multi-stage adaptive experiments with time-varying treatments. Our approach builds on the literature on targeted maximum likelihood estimation (TMLE) for adaptively collected data [53, 13, 14, 61, 62, 93], and recent work that establishes semiparametric efficiency of the ADL-TMLE approach under adaptive experiments [92].

We illustrate the framework through longitudinal adaptive experiments for estimating multiple treatment-initiation effects. Simulations show that the proposed adaptive design yields substantial variance reduction relative to non-adaptive designs, especially when combined with the efficient ADL-LTMLE approach, with its performance close to that of the oracle design.

Overall, this work contributes a general maximal-information-type framework for the design and analysis of adaptive experiments with longitudinal, time-varying treatments. The framework accommodates a wide range of target estimands and their joint collections, including estimands of causal effects and mean outcomes under dynamic or stochastic treatment rules, and recovers single-stage adaptive designs as a special case.

The remainder of this chapter is organized as follows. Section 3.2 describes the statistical setup. Section 3.3 introduces the proposed efficiency-oriented design criterion, and develops a backward-recursive strategy for characterizing oracle designs for either a single target

estimand or multiple target estimands and constructing the corresponding longitudinal adaptive designs. Section 3.4 develops semiparametric efficient adaptive-design-likelihood-based longitudinal TMLE methods for estimating target statistical estimands and providing inference based on data collected from longitudinal adaptive experiments. Section 3.5 presents treatment-initiation effects as an illustrative application of the general framework and derives oracle and adaptive designs for their efficient joint estimation. Section 3.6 reports the simulation studies. Section 3.7 concludes with a discussion.

## 3.2 Statistical Setup

### 3.2.1 Data Structure

Consider a longitudinal data structure $O = (L_1, A_1, \ldots, L_K, A_K, L_{K+1} \equiv Y)$ over $K+1$ follow-up stages for each participant. At each stage $k \in [K] := \{1, \ldots, K\}$, $L_k \in \mathcal{L}_k$ denotes covariates (which can also be intermediate responses for $k > 1$), $A_k \in \mathcal{A}_k$ denotes the categorical treatment assigned at stage $k$. The final outcome $Y \in \mathcal{Y}$, also denoted by $L_{K+1}$, is a bounded outcome observed at stage $K+1$, which we assume without loss of generality that $\mathcal{Y} = [0, 1]$. Let $\bar{A}_k := (A_1, \ldots, A_k)$ and $\bar{L}_k := (L_1, \ldots, L_k)$ denote the treatment and covariate histories up to stage $k$, and define $H_k := (\bar{A}_{k-1}, \bar{L}_k) \in \mathcal{H}_k$ as the participant's observed history immediately before treatment $A_k$. Let $Q_1$ denote the marginal distribution of baseline covariates $L_1$, and for $k = 1, \ldots, K$, let $Q_{k+1}$ denote the conditional distribution of $L_{k+1}$ given $(A_k, H_k)$. We collect these distributions into $\mathbf{Q} := (Q_1, \ldots, Q_{K+1})$. Let $\mathbf{g} := (g_1, \ldots, g_K)$ denote a collection of stage-wise treatment randomization mechanisms, where each $g_k : \mathcal{H}_k \times \mathcal{A}_k \to [0, 1]$ maps a history-treatment pair $(h_k, a_k)$ to the conditional probability of assigning treatment $a_k$ given history $h_k$. We write $g_k(a_k \mid h_k)$ as shorthand for $P(A_k = a_k \mid H_k = h_k)$ under the randomization mechanism $g_k$. We consider the nonparametric statistical model $\mathcal{M} = \{P_{\mathbf{Q},\mathbf{g}} : \mathbf{Q} \in \mathcal{Q},\ \mathbf{g} \in \mathcal{G}\}$, where each $P_{\mathbf{Q},\mathbf{g}}$ admits the following likelihood factorization:

$$p_{\mathbf{Q},\mathbf{g}}(o) = q_1(l_1) \prod_{k=1}^{K} g_k(a_k \mid h_k)\, q_{k+1}(l_{k+1} \mid a_k, h_k), \tag{3.1}$$

which defines the joint density of the longitudinal data $O$ under $P_{\mathbf{Q},\mathbf{g}}$. Here, $\mathcal{Q}$ denotes the nonparametric model for $\mathbf{Q} = (Q_1, \ldots, Q_{K+1})$, and $\mathcal{G}$ denotes the class of admissible treatment randomization functions.

### 3.2.2 Longitudinal Adaptive Experiment

Consider a sequential longitudinal adaptive experiment with participants indexed by $i = 1, \ldots, n$, where each participant is followed for $K+1$ stages. For each participant $i$, we observe longitudinal data $O_i = (L_{1,i}, A_{1,i}, \ldots, L_{K,i}, A_{K,i}, Y_i)$, where $L_{k,i}$ denotes time-varying covariates at stage $k$, $A_{k,i} \in \mathcal{A}_k$ denotes the categorical treatment assigned at stage $k$,

and $Y_i \equiv L_{K+1,i}$ is the final outcome. Define the accumulated data up to participant $i$ by $\bar{O}_i := \{O_1, \ldots, O_i\}$, where $\bar{O}_0 := \emptyset$. Let $\mathcal{F}_i = \sigma(O_1, \ldots, O_i)$ denote the filtration generated by the sequentially observed data.

Before participant $i$ is enrolled, the experimenter uses the accumulated data $\bar{O}_{i-1}$ to construct a collection of stage-wise randomization mechanisms $\mathbf{g}_i = (g_{1,i}, \ldots, g_{K,i})$. Conditional on $\bar{O}_{i-1}$, treatment at stage $k$ for participant $i$ is assigned according to the known randomization mechanism $g_{k,i}(\cdot \mid H_{k,i})$. These mechanisms govern the participant's treatment assignments sequentially across stages throughout longitudinal follow-up. This formulation also includes cohort-wise adaptation as a special case. Specifically, if participants are enrolled in cohorts, the same collection of randomization mechanisms $\mathbf{g}_i$ may be applied to all participants within a cohort and updated only between cohorts.

Let $\mathbf{g}_{1:n} := (\mathbf{g}_1, \ldots, \mathbf{g}_n)$ denote the known adaptive design sequence applied to the $n$ participants. For a given $(\mathbf{Q}, \mathbf{g}_{1:n})$, we write $P_{\mathbf{Q},\mathbf{g}_{1:n}}$ for the joint distribution of $(O_1, \ldots, O_n)$. The likelihood of $(o_1, \ldots, o_n)$ is given by

$$p_{\mathbf{Q},\mathbf{g}_{1:n}}(o_1, \ldots, o_n) \;=\; \prod_{i=1}^{n} \left[ q_1(l_{1,i}) \prod_{k=1}^{K} \Big\{ g_{k,i}(a_{k,i} \mid h_{k,i})\, q_{k+1}(l_{k+1,i} \mid a_{k,i}, h_{k,i}) \Big\} \right].$$

## 3.3 A General Framework for Efficiency-Oriented Longitudinal Adaptive Design

### 3.3.1 Oracle Longitudinal Design for Estimation Efficiency

Let $\psi(\mathbf{Q})$ denote a pathwise differentiable target estimand of interest. We introduce a semiparametric efficiency framework for defining an efficiency-oriented design criterion. Specifically, for a fixed design $\mathbf{g}$, define

$$C(\mathbf{Q}, \mathbf{g}) := Var_{P_{\mathbf{Q},\mathbf{g}}} \left\{ D(P_{\mathbf{Q},\mathbf{g}})(O) \right\}, \tag{3.2}$$

where $D(P_{\mathbf{Q},\mathbf{g}})$ denotes the efficient influence curve (EIC) of $\psi(\mathbf{Q})$ under the observed-data distribution $P_{\mathbf{Q},\mathbf{g}}$. Under standard regularity conditions, the pathwise derivative of $\psi(\mathbf{Q})$ along any regular parametric submodel can be represented as the inner product of its score with an influence function. By the Riesz representation theorem, the EIC is the unique element of the tangent space that represents this derivative and has the smallest $L^2(P_{\mathbf{Q},\mathbf{g}})$ norm among all influence functions. Consequently, (3.2) is the semiparametric efficiency bound that characterizes the smallest asymptotic variance attainable by a regular asymptotically linear estimator of $\psi(\mathbf{Q})$ [48].

While the collection of randomization mechanisms $\mathbf{g} = (g_1, \ldots, g_K)$ does not alter the definition of $\psi(\mathbf{Q})$, it determines the induced observed-data distribution $P_{\mathbf{Q},\mathbf{g}}$ and, consequently, the statistical information available for estimating $\psi(\mathbf{Q})$. This motivates the oracle maximal-information longitudinal design, defined as

$$\mathbf{g}^\star(\mathbf{Q}) := \arg\min_{\mathbf{g} \in \mathcal{G}} C(\mathbf{Q}, \mathbf{g}). \tag{3.3}$$

For many longitudinal target estimands, the EIC admits a stage-wise decomposition $D(P_{\mathbf{Q},\mathbf{g}}) = \sum_{k=0}^{K} D_k(P_{\mathbf{Q},\mathbf{g}})$, where $D_0(P_{\mathbf{Q},\mathbf{g}})$ is the mean-zero baseline component associated with the marginal law of $L_1$. For each $k \in [K]$, $D_k(P_{\mathbf{Q},\mathbf{g}})(O)$ is the component associated with the conditional law of $L_{k+1}$ given $(H_k, A_k)$; it is measurable with respect to $(H_k, A_k, L_{k+1})$ and satisfies $E_{P_{\mathbf{Q},\mathbf{g}}}\{D_k(P_{\mathbf{Q},\mathbf{g}})(O) \mid H_k, A_k\} = 0$. These stage-specific components form martingale increments and are mutually orthogonal in $L^2(P_{\mathbf{Q},\mathbf{g}})$. Consequently, the variance of the EIC decomposes as $Var_{P_{\mathbf{Q},\mathbf{g}}}\{D(P_{\mathbf{Q},\mathbf{g}})(O)\} = \sum_{k=0}^{K} E_{P_{\mathbf{Q},\mathbf{g}}}\left[\{D_k(P_{\mathbf{Q},\mathbf{g}})(O)\}^2\right]$. Thus, the semiparametric efficiency bound separates into stage-specific contributions through which the longitudinal treatment mechanisms affect the total estimation variance.

### 3.3.2 Backward-Recursive Structure of the Oracle Design

We consider a broad class of longitudinal target estimands whose stage-specific EIC components have the following representation:

$$D_k(P_{\mathbf{Q},\mathbf{g}})(O) = \frac{\Phi_k(\mathbf{Q})(O)}{\prod_{s=1}^{k} g_s(A_s \mid H_s)}, \qquad k = 1, \ldots, K, \tag{3.4}$$

where we require that each $\Phi_k(\mathbf{Q})$ is an estimand-specific function that does not depend on the experimental mechanism $\mathbf{g}$.

This representation characterizes a general class of statistical estimands supported by the proposed design optimization framework, including many classical causal estimands in longitudinal settings such as mean outcomes under static, dynamic, or stochastic treatment rules and their linear combinations (see, e.g., [73, 48, 66, 45, 85]).

Importantly, for this class of estimands, the stage-wise EIC components exhibit a nested dependence on the longitudinal treatment mechanisms, as indicated by the following simplified indexing:

$$D_1(\Phi_1, g_1), \quad D_2(\Phi_2, g_1, g_2), \quad \ldots, \quad D_K(\Phi_K, g_1, \ldots, g_K).$$

Specifically, conditional on the realized history $H_k$ available before treatment assignment at stage $k$, the randomization mechanism $g_k$ affects the variance contribution of $D_k$ and all subsequent components $D_{k+1}, \ldots, D_K$, where the latter contributions further depend on the future treatment mechanisms $g_{k+1}, \ldots, g_K$.

This nested dependence gives rise to a forward-looking, backward-recursive optimization of the longitudinal design. The recursion starts by optimizing the terminal randomization mechanism $g_K$ and then proceeds backward through $g_{K-1}, \ldots, g_1$, so that each earlier-stage randomization mechanism accounts for its downstream contribution to the variance through the subsequent data trajectories it induces. The procedure is as follows.

First, at the final stage $K$, let $\sigma_K^2(a_K \mid h_K) = E\left[\{\Phi_K(\mathbf{Q})(O)\}^2 \mid H_K = h_K, A_K = a_K\right]$. Conditional on $H_K = h_K$, minimizing the final-stage variance contribution of $D_K$ yields the

allocation:

$$g_K^\star(a_K \mid h_K) = \frac{\sigma_K(a_K \mid h_K)}{\sum_{a'_K \in \mathcal{A}_K} \sigma_K(a'_K \mid h_K)}. \tag{3.5}$$

Intuitively, the oracle design assigns greater probability to treatment arms associated with larger current and future uncertainty conditional on the current history. Collecting more observations under these arms reduces their contribution to the overall uncertainty and improves the efficiency of estimating the target estimand.

Proceeding backward, for $k = K-1, \ldots, 1$, define

$$\begin{aligned}\sigma_k^2(a_k \mid h_k) = E\left[\{\Phi_k(\mathbf{Q})(O)\}^2 \mid H_k = h_k, A_k = a_k\right] \\ + E\left[\sum_{a_{k+1} \in \mathcal{A}_{k+1}} \frac{\sigma_{k+1}^2(a_{k+1} \mid H_{k+1})}{g_{k+1}^\star(a_{k+1} \mid H_{k+1})} \middle| H_k = h_k, A_k = a_k\right],\end{aligned} \tag{3.6}$$

which aggregates the conditional variance contributions of $D_k, \ldots, D_K$ under the future oracle mechanisms $(g_{k+1}^\star, \ldots, g_K^\star)$.

The resulting stage-$k$ oracle allocation is given by:

$$g_k^\star(a_k \mid h_k) = \frac{\sigma_k(a_k \mid h_k)}{\sum_{a'_k \in \mathcal{A}_k} \sigma_k(a'_k \mid h_k)}, \qquad k = K, \ldots, 1. \tag{3.7}$$

The final collection $\mathbf{g}^\star := (g_1^\star, \ldots, g_K^\star)$ thus constitutes the oracle longitudinal design across all stages.

### 3.3.3 Oracle Longitudinal Design for Multiple Target Estimands

With rich data trajectories involving time-varying treatments and responses, longitudinal studies create opportunities to address a wide range of scientific questions through multiple target estimands of interest. However, they also give rise to complex trade-offs in experimental design, as a design optimized for one estimand may provide less favorable data support for estimating others. This challenge is particularly substantial in longitudinal experiments, where efficiency losses due to limited support may propagate multiplicatively through the stage-wise dependence of the longitudinal data. This motivates a joint design criterion that explicitly accounts for the trade-offs in estimation efficiency across multiple target estimands.

Consider a collection of target estimands of interest $\{\psi^j(\mathbf{Q}) : j \in [J]\}$. We define the joint efficiency-oriented criterion

$$C_w(\mathbf{Q}, \mathbf{g}) := \sum_{j \in [J]} w^j Var_{P_{\mathbf{Q},\mathbf{g}}}\left\{D^j(P_{\mathbf{Q},\mathbf{g}})(O)\right\}, \tag{3.8}$$

where $D^j(P_{\mathbf{Q},\mathbf{g}})$ denotes the EIC corresponding to $\psi^j(\mathbf{Q})$ under $P_{\mathbf{Q},\mathbf{g}}$, and the user-specified nonnegative weights $\{w^j : j \in [J]\}$ satisfying $\sum_{j=1}^{J} w^j = 1$ determine the relative emphasis placed on the estimation efficiency of each target estimand.

When the EIC of each target estimand admits the stage-wise representation in (3.4), the optimization for the joint criterion also preserves the same backward-recursive structure. Specifically, for each estimand $\psi^j$, let $\sigma_k^{2,j}(a_k \mid h_k)$ denote the conditional second-moment contribution of its stage-$k$ and subsequent EIC components under the common future oracle mechanisms. Define $\Sigma_k^2(a_k \mid h_k) = \sum_{j\in[J]} w^j \sigma_k^{2,j}(a_k \mid h_k)$. The oracle allocation conditional on the history of stage $k$ is then $g_k^\star(a_k \mid h_k) = \frac{\Sigma_k(a_k|h_k)}{\sum_{a_k'\in\mathcal{A}_k} \Sigma_k(a_k' \mid h_k)}$ for $k = K, \ldots, 1$. Thus, the joint criterion retains the same stage-wise allocation form, while aggregating uncertainty across the target estimands according to user-specified weights that reflect their relative estimation priorities.

More generally, let $\boldsymbol{D}(P_{\mathbf{Q},\mathbf{g}}) = \left\{D^1(P_{\mathbf{Q},\mathbf{g}}), \ldots, D^J(P_{\mathbf{Q},\mathbf{g}})\right\}^\top$ denote the vector-valued EIC, and consider $C_{\boldsymbol{\Omega}}(\mathbf{Q}, \mathbf{g}) = \mathrm{tr}\left[\boldsymbol{\Omega}\, Var_{P_{\mathbf{Q},\mathbf{g}}}\{\boldsymbol{D}(P_{\mathbf{Q},\mathbf{g}})(O)\}\right]$ as a design criterion, where $Var_{P_{\mathbf{Q},\mathbf{g}}}\{\boldsymbol{D}(P_{\mathbf{Q},\mathbf{g}})(O)\}$ is the covariance matrix of the EICs and $\boldsymbol{\Omega}$ is a user-specified symmetric positive-semidefinite matrix. In particular, let $\boldsymbol{\Omega} := B^\top \Lambda B$, where $B \in \mathbb{R}^{R\times J}$ is a prespecified contrast matrix whose rows define $R$ linear combinations of the $J$ target estimands, and let $\Lambda = \mathrm{diag}(\lambda^1, \ldots, \lambda^R)$ be a diagonal matrix of nonnegative weights. This criterion represents a weighted sum of the semiparametric efficiency bounds for the user-specified linear combinations of the estimands. The backward-recursive optimization procedure continues to apply.

### 3.3.4 Efficiency-Oriented Longitudinal Adaptive Design

However, oracle designs depend on unknown features of the data-generating distribution and therefore cannot be implemented *a priori*. This motivates a longitudinal adaptive design that estimates the relevant quantities sequentially from accumulating data and updates the stage-specific randomization mechanisms accordingly.

Specifically, before enrolling participant $i$, we can use the accumulated observations $\bar{O}_{i-1}$ to estimate $\{\sigma_k^2(a_k \mid H_k) : k \in [K]\}$ for a single target estimand, or $\{\Sigma_k^2(a_k \mid H_k) : k \in [K]\}$ for multiple target estimands. Substituting these estimates into the backward-recursive solution yields the estimated oracle mechanisms $\{\hat{g}_{k,i}^\star(\cdot \mid H_k) : k \in [K]\}$. The stage-specific treatment mechanisms for participant $i$ are then given by $g_{k,i}(\cdot \mid H_k) = \hat{g}_{k,i}^\star(\cdot \mid H_k)$ for $k \in [K]$. By doing so, the longitudinal adaptive design updates the randomization mechanisms as information accumulates, thereby steering the distribution of the accruing data toward that under the oracle design with the intended efficiency objective.

### 3.3.5 Inference After Longitudinal Adaptive Experiments

While adaptive experiments are attractive because they can steer the randomization mechanisms toward the design objective, the resulting dependence across participants invalidates standard inferential procedures developed for i.i.d. data, particularly when inference is conducted without parametric assumptions. In longitudinal experiments, statistical inference is further complicated by the sequential dependence among time-varying covariates, treatments, and outcomes within each participant. In the next section, we provide a targeted maximum likelihood estimation (TMLE) framework to address these challenges.

## 3.4 Adaptive-Design-Likelihood-Based Longitudinal TMLE

We develop an adaptive-design-likelihood-based longitudinal TMLE (ADL-LTMLE) approach for semiparametric efficient inference under longitudinal adaptive experiments, building on recent advances in adaptive-design-likelihood-based TMLE (ADL-TMLE) framework [92]. This framework relates semiparametric efficiency bound under an adaptive experiment to the corresponding bound under an i.i.d. reference distribution induced by the average of the distributions generated under the adaptive design. It thereby yields an asymptotically normal and semiparametric efficient estimator for adaptively collected data without relying on parametric model assumptions.

This framework applies to a broad class of longitudinal target estimands, including those defined under dynamic or stochastic treatment rules and other estimands that admit analogous TMLE procedures based on martingale and semiparametric efficiency theory. More generally, it applies to target estimands of the form $\psi(\mathbf{Q})$ for data likelihoods factorizing as $p(o_1, \ldots, o_n) = \prod_{i=1}^{n} \mathbf{q}(o_i)\mathbf{g}_i(o_i)$, where $\mathbf{q}$ is the density of some $\mathbf{Q} \in \mathcal{Q}$ that defines the target estimand. Thus, the framework is not restricted to a particular type of adaptive experiment or to the estimands used to guide the adaptive design, provided that the observed data offer sufficient support for the target estimand of interest.

For the rest of this section, we use $\psi^{\mathbf{d}}(\mathbf{Q})$, the mean outcome under a dynamic treatment rule $\mathbf{d}$, as a representative example of the proposed framework in longitudinal settings. We first formalize this estimand and introduce the proposed ADL-LTMLE procedure, and then establish its asymptotic normality and semiparametric efficiency, with proofs provided in the Supplementary Material (Section 3.8).

### 3.4.1 Expected Outcome under a Dynamic Treatment Rule

A dynamic treatment rule is a sequence of decision rules $\mathbf{d} = (d_1, \ldots, d_K)$, where each $d_k : \mathcal{H}_k \to \mathcal{A}_k$ maps the observed history $H_k$ to a treatment at stage $k$. For each $k \in [K]$, let $L^{\mathbf{d}}_{k+1:K+1} = \left(L^{\mathbf{d}}_{k+1}, \ldots, L^{\mathbf{d}}_{K+1}\right)$ denote the counterfactual trajectory of covariates/outcomes from stage $k+1$ onward under rule $\mathbf{d}$, with $L^{\mathbf{d}}_{K+1} = Y^{\mathbf{d}}$. The causal estimand of interest

is $\Psi^{\mathbf{d}} := E\left(Y^{\mathbf{d}}\right)$, the mean counterfactual outcome under rule $\mathbf{d}$. Let $P^{\mathbf{d}}_{\mathbf{Q}_0}$ denote the intervention distribution induced by $\mathbf{Q}_0$ when treatment is assigned by rule $\mathbf{d}$, and define $\bar{G}^{\mathbf{d}}_{k,n}(h_k) = \frac{1}{n}\sum_{i=1}^{n}\prod_{s=1}^{k} g_{s,i}\{d_s(h_s) \mid h_s\}$ as the average cumulative probability of following $\mathbf{d}$ across participants in the adaptive experiment.

**Assumption 3.1** (Identification)**.** *For a treatment rule $\mathbf{d} = (d_1, \ldots, d_K)$, suppose that:*

*(i) Consistency: if $A_{s,i} = d_s(H_{s,i})$ for every $s \le k$, then $L_{k+1,i} = L^{\mathbf{d}}_{k+1,i}$;*

*(ii) Sequential randomization: for every $k \in [K]$, $L^{\mathbf{d}}_{k+1:K+1} \perp\!\!\!\perp A_{k,i} \mid H_{k,i}, \bar{O}_{i-1}$, where $\bar{O}_{i-1}$ denotes accumulated data before participant $i$ is enrolled;*

*(iii) Positivity: $\forall k \in [K]$, $\bar{G}^{\mathbf{d}}_{k,n}(h_k) > 0$ for $P^{\mathbf{d}}_{\mathbf{Q}_0}$-almost every history $h_k$.*

Here, consistency links the observed and counterfactual variables along the target treatment rule, sequential randomization removes treatment confounding conditional on the observed history and prior experimental data, and positivity ensures that the adaptive experiment provides support for the treatment trajectory required to evaluate $\mathbf{d}$.

To state the identification formula, for notational convenience, we define $d_{K+1}(H_{K+1}) \equiv \emptyset$ and set $m^{d}_{K+1}(\emptyset, H_{K+1}) \equiv Y$ throughout. Then, for $k = K, \ldots, 1$, define

$$m^{\mathbf{d}}_k(a_k, h_k) = E_{\mathbf{Q}}\left[m^{\mathbf{d}}_{k+1}\{d_{k+1}(H_{k+1}), H_{k+1}\} \mid A_k = a_k, H_k = h_k\right].$$

Under the conditions above, the longitudinal $g$-computation formula [73] identifies $\Psi^{\mathbf{d}}$ through

$$\psi^{\mathbf{d}}(\mathbf{Q}) := E_{\mathbf{Q}}\left[m^{\mathbf{d}}_1\{d_1(H_1), H_1\}\right],$$

which serves as the observed-data target estimand for statistical estimation. The EIC of $\psi^{\mathbf{d}}(\mathbf{Q})$ at $P_{\mathbf{Q},\mathbf{g}}$ is

$$\begin{aligned} & D^{\mathbf{d}}(P_{\mathbf{Q},\mathbf{g}})(O) \\ = \; & m^{\mathbf{d}}_1\{d_1(H_1), H_1\} - \psi^{\mathbf{d}}(\mathbf{Q}) \\ & + \sum_{k=1}^{K} \frac{\prod_{s=1}^{k} I\{A_s = d_s(H_s)\}}{\prod_{s=1}^{k} g_s\{d_s(H_s) \mid H_s\}} \left[m^{\mathbf{d}}_{k+1}\{d_{k+1}(H_{k+1}), H_{k+1}\} - m^{\mathbf{d}}_k(A_k, H_k)\right]. \end{aligned}$$

### 3.4.2 ADL-LTMLE Procedure

The ADL-LTMLE is constructed recursively as follows.

First, we initialize the final pseudo-outcome by $m^{\mathbf{d},*}_{n,K+1}\{d_{K+1}(H_{K+1}), H_{K+1}\} \equiv Y$. Then, for each $k = K, \ldots, 1$, we proceed as follows.

At stage $k$, we use the targeted regression $m^{\mathbf{d},*}_{n,k+1}$ obtained at stage $k+1$ to construct an estimate $m^{\mathbf{d}}_{n,k}(a_k, h_k)$ by regressing $m^{\mathbf{d},*}_{n,k+1}\{d_{k+1}(H_{k+1}), H_{k+1}\}$ on $A_k$ and $H_k$.

We then perform a targeting step at stage $k$. Using the average-reference cumulative treatment probability $\bar{G}^{\mathbf{d}}_{k,n}$ defined above, let $H^{\mathbf{d}}_{k,n}(O) = \frac{\prod_{s=1}^{k} I\{A_s = d_s(H_s)\}}{\bar{G}^{\mathbf{d}}_{k,n}(H_k)}$. We fluctuate $m^{\mathbf{d}}_{n,k}$ along a logistic model:

$$\operatorname{logit} m^{\mathbf{d}}_{n,k,\epsilon}(A_k, H_k) = \operatorname{logit} m^{\mathbf{d}}_{n,k}(A_k, H_k) + \epsilon H^{\mathbf{d}}_{k,n}(O), \tag{3.9}$$

where we use $m^{\mathbf{d},*}_{n,k+1}\{d_{k+1}(H_{k+1}), H_{k+1}\}$ as the outcome. Let $\hat{\epsilon}_{k,n}$ denote the estimated fluctuation parameter and set $m^{\mathbf{d},*}_{n,k} = m^{\mathbf{d}}_{n,k,\hat{\epsilon}_{k,n}}$. The targeted prediction $m^{\mathbf{d},*}_{n,k}\{d_k(H_k), H_k\}$ then becomes the pseudo-outcome for estimating the regression at stage $k-1$.

After completing the recursion, let $\mathbf{m}^{\mathbf{d},*}_n = \left(m^{\mathbf{d},*}_{n,1}, \ldots, m^{\mathbf{d},*}_{n,K}\right)$ denote the resulting collection of targeted sequential regressions, and let $Q^*_{1,n}$ denote the empirical distribution of $L_1$. We use $\mathbf{Q}^*_n = \left(Q^*_{1,n}, Q^*_{2,n}, \ldots, Q^*_{K+1,n}\right)$ to denote the longitudinal distribution that induces the sequential regressions $\mathbf{m}^{\mathbf{d},*}_n$. The resulting ADL-LTMLE is

$$\psi^{\mathbf{d}}(\mathbf{Q}^*_n) := \frac{1}{n}\sum_{i=1}^{n} m^{\mathbf{d},*}_{n,1}\{d_1(H_{1,i}), H_{1,i}\}.$$

### 3.4.3 Asymptotic Normality

We now state the asymptotic normality of the ADL-LTMLE. For any integrable function $f$, we use $P_{\mathbf{Q}_0,\mathbf{g}_i} f = E\{f(O_i) \mid \mathcal{F}_{i-1}\}$ to denote the conditional expectation under the participant-specific randomization mechanism $\mathbf{g}_i$, given the previously observed data. Define an i.i.d. reference law by the average of the participant-specific observed-data laws across participants in the adaptive experiment, i.e., $\bar{P}_{\mathbf{Q},\mathbf{g}_{1:n}} := \frac{1}{n}\sum_{i=1}^{n} P_{\mathbf{Q},\mathbf{g}_i}$, which serves as the reference law for characterizing the semiparametric efficiency bound for regular estimators under adaptive experiments and for constructing the ADL-LTMLE. Moreover, we write $\bar{P}_{\mathbf{Q},\mathbf{g}_{1:n}} = P_{\mathbf{Q},\bar{\mathbf{g}}_n}$ as shorthand, where $\bar{\mathbf{g}}_n = (\bar{g}_{1,n}, \ldots, \bar{g}_{K,n})$ denotes the longitudinal treatment mechanism induced by the average-reference law. In particular, $\bar{g}_{k,n}(a_k \mid h_k) := \frac{\bar{G}_{k,n}(\bar{a}_k, \bar{h}_k)}{\bar{G}_{k-1,n}(\bar{a}_{k-1}, \bar{h}_{k-1})}$, where $\bar{G}_{k,n}(\bar{a}_k, \bar{h}_k) := \frac{1}{n}\sum_{i=1}^{n}\prod_{s=1}^{k} g_{s,i}(a_s \mid h_s)$, and $\bar{G}_{0,n} := 1$. Further details are provided in Section 3.8.1.

Let $\mathbf{Q}_\infty$ and $\bar{\mathbf{g}}_\infty$ denote fixed limits of $\mathbf{Q}^*_n$ and $\bar{\mathbf{g}}_n$, respectively, and define $G^{\bar{\mathbf{g}}_\infty,\mathbf{d}}_k(h_k) := \prod_{s=1}^{k} \bar{g}_{s,\infty}\{d_s(h_s) \mid h_s\}$. Beyond the positivity and variance stabilization assumptions below, we additionally impose entropy and convergence conditions governing the relevant function classes and the asymptotic behavior of $\mathbf{Q}^*_n$ and $\bar{\mathbf{g}}_n$; their formal statements are provided in Assumptions 3.5–3.7 of Section 3.8.1.

**Assumption 3.2** (Strong Positivity of Average Cumulative Treatment Probabilities)**.** *There exists* $\zeta > 0$ *such that, for every* $n$, $\min_{k\in[K]} \inf_{h_k \in \mathcal{H}_k} \bar{G}^{\mathbf{d}}_{k,n}(h_k) \geq \zeta$,
*and* $\min_{k\in[K]} \inf_{h_k \in \mathcal{H}_k} G^{\bar{\mathbf{g}}_\infty,\mathbf{d}}_k(h_k) \geq \zeta$.

**Assumption 3.3** (Variance Stabilization)**.** *There exists a constant* $\sigma^2_{\mathbf{d}} > 0$ *such that* $\frac{1}{n}\sum_{i=1}^{n} \operatorname{Var}_{P_{\mathbf{Q}_0,\mathbf{g}_i}}\left[D^{\mathbf{d}}(P_{\mathbf{Q}_\infty,\bar{\mathbf{g}}_\infty})(O)\right] \xrightarrow{p} \sigma^2_{\mathbf{d}}$.

**Theorem 3.1** (Asymptotic Normality of the ADL-LTMLE)**.** *Suppose Assumptions 3.2, 3.3, 3.5, 3.6, and 3.7 hold. Then*

$$\sqrt{n}\left\{\psi^{\mathbf{d}}(\mathbf{Q}_n^*) - \psi^{\mathbf{d}}(\mathbf{Q}_0)\right\} \xrightarrow{d} N(0, \sigma_{\mathbf{d}}^2).$$

Let $\hat{\sigma}_{\mathbf{d},n}^2$ be a consistent estimate of $\sigma_{\mathbf{d}}^2$. The standard error of $\psi^{\mathbf{d}}(\mathbf{Q}_n^*)$ is given by $\hat{\sigma}_{\mathbf{d},n}/\sqrt{n}$. The $(1-\alpha)$ Wald confidence interval is $\psi^{\mathbf{d}}(\mathbf{Q}_n^*) \pm z_{1-\alpha/2}\hat{\sigma}_{\mathbf{d},n}/\sqrt{n}$, where $z_{1-\alpha/2}$ is the $(1-\alpha/2)$-quantile of the standard normal distribution.

### 3.4.4 Semiparametric Efficiency

We further establish the semiparametric efficiency of the ADL-LTMLE following the efficiency result of [92], which relates the semiparametric efficiency notion for the adaptive experiment $P_{\mathbf{Q},\mathbf{g}_{1:n}}$ to that of a limiting i.i.d. reference experiment with law $P_{\mathbf{Q},\bar{\mathbf{g}}_\infty}$.

For any $\mathbf{Q} = (Q_1, \ldots, Q_{K+1}) \in \mathcal{Q}$, define the tangent space

$$\mathcal{H}(\mathbf{Q}, \bar{\mathbf{g}}_\infty) = \left\{ h = \sum_{k=1}^{K+1} h_k \in L^2(P_{\mathbf{Q},\bar{\mathbf{g}}_\infty}) : \begin{array}{l} E_{\mathbf{Q}}\{h_1(L_1)\} = 0, \\ E_{\mathbf{Q}}\{h_{k+1}(L_{k+1}, A_k, H_k) \mid A_k, H_k\} = 0, \quad k \in [K] \end{array} \right\},$$

equipped with the inner product $\langle h, \widetilde{h}\rangle_{\mathcal{H}(\mathbf{Q},\bar{\mathbf{g}}_\infty)} = P_{\mathbf{Q},\bar{\mathbf{g}}_\infty}(h\widetilde{h})$.

**Assumption 3.4** (Reference-Law Norm Stability)**.** *For every $h \in \mathcal{H}(\mathbf{Q}_0, \bar{\mathbf{g}}_\infty)$, $P_{\mathbf{Q}_0,\bar{\mathbf{g}}_n} h^2 \xrightarrow{p} P_{\mathbf{Q}_0,\bar{\mathbf{g}}_\infty} h^2$.*

**Theorem 3.2** (Semiparametric Efficiency of the ADL-LTMLE)**.** *Suppose Assumptions 3.2, 3.3, 3.4, 3.5, 3.6, and 3.7 hold, and $\mathbf{Q}_\infty = \mathbf{Q}_0$. Then the ADL-LTMLE is regular and efficient,*

$$\sqrt{n}\left\{\psi^{\mathbf{d}}(\mathbf{Q}_n^*) - \psi^{\mathbf{d}}(\mathbf{Q}_0)\right\} \xrightarrow{d} N\left(0, P_{\mathbf{Q}_0,\bar{\mathbf{g}}_\infty}\left[D^{\mathbf{d}}(P_{\mathbf{Q}_0,\bar{\mathbf{g}}_\infty})^2\right]\right),$$

*i.e., the asymptotic variance of $\sqrt{n}\{\psi^{\mathbf{d}}(\mathbf{Q}_n^*) - \psi^{\mathbf{d}}(\mathbf{Q}_0)\}$ equals the minimum attainable asymptotic variance among all regular estimators.*

## 3.5 Application to Multiple Time-to-Treatment-Initiation Effects

This section illustrates the general framework developed in Sections 3.3 using a class of time-to-treatment-initiation effects as a motivating example. These effects compare initiating treatment at a given stage with delaying initiation until the subsequent stage. Such comparisons reflect clinically relevant decision problems in many settings, such as cancer and HIV treatments (see, e.g., [17, 95, 90]).

**Time-to-treatment-initiation effects.** Consider a class of treatment rules that vary in the timing of treatment initiation:

$$\mathbf{d}^\tau = (\underbrace{0,\dots,0}_{\tau-1},\underbrace{1,\dots,1}_{K-\tau+1}), \qquad \mathbf{d}^{K+1} = (0,\dots,0),$$

and consider the $\tau$-specific causal estimand $\Psi^\tau(\mathbf{Q}) = E(Y^{\mathbf{d}^\tau}) - E(Y^{\mathbf{d}^{\tau+1}})$, which compares initiating treatment at stage $\tau$ with delaying initiation until stage $\tau+1$. For $\tau = K$, this compares initiating treatment at the final stage $K$ with never initiating treatment. Under the identification conditions in Assumption 3.1, the target estimands of interest are

$$\psi^\tau(\mathbf{Q}) = E\left[m_1^{\mathbf{d}^\tau}\{d_1^\tau(H_1), H_1\} - m_1^{\mathbf{d}^{\tau+1}}\{d_1^{\tau+1}(H_1), H_1\}\right], \quad \tau \in [K].$$

**Longitudinal adaptive design.** For a treatment rule $\mathbf{d}$, define the stage-$k$ sequential-regression residual $\Delta_k^{\mathbf{d}}(O) = m_{k+1}^{\mathbf{d}}\{d_{k+1}(H_{k+1}), H_{k+1}\} - m_k^{\mathbf{d}}(A_k, H_k)$, where, for $k = K$, we set $m_{k+1}^{\mathbf{d}}\{d_{k+1}(H_{k+1}), H_{k+1}\} \equiv Y$. The EIC of $\psi^\tau$ admits the decomposition $D^\tau(P_{\mathbf{Q},\mathbf{g}}) = D_0^\tau(P_{\mathbf{Q},\mathbf{g}}) + \sum_{k=1}^K D_k^\tau(P_{\mathbf{Q},\mathbf{g}})$, where $D_0^\tau = m_1^{\mathbf{d}^\tau}\{d_1^\tau(H_1), H_1\} - m_1^{\mathbf{d}^{\tau+1}}\{d_1^{\tau+1}(H_1), H_1\} - \psi^\tau$, $D_k^\tau(P_{\mathbf{Q},\mathbf{g}})(O) = \frac{\Phi_k^\tau(\mathbf{Q})(O)}{\prod_{s=1}^k g_s(A_s|H_s)}$, and $\Phi_k^\tau(\mathbf{Q})(O) = I(\bar{A}_k = \mathbf{d}_{1:k}^\tau)\Delta_k^{\mathbf{d}^\tau}(O) - I(\bar{A}_k = \mathbf{d}_{1:k}^{\tau+1})\Delta_k^{\mathbf{d}^{\tau+1}}(O)$.

This illustrates that this class of estimands admits the general stage-wise representation required for applying the design optimization framework in Section 3.3.2. Specifically, for each stage $k$ and action $a$, let $\{\sigma_{a,k}^\tau(h_k)\}^2$ denote the conditional variance of the downstream EIC components from $D_k^\tau$ to $D_K^\tau$ after separating out $g_k(a \mid h_k)^{-1}$. Conditional on $H_k = h_k$, the component of the efficiency bound depending on $g_k$ is $\frac{\{\sigma_{1,k}^\tau(h_k)\}^2}{g_k(1|h_k)} + \frac{\{\sigma_{0,k}^\tau(h_k)\}^2}{g_k(0|h_k)}$. Using the backward-recursive strategy to optimize the longitudinal design, the oracle randomization scheme at stage $k = \tau$ is

$$g_k^*(1 \mid h_k) = \frac{\sigma_{1,k}^\tau(h_k)}{\sigma_{0,k}^\tau(h_k) + \sigma_{1,k}^\tau(h_k)}, \quad g_k^*(0 \mid h_k) = \frac{\sigma_{0,k}^\tau(h_k)}{\sigma_{0,k}^\tau(h_k) + \sigma_{1,k}^\tau(h_k)}. \tag{3.10}$$

By contrast, for stages $k \neq \tau$, the oracle design becomes degenerate:

$$g_k^*(1 \mid H_k) \equiv 0 \quad \text{for } k < \tau, \quad g_k^*(1 \mid H_k) \equiv 1 \quad \text{for } k > \tau.$$

Intuitively, the oracle design keeps all participants on the treatment history shared by the two regimes before stage $\tau$, randomizes at the stage where the regimes diverge according to the variance-minimizing allocation determined by the relative action-specific conditional variability, and assigns the common treatment prescribed by both regimes for later stages. As a result, it maximally preserves the longitudinal support needed to learn both treatment rules while randomizing treatment at the stage where they diverge according to the optimal allocation defined by the efficiency criterion for estimating their contrast.

However, when the goal is to learn multiple target estimands concurrently, such as the collection of treatment-initiation effects $\{\psi^\tau : \tau \in [K]\}$, an oracle design optimized for one estimand may substantially reduce the information available for others. These trade-offs are particularly substantial in longitudinal settings, where treatment is assigned sequentially and each stage-specific randomization decision directly shapes the data trajectories at subsequent stages. Consequently, favoring the optimal design for one estimand can irreversibly restrict the trajectories needed to estimate estimands that rely on sufficient data support at other stages, leading to limited support, inflated variance, or even loss of identification. This tension is explicit in the treatment-initiation-effect example, where estimand-specific oracle designs concentrate randomization at different stages and thereby impose competing demands on longitudinal data support. The joint optimization and the corresponding adaptive design framework in Section 3.3.3 therefore provide a principled approach to navigating these trade-offs.

## 3.6 Simulation Studies

We conduct simulation studies to evaluate the proposed longitudinal designs and estimators in the illustrative example of learning multiple treatment-initiation effects $(\psi^1, \ldots, \psi^K)$. Specifically, we assess the efficiency gains of the proposed adaptive designs developed in Section 3.3 relative to fixed non-adaptive randomization and the oracle design, as well as the finite-sample performance of the proposed estimators.

**Setup.** In the primary simulations, $n = 2000$ participants are enrolled sequentially in $T = 5$ equally sized cohorts of 400 participants. We consider longitudinal settings with either two or three treatment stages ($K = 2$ and $K = 3$), with discrete time-varying covariates and a continuous outcome. The complete data-generating mechanisms are provided in Section 3.8.2.

For each setting, we consider three user-specified weight vectors $w = (w^1, \ldots, w^K)$ in the joint efficiency criterion. For $K = 2$, we use $(1/2, 1/2)$, $(1/3, 2/3)$, and $(2/3, 1/3)$; for $K = 3$, we use $(1/3, 1/3, 1/3)$, $(1/6, 1/3, 1/2)$ and $(1/2, 1/3, 1/6)$.

We compare the following three designs: (a) a fixed non-adaptive design that assigns treatment with probability 0.5 at every stage; (b) an oracle design defined by recursively computing the optimal stage-specific randomization probabilities using the true data-generating mechanism; and (c) the proposed adaptive design, which starts from equal randomization and then estimates and tracks the oracle design using the accrued data.

**Estimation.** For each Monte Carlo run, we estimate the collection of treatment-initiation effects $\{\psi^\tau : \tau \in [K]\}$ by applying the ADL-LTMLE framework of Section 3.4 using the EIC corresponding to each contrast between two treatment-initiation rules, after rescaling the outcome to $[0, 1]$ using the sample minimum and maximum. In the primary simulations, the initial sequential outcome regressions are estimated using empirical means within strata defined by the observed covariate and treatment histories available before each treatment stage.

As a comparator, we also implement an adaptive-design LTMLE (AD-LTMLE) based on the participant-specific randomization mechanisms $\mathbf{g}_i$, which solves $\frac{1}{n}\sum_{i=1}^{n} D^{\tau}\left(P_{\mathbf{Q}_n^*,\mathbf{g}_i}\right)(O_i) = 0$. This estimator is asymptotically normal under standard conditions for adaptive experiments, but it may exhibit greater finite-sample variability and require stronger design-stability conditions than the ADL-LTMLE based on the design $\bar{\mathbf{g}}_n$ induced by the average-reference law $\bar{P}_{\mathbf{Q},\mathbf{g}_{1:n}}$ [92]. The variance estimators for the ADL-LTMLE and AD-LTMLE are $\frac{1}{n}\sum_{t=1}^{T}\frac{n_t}{n}\widehat{\mathrm{Var}}_{i\in\mathcal{I}_t} D^{\tau}\left(P_{\mathbf{Q}_n^*,\bar{\mathbf{g}}_n}\right)$ and $\frac{1}{n}\sum_{t=1}^{T}\frac{n_t}{n}\widehat{\mathrm{Var}}_{i\in\mathcal{I}_t} D^{\tau}\left(P_{\mathbf{Q}_n^*,\mathbf{g}_i}\right)$, respectively, where $\mathcal{I}_t$ denotes the set of participant indices in cohort $t$, and $n_t$ denotes the corresponding cohort size. Under the non-adaptive and oracle designs, where the randomization strategy is fixed, the two estimators coincide with the standard i.i.d. LTMLE [44]. We also evaluate two one-step estimators constructed using $\bar{\mathbf{g}}_n$ and $\mathbf{g}_i$, respectively, as analogues of ADL-LTMLE and AD-LTMLE, with more details provided in the Supplementary Material.

**Results.** Tables 3.1 and 3.2 summarize the results over 500 Monte Carlo replicates for $K = 2$ and $K = 3$, respectively. The columns $\mathrm{Var}_j$ report the empirical variances of the estimators of $\psi^j$, whereas TarVar reports the corresponding weighted average of these variances using the weight vector specified in the same row. The columns $\mathrm{Cov}_j$ report the empirical coverage of 95% Wald confidence intervals, whereas $\mathrm{OCov}_j$ report the coverage obtained by using the Monte Carlo standard deviation of the point estimates to construct the confidence intervals.

For $K = 2$, the proposed adaptive design yields substantial efficiency gains relative to the fixed design under both estimators. For $w = (1/2, 1/2)$, the values of TarVar under the fixed design, adaptive design with AD-LTMLE, adaptive design with ADL-LTMLE, and the oracle design are 0.029, 0.022, 0.019, and 0.017, respectively. For the other weights, the corresponding values are 0.026, 0.019, 0.017, and 0.015 for $w = (1/3, 2/3)$, and 0.033, 0.025, 0.021, and 0.019 for $w = (2/3, 1/3)$. Thus, even when paired with the AD-LTMLE, the adaptive design reduces TarVar by approximately 24%–26% relative to the fixed design. Under the same adaptive design, the ADL-LTMLE further reduces TarVar, resulting in overall reductions of approximately 34%–36% relative to the fixed design and bringing performance closer to that of the oracle design, which yields variance reduction by approximately 41%–42%.

The efficiency gains of the proposed designs are more substantial for $K = 3$. For $w = (1/3, 1/3, 1/3)$, the TarVar values under the fixed design, adaptive design with AD-LTMLE, adaptive design with ADL-LTMLE, and oracle design are 0.056, 0.031, 0.026, and 0.022, respectively. The corresponding values are 0.062, 0.032, 0.024, and 0.021 for $w = (1/6, 1/3, 1/2)$, and 0.050, 0.029, 0.024, and 0.021 for $w = (1/2, 1/3, 1/6)$. Pairing the proposed adaptive design with the AD-LTMLE therefore reduces TarVar by approximately 42%–48% relative to the fixed design. The ADL-LTMLE provides a further variance reduction under the same adaptive design, yielding overall reductions of approximately 52%–61%. The resulting performance is close to the corresponding oracle benchmark, which reduces TarVar by approximately 58%–66%.

The results also show that changing the weight vector shifts estimation efficiency toward the estimands assigned larger weights. For $K = 2$, under the oracle design, increasing the

weight on $\psi^2$ from $w = (2/3, 1/3)$ to $w = (1/3, 2/3)$ reduces $\text{Var}_2$ from 0.012 to 0.010, while $\text{Var}_1$ increases slightly from 0.023 to 0.024. The trade-offs are more pronounced for $K = 3$. Shifting from equal weights to $w = (1/6, 1/3, 1/2)$ prioritizes $\psi^3$ and reduces $\text{Var}_3$ from 0.020 to 0.017 in the oracle design, while $\text{Var}_1$ increases from 0.020 to 0.026. Conversely, shifting from equal weights to $w = (1/2, 1/3, 1/6)$ prioritizes $\psi^1$ and reduces $\text{Var}_1$ from 0.020 to 0.016, while $\text{Var}_3$ increases from 0.020 to 0.025. The adaptive designs follow the same qualitative pattern and achieve performance close to that of the oracle design.

Supplementary Tables 3.5–3.8 in Section 3.8 summarize sensitivity analyses with respect to the sample size and the model specification. For both $K = 2$ and $K = 3$, TarVar decreases substantially as the total sample size increases from 1,000 to 2,000 and then to 4,000. While misspecification of the outcome regressions increases variance, coverage remains near the nominal level across model specifications and sample sizes, supporting the robustness of ADL-LTMLE to outcome-regression misspecification.

Table 3.1: Simulation results for different longitudinal designs and estimators for $K = 2$. We report the empirical variance of the estimates of $\psi^1$ ($\text{Var}_1$) and $\psi^2$ ($\text{Var}_2$), the weighted sum of variances using user-specified weight $w$ (TarVar), the empirical 95% coverage probabilities ($\text{Cov}_1$, $\text{Cov}_2$), and the corresponding oracle coverage probabilities ($\text{OCov}_1$, $\text{OCov}_2$).

| $w$ | Design | Estimator | TarVar | $\text{Var}_1$ | $\text{Var}_2$ | $\text{Cov}_1$(%) | $\text{Cov}_2$(%) | $\text{OCov}_1$(%) | $\text{OCov}_2$(%) |
|---|---|---|---|---|---|---|---|---|---|
| (1/2, 1/2) | Non-adaptive | LTMLE | 0.029 | 0.040 | 0.018 | 95.0 | 97.0 | 95.4 | 95.6 |
| | Adaptive | AD-LTMLE | 0.022 | 0.030 | 0.013 | 92.6 | 94.4 | 94.8 | 94.2 |
| | Adaptive | ADL-LTMLE | 0.019 | 0.026 | 0.012 | 93.0 | 94.6 | 94.2 | 95.4 |
| | Oracle | LTMLE | 0.017 | 0.024 | 0.011 | 94.2 | 94.6 | 95.0 | 95.0 |
| (1/3, 2/3) | Non-adaptive | LTMLE | 0.026 | 0.040 | 0.018 | 95.0 | 97.0 | 95.4 | 95.6 |
| | Adaptive | AD-LTMLE | 0.019 | 0.032 | 0.013 | 92.0 | 96.0 | 94.2 | 96.6 |
| | Adaptive | ADL-LTMLE | 0.017 | 0.028 | 0.012 | 92.6 | 95.6 | 95.0 | 96.0 |
| | Oracle | LTMLE | 0.015 | 0.024 | 0.010 | 94.8 | 95.6 | 95.4 | 95.6 |
| (2/3, 1/3) | Non-adaptive | LTMLE | 0.033 | 0.040 | 0.018 | 95.0 | 97.0 | 95.4 | 95.6 |
| | Adaptive | AD-LTMLE | 0.025 | 0.030 | 0.014 | 92.2 | 94.4 | 95.4 | 95.2 |
| | Adaptive | ADL-LTMLE | 0.021 | 0.026 | 0.012 | 93.2 | 94.8 | 95.4 | 95.6 |
| | Oracle | LTMLE | 0.019 | 0.023 | 0.012 | 94.6 | 94.6 | 95.2 | 94.8 |

## 3.7 Conclusions

In this work, we develop a general framework for constructing longitudinal adaptive experiments aimed at learning one or more statistical estimands with semiparametric efficient inference. We characterize oracle longitudinal designs that minimize semiparametric efficiency bounds using backward-recursive solutions of treatment randomization schemes across treatment stages, and propose an extended design optimization criterion that balances estimation efficiency of multiple estimands and navigates their trade-offs. This approach applies

Table 3.2: Simulation results for different designs and estimators for $K = 3$. We report the empirical variance of the estimates of $\psi^1$ ($\text{Var}_1$), $\psi^2$ ($\text{Var}_2$), and $\psi^3$ ($\text{Var}_3$), the weighted sum of variances using user-specified weight $w$ (TarVar), the empirical 95% coverage probabilities ($\text{Cov}_1$, $\text{Cov}_2$, $\text{Cov}_3$), and the corresponding oracle coverage probabilities ($\text{OCov}_1$, $\text{OCov}_2$, $\text{OCov}_3$).

| $w$ | Design | Estimator | TarVar | $\text{Var}_1$ | $\text{Var}_2$ | $\text{Var}_3$ | $\text{Cov}_1$(%) | $\text{Cov}_2$(%) | $\text{Cov}_3$(%) | $\text{OCov}_1$(%) | $\text{OCov}_2$(%) | $\text{OCov}_3$(%) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| (1/3, 1/3, 1/3) | Non-adaptive | LTMLE | 0.056 | 0.026 | 0.079 | 0.063 | 94.2 | 94.6 | 96.2 | 94.0 | 94.2 | 95.4 |
| | Adaptive | AD-LTMLE | 0.031 | 0.022 | 0.039 | 0.033 | 95.2 | 94.6 | 94.6 | 95.4 | 94.8 | 94.0 |
| | Adaptive | ADL-LTMLE | 0.026 | 0.020 | 0.031 | 0.026 | 95.2 | 94.4 | 95.2 | 95.2 | 95.0 | 95.4 |
| | Oracle | LTMLE | 0.022 | 0.020 | 0.025 | 0.020 | 95.0 | 94.8 | 96.4 | 94.8 | 94.0 | 94.6 |
| (1/6, 1/3, 1/2) | Non-adaptive | LTMLE | 0.062 | 0.026 | 0.079 | 0.063 | 94.2 | 94.6 | 96.2 | 94.0 | 94.2 | 95.4 |
| | Adaptive | AD-LTMLE | 0.032 | 0.027 | 0.038 | 0.030 | 94.4 | 95.4 | 94.2 | 94.4 | 95.4 | 94.8 |
| | Adaptive | ADL-LTMLE | 0.024 | 0.024 | 0.028 | 0.022 | 94.4 | 95.6 | 95.2 | 94.8 | 95.0 | 94.8 |
| | Oracle | LTMLE | 0.021 | 0.026 | 0.024 | 0.017 | 95.8 | 95.8 | 97.2 | 95.4 | 94.4 | 96.2 |
| (1/2, 1/3, 1/6) | Non-adaptive | LTMLE | 0.050 | 0.026 | 0.079 | 0.063 | 94.2 | 94.6 | 96.2 | 94.0 | 94.2 | 95.4 |
| | Adaptive | AD-LTMLE | 0.029 | 0.019 | 0.039 | 0.038 | 95.0 | 93.4 | 93.0 | 95.0 | 93.6 | 94.4 |
| | Adaptive | ADL-LTMLE | 0.024 | 0.017 | 0.031 | 0.029 | 94.8 | 94.2 | 94.6 | 95.4 | 94.4 | 95.4 |
| | Oracle | LTMLE | 0.021 | 0.016 | 0.026 | 0.025 | 96.2 | 95.6 | 95.2 | 95.0 | 95.0 | 94.4 |

broadly to the identification of oracle designs and the construction of adaptive designs for a wide class of target estimands, including various timing-of-treatment-initiation effects, treatment-specific means at multiple time points, and mean outcomes under dynamic or stochastic treatment rules. We further establish an ADL-LTMLE framework for asymptotically normal and semiparametric efficient estimation of statistical estimands from dependent data collected from adaptive experiments, without relying on parametric model assumptions. Simulation studies show that the proposed longitudinal designs achieve substantial efficiency gains over non-adaptive designs with the performance close to that of the unknown oracle design, especially when combined with the efficient ADL-LTMLE.

Overall, this work contributes a general efficiency-oriented framework for the design and estimation of adaptive experiments in longitudinal settings involving multi-stage, time-varying treatments. Future work may explore applications of the proposed framework to other statistical learning objectives and extend it to a broader range of experimental settings and application domains.

## 3.8 Supplementary Material

### 3.8.1 Theoretical Analyses of ADL-LTMLE

This section provides theoretical analyses of ADL-LTMLE. We first characterize the treatment randomization mechanisms induced by the reference distribution used to construct the estimator, then establish its asymptotic normality and semiparametric efficiency.

**Stage-wise treatment distribution under the reference distribution.** Recall that $\bar{P}_{\mathbf{Q},\mathbf{g}_{1:n}} := \frac{1}{n}\sum_{i=1}^{n} P_{\mathbf{Q},\mathbf{g}_i}$, where $\mathbf{g}_i = (g_{1,i}, \ldots, g_{K,i})$ denotes the collection of randomization mechanisms used to generate participant $i$'s treatment sequence. Here, $\mathbf{Q} = (Q_1, Q_2, \ldots, Q_{K+1})$ collects the stage-wise nuisance components, with densities denoted by $q_1, \ldots, q_{K+1}$. The density of $\bar{P}_{\mathbf{Q},\mathbf{g}_{1:n}}$ is

$$\bar{p}_{\mathbf{Q},\mathbf{g}_{1:n}}(o) = \frac{1}{n}\sum_{i=1}^{n}\left[q_1(l_1)\prod_{k=1}^{K} g_{k,i}(a_k \mid h_k) q_{k+1}(l_{k+1} \mid a_k, h_k)\right].$$

For any fixed longitudinal treatment mechanism $\mathbf{g} = (g_1, \ldots, g_K)$, define its cumulative treatment probability through stage $k$ by $G_k^{\mathbf{g}}(\bar{a}_k \mid h_k) := \prod_{s=1}^{k} g_s(a_s \mid h_s)$. For participant $i$, write

$$G_{k,i}(\bar{a}_k \mid h_k) := G_k^{\mathbf{g}_i}(\bar{a}_k \mid h_k) = \prod_{s=1}^{k} g_{s,i}(a_s \mid h_s).$$

For any $1 \le k \le K$, the joint density of $(\bar{L}_k, \bar{A}_k)$ under $\bar{P}_{\mathbf{Q},\mathbf{g}_{1:n}}$, obtained by marginalizing over future variables, is

$$p_{\bar{P}}(\bar{l}_k, \bar{a}_k) = \left[q_1(l_1)\prod_{s=2}^{k} q_s(l_s \mid a_{s-1}, h_{s-1})\right]\left[\frac{1}{n}\sum_{i=1}^{n} G_{k,i}(\bar{a}_k \mid h_k)\right].$$

Similarly,

$$p_{\bar{P}}(\bar{l}_k, \bar{a}_{k-1}) = \left[q_1(l_1)\prod_{s=2}^{k} q_s(l_s \mid a_{s-1}, h_{s-1})\right]\left[\frac{1}{n}\sum_{i=1}^{n} G_{k-1,i}(\bar{a}_{k-1} \mid h_{k-1})\right]$$

for $2 \le k \le K$. For $k = 1$, $p_{\bar{P}}(\bar{l}_k, \bar{a}_{k-1})$ reduces to $p_{\bar{P}}(\bar{l}_k)$, which equals $q_1(l_1)$.

Define the cumulative treatment probability under the finite-sample average-reference distribution by $\bar{G}_{k,n}(\bar{a}_k \mid h_k) := \frac{1}{n}\sum_{i=1}^{n} G_{k,i}(\bar{a}_k \mid h_k)$. Therefore, the stage-wise treatment distribution induced by the average-reference law for every $k \in [K]$ is

$$\bar{g}_{k,n}(a_k \mid h_k) = \frac{p_{\bar{P}}(\bar{l}_k, \bar{a}_k)}{p_{\bar{P}}(\bar{l}_k, \bar{a}_{k-1})} = \frac{\bar{G}_{k,n}(\bar{a}_k \mid h_k)}{\bar{G}_{k-1,n}(\bar{a}_{k-1} \mid h_{k-1})}. \tag{3.11}$$

In particular, $\bar{g}_{1,n}(a_1 \mid h_1) = \frac{1}{n}\sum_{i=1}^{n} g_{1,i}(a_1 \mid h_1)$.

Let $\bar{\mathbf{g}}_n := (\bar{g}_{1,n}, \ldots, \bar{g}_{K,n})$. By construction,

$$G_k^{\bar{\mathbf{g}}_n}(\bar{a}_k \mid h_k) = \bar{G}_{k,n}(\bar{a}_k \mid h_k) = \frac{1}{n}\sum_{i=1}^{n} G_k^{\mathbf{g}_i}(\bar{a}_k \mid h_k).$$

Thus, the cumulative treatment probability induced by the average-reference distribution $\bar{P}_{\mathbf{Q},\mathbf{g}_{1:n}}$ is the average of the participant-specific cumulative treatment probabilities. For notational convenience in the following sections, we therefore write $P_{\mathbf{Q},\bar{\mathbf{g}}_n} := \bar{P}_{\mathbf{Q},\mathbf{g}_{1:n}}$, which defines a fixed reference experiment with independent observations drawn from $P_{\mathbf{Q},\bar{\mathbf{g}}_n}$ where each participant has stage-wise treatments sequentially generated according to the fixed mechanism $\bar{\mathbf{g}}_n$.

**Asymptotic normality.** We use the example of the mean outcome under a longitudinal dynamic treatment rule $\mathbf{d} = (d_1, \ldots, d_K)$ as the target estimand of interest, which falls within the broader ADL-TMLE framework of [92] that considers general target parameters of the form $\Psi(\mathbf{Q})$ for data likelihoods factorizing as $p(o_1, \ldots, o_n) = \prod_{i=1}^{n} \mathbf{q}(o_i)\mathbf{g}_i(o_i)$, where $\mathbf{q}$ is the density of some $\mathbf{Q} \in \mathcal{Q}$ that defines the target estimand.

Let $\psi^{\mathbf{d}}(P_{\mathbf{Q},\mathbf{g}}) = E_{P_{\mathbf{Q},\mathbf{g}}}(Y^{\mathbf{d}})$ denote the mean counterfactual outcome under $\mathbf{d}$. Since the estimand is only defined on $P_{\mathbf{Q},\mathbf{g}}$ through $\mathbf{Q}$, we also express the estimand by $\psi^{\mathbf{d}}(\mathbf{Q})$.

**Theorem 3.3** (Decomposition of the ADL-LTMLE)**.**

$$\psi^{\mathbf{d}}(\mathbf{Q}_n^*) - \psi^{\mathbf{d}}(\mathbf{Q}_0) = M_{1,n}^{\mathbf{d}}(\mathbf{Q}_\infty, \bar{\mathbf{g}}_\infty) + M_{2,n}^{\mathbf{d}}(\mathbf{Q}_n^*, \mathbf{Q}_\infty, \bar{\mathbf{g}}_n) + M_{3,n}^{\mathbf{d}}(\mathbf{Q}_\infty, \bar{\mathbf{g}}_n, \bar{\mathbf{g}}_\infty),$$

*where*

$$M_{1,n}^{\mathbf{d}}(\mathbf{Q}_\infty, \bar{\mathbf{g}}_\infty) := \frac{1}{n}\sum_{i=1}^{n} \left[D^{\mathbf{d}}(P_{\mathbf{Q}_\infty,\bar{\mathbf{g}}_\infty})(O_i) - P_{\mathbf{Q}_0,\mathbf{g}_i} D^{\mathbf{d}}(P_{\mathbf{Q}_\infty,\bar{\mathbf{g}}_\infty})\right],$$

$$M_{2,n}^{\mathbf{d}}(\mathbf{Q}_n^*, \mathbf{Q}_\infty, \bar{\mathbf{g}}_n) := \frac{1}{n}\sum_{i=1}^{n} \Bigg\{ \left[D^{\mathbf{d}}(P_{\mathbf{Q}_n^*,\bar{\mathbf{g}}_n})(O_i) - P_{\mathbf{Q}_0,\mathbf{g}_i} D^{\mathbf{d}}(P_{\mathbf{Q}_n^*,\bar{\mathbf{g}}_n})\right] - \left[D^{\mathbf{d}}(P_{\mathbf{Q}_\infty,\bar{\mathbf{g}}_n})(O_i) - P_{\mathbf{Q}_0,\mathbf{g}_i} D^{\mathbf{d}}(P_{\mathbf{Q}_\infty,\bar{\mathbf{g}}_n})\right] \Bigg\},$$

*and*

$$M_{3,n}^{\mathbf{d}}(\mathbf{Q}_\infty, \bar{\mathbf{g}}_n, \bar{\mathbf{g}}_\infty) := \frac{1}{n}\sum_{i=1}^{n} \Bigg\{ \left[D^{\mathbf{d}}(P_{\mathbf{Q}_\infty,\bar{\mathbf{g}}_n})(O_i) - P_{\mathbf{Q}_0,\mathbf{g}_i} D^{\mathbf{d}}(P_{\mathbf{Q}_\infty,\bar{\mathbf{g}}_n})\right] - \left[D^{\mathbf{d}}(P_{\mathbf{Q}_\infty,\bar{\mathbf{g}}_\infty})(O_i) - P_{\mathbf{Q}_0,\mathbf{g}_i} D^{\mathbf{d}}(P_{\mathbf{Q}_\infty,\bar{\mathbf{g}}_\infty})\right] \Bigg\}.$$

*Proof.* Define the sequential regressions recursively by setting

$$m_{K+1}^{\mathbf{d}}\{d_{K+1}(H_{K+1}), H_{K+1}\} = Y,$$

and, for $k = 1, \ldots, K$,

$$m_k^{\mathbf{d}}(a_k, h_k) = E_{\mathbf{Q}}\left[m_{k+1}^{\mathbf{d}}\{d_{k+1}(H_{k+1}), H_{k+1}\} \mid A_k = a_k,\ H_k = h_k\right].$$

For any longitudinal treatment mechanism $\mathbf{g}$, define $G_k^{\mathbf{g},\mathbf{d}}(H_k) := \prod_{s=1}^{k} g_s\{d_s(H_s) \mid H_s\}$. For participant $i$, we write $G_{k,i}^{\mathbf{d}}(H_k) := G_k^{\mathbf{g}_i,\mathbf{d}}(H_k)$. The corresponding cumulative probability under the finite-sample average-reference mechanism is $\bar{G}_{k,n}^{\mathbf{d}}(H_k) := G_k^{\bar{\mathbf{g}}_n,\mathbf{d}}(H_k) = \frac{1}{n}\sum_{i=1}^{n} G_{k,i}^{\mathbf{d}}(H_k)$.

The treatment-rule-specific EIC evaluated under any fixed $P_{\mathbf{Q},\mathbf{g}}$ is

$$D^{\mathbf{d}}(P_{\mathbf{Q},\mathbf{g}})(O) = m_1^{\mathbf{d}}\{d_1(H_1), H_1\} - \psi^{\mathbf{d}}(\mathbf{Q}) + \sum_{k=1}^{K} \frac{I\{\bar{A}_k = \bar{d}_k\}}{G_k^{\mathbf{g},\mathbf{d}}(H_k)} \left[m_{k+1}^{\mathbf{d}}\{d_{k+1}(H_{k+1}), H_{k+1}\} - m_k^{\mathbf{d}}(A_k, H_k)\right].$$

Under the event $I\{\bar{A}_k = \bar{d}_k\} = 1$, the term $m_k^{\mathbf{d}}(A_k, H_k)$ equals $m_k^{\mathbf{d}}\{d_k(H_k), H_k\}$.

For participant $i$, define the cumulative probability of following $\mathbf{d}$ through stage $k$ under that participant's randomization mechanism by $G_{k,i}^{\mathbf{d}}(H_k) = \prod_{s=1}^{k} g_{s,i}\{d_s(H_s) \mid H_s\}$. We also define the stage-$k$ residual under $\mathbf{Q}$ by $\Delta_k^{\mathbf{d}}(\mathbf{Q}) := m_{k+1}^{\mathbf{d}}\{d_{k+1}(H_{k+1}), H_{k+1}\} - m_k^{\mathbf{d}}\{d_k(H_k), H_k\}$.

For each $i$, applying von Mises expansion with respect to $P_{\mathbf{Q},\bar{\mathbf{g}}_n}$ and $P_{\mathbf{Q}_0,\mathbf{g}_i}$ yields

$$\psi^{\mathbf{d}}(P_{\mathbf{Q},\bar{\mathbf{g}}_n}) - \psi^{\mathbf{d}}(P_{\mathbf{Q}_0,\mathbf{g}_i}) = \int D^{\mathbf{d}}(P_{\mathbf{Q},\bar{\mathbf{g}}_n})(o)\, d\{P_{\mathbf{Q},\bar{\mathbf{g}}_n} - P_{\mathbf{Q}_0,\mathbf{g}_i}\}(o) + R_i^{\mathbf{d}}(\mathbf{Q}, \mathbf{Q}_0; \bar{\mathbf{g}}_n, \mathbf{g}_i).$$

For the mean outcome under a longitudinal dynamic treatment rule $\mathbf{d}$, the remainder takes the form

$$R_i^{\mathbf{d}}(\mathbf{Q}, \mathbf{Q}_0; \bar{\mathbf{g}}_n, \mathbf{g}_i) = \sum_{k=1}^{K} E_{\mathbf{Q}_0}^{\mathbf{d}} \left[\frac{G_{k,i}^{\mathbf{d}}(H_k) - \bar{G}_{k,n}^{\mathbf{d}}(H_k)}{\bar{G}_{k,n}^{\mathbf{d}}(H_k)} \Delta_k^{\mathbf{d}}(\mathbf{Q})\right],$$

where $G_{k,i}^{\mathbf{d}}(H_k) = \prod_{s=1}^{k} g_{s,i}\{d_s(H_s) \mid H_s\}$, $\bar{G}_{k,n}^{\mathbf{d}}(H_k) = \frac{1}{n}\sum_{i=1}^{n} G_{k,i}^{\mathbf{d}}(H_k)$, and $\Delta_k^{\mathbf{d}}(\mathbf{Q}) = m_{k+1}^{\mathbf{d}}\{d_{k+1}(H_{k+1}), H_{k+1}\} - m_k^{\mathbf{d}}\{d_k(H_k), H_k\}$.

Here, $E_{\mathbf{Q}_0}^{\mathbf{d}}$ denotes expectation under the distribution that sets treatment according to $\mathbf{d}$ while keeping the longitudinal covariate and outcome law at $\mathbf{Q}_0$. By averaging over $i = 1, \ldots, n$, we obtain

$$\begin{aligned}
\psi^{\mathbf{d}}(\mathbf{Q}) - \psi^{\mathbf{d}}(\mathbf{Q}_0) &= \frac{1}{n}\sum_{i=1}^{n} \int D^{\mathbf{d}}(P_{\mathbf{Q},\bar{\mathbf{g}}_n})(o)\, d\{P_{\mathbf{Q},\bar{\mathbf{g}}_n} - P_{\mathbf{Q}_0,\mathbf{g}_i}\}(o) \\
&\quad + \frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{K} E_{\mathbf{Q}_0}^{\mathbf{d}} \left[\frac{G_{k,i}^{\mathbf{d}}(H_k) - \bar{G}_{k,n}^{\mathbf{d}}(H_k)}{\bar{G}_{k,n}^{\mathbf{d}}(H_k)} \Delta_k^{\mathbf{d}}(\mathbf{Q})\right] \\
&= \frac{1}{n}\sum_{i=1}^{n} \int D^{\mathbf{d}}(P_{\mathbf{Q},\bar{\mathbf{g}}_n})(o)\, d\{P_{\mathbf{Q},\bar{\mathbf{g}}_n} - P_{\mathbf{Q}_0,\mathbf{g}_i}\}(o) \\
&= -P_{\mathbf{Q}_0,\bar{\mathbf{g}}_n} D^{\mathbf{d}}(P_{\mathbf{Q},\bar{\mathbf{g}}_n}),
\end{aligned}$$

where the second equality holds since $\bar{G}_{k,n}^{\mathbf{d}}(H_k) = \frac{1}{n}\sum_{i=1}^{n} G_{k,i}^{\mathbf{d}}(H_k)$, and the last equality uses the fact that $P_{\mathbf{Q},\mathbf{g}} D^{\mathbf{d}}(P_{\mathbf{Q},\mathbf{g}}) = 0$ for any $P_{\mathbf{Q},\mathbf{g}}$.

Since an ADL-LTMLE for estimating $\psi^{\mathbf{d}}(\mathbf{Q}_0)$ updates an initial estimate $\mathbf{Q}_n$ to obtain a $\mathbf{Q}_n^*$ that solves $\frac{1}{n}\sum_{i=1}^n D^{\mathbf{d}}(P_{\mathbf{Q}_n^*,\bar{\mathbf{g}}_n})(O_i) = 0$, we have

$$\psi^{\mathbf{d}}(\mathbf{Q}_n^*) - \psi^{\mathbf{d}}(\mathbf{Q}_0) = \frac{1}{n}\sum_{i=1}^n D^{\mathbf{d}}(P_{\mathbf{Q}_n^*,\bar{\mathbf{g}}_n}) - P_{\mathbf{Q}_0,\bar{\mathbf{g}}_n} D^{\mathbf{d}}(P_{\mathbf{Q}_n^*,\bar{\mathbf{g}}_n}).$$

The final decomposition follows from some algebra that adds and subtracts other corresponding terms. □

**Theorem 3.4** (Asymptotic Normality of $M_{1,n}^{\mathbf{d}}$)**.** *Suppose Assumptions 3.2 and 3.3 hold, then* $\sqrt{n}M_{1,n}^{\mathbf{d}}(\mathbf{Q}_\infty, \bar{\mathbf{g}}_\infty) \xrightarrow{d} N(0, \sigma_{\mathbf{d}}^2)$.

*Proof.* Note that $M_{1,n}^{\mathbf{d}}(\mathbf{Q}_\infty, \bar{\mathbf{g}}_\infty)$ is an average of martingale differences sequences. Under Assumption 3.2, $D^{\mathbf{d}}(P_{\mathbf{Q}_\infty,\bar{\mathbf{g}}_\infty})$ is uniformly bounded, therefore satisfying the conditional Lindeberg condition, i.e., for every $\epsilon > 0$, $\frac{1}{n}\sum_{i=1}^n E\left[Z_{n,i}^2 I\left\{|Z_{n,i}| > \epsilon\sqrt{n}\right\} \mid \bar{O}(i-1)\right] \xrightarrow{p} 0$, where $Z_{n,i} = D(P_{\mathbf{Q}_\infty,\bar{\mathbf{g}}_\infty})(O_i) - P_{\mathbf{Q}_0,\mathbf{g}_i} D(P_{\mathbf{Q}_\infty,\bar{\mathbf{g}}_\infty})$.

Combined with the variance stabilization assumption stated in Assumption 3.3, one can directly apply Martingale Central Limit Theorem (see, e.g., [29]) and obtain the asymptotic normality. □

Now we establish equicontinuity conditions for $M_{2,n}^{\mathbf{d}}$ and $M_{3,n}^{\mathbf{d}}$.

For $\mathbf{Q} = (Q_1, \ldots, Q_{K+1}) \in \mathcal{Q}$, define the collection of sequential regression functions associated with rule $\mathbf{d}$ by $\mathbf{m}^{\mathbf{d}}(\mathbf{Q}) = \left(m_1^{\mathbf{d}}(\mathbf{Q}), \ldots, m_K^{\mathbf{d}}(\mathbf{Q})\right)$, and let $\mathcal{M}^{\mathbf{d}} = \left\{\mathbf{m}^{\mathbf{d}}(\mathbf{Q}) : \mathbf{Q} \in \mathcal{Q}\right\}$ equipped with the supremum norm defined by $\left\|\mathbf{m}^{\mathbf{d}}\right\|_\infty = \max_{k\in[K]} \left\|m_k^{\mathbf{d}}\right\|_\infty$. Let $\mathcal{G}$ denote the fixed class of treatment mechanisms satisfying Assumption 3.2, equipped with $\|\mathbf{g}\|_\infty = \max_{k\in[K]} \sup_{a_k,h_k} |g_k(a_k \mid h_k)|$. For a class of functions $\mathcal{H}$, let $N_\infty(u, \mathcal{H})$ denote its $u$-covering number under the corresponding supremum norm, and define the entropy integral $J_\infty(\epsilon, \mathcal{H}) := \int_0^\epsilon \sqrt{\log\left[1 + N_\infty(u, \mathcal{H})\right]}\, du$.

**Assumption 3.5** (Entropy Conditions)**.** *$J_\infty(\epsilon, \mathcal{M}^{\mathbf{d}})$ and $J_\infty(\epsilon, \mathcal{G})$ are finite and converge to zero as $\epsilon \to 0$.*

**Assumption 3.6** (Convergence of $\mathbf{Q}_n^*$)**.** *Define $\rho(x) = e^x - x - 1$ and $\widetilde{D}^{\mathbf{d}}(P_{\mathbf{Q},\bar{\mathbf{g}}_n})(O_i) = m_1^{\mathbf{d}}(\mathbf{Q})(O_i) + \sum_{k=1}^K D_k^{\mathbf{d}}(P_{\mathbf{Q},\bar{\mathbf{g}}_n})(O_i)$ for any $\mathbf{Q} \in \mathcal{Q}$. Let*

$$s_{2,n}^{\mathbf{d}} := \left\{\frac{1}{n}\sum_{i=1}^n E\left[\rho\left(\frac{\left|D^{\mathbf{d}}(P_{\mathbf{Q}_n^*,\bar{\mathbf{g}}_n})(O_i) - D^{\mathbf{d}}(P_{\mathbf{Q}_\infty,\bar{\mathbf{g}}_n})(O_i)\right|}{c_2}\right) \,\middle|\, \mathcal{F}_{i-1}\right]\right\}^{1/2},$$

*where $c_2 = 8K/\zeta + 8K^2/\zeta^2$. Assume that $s_{2,n}^{\mathbf{d}} = O_P(\delta_{2,n})$ for some $\delta_{2,n} \to 0$.*

**Assumption 3.7** (Convergence of $\bar{\mathbf{g}}_n$)**.** *Define*

$$s_{3,n}^{\mathbf{d}} := \left\{\frac{1}{n}\sum_{i=1}^n E\left[\rho\left(\frac{\left|D^{\mathbf{d}}(P_{\mathbf{Q}_\infty,\bar{\mathbf{g}}_n})(O_i) - D^{\mathbf{d}}(P_{\mathbf{Q}_\infty,\bar{\mathbf{g}}_\infty})(O_i)\right|}{c_3}\right) \,\middle|\, \mathcal{F}_{i-1}\right]\right\}^{1/2},$$

*where $c_3 = 4K(K+1)/\zeta^2$. Assume that $s_{3,n}^{\mathbf{d}} = O_P(\delta_{3,n})$ for some $\delta_{3,n} \to 0$.*

We also refer to the following lemma of a general martingale equicontinuity result of [93].

**Lemma 3.1.** *Let* $M_n f = \frac{1}{n}\sum_{i=1}^n [f(O_i) - E\{f(O_i) \mid \mathcal{F}_{i-1}\}]$. *Let* $\Theta$ *be an index set and let* $\{f_\theta : \theta \in \Theta\}$ *be a uniformly bounded class satisfying* $\|f_\theta - f_{\theta'}\|_\infty \leq L\|\theta - \theta'\|_\infty$. *Define the entropy integral as* $J_\infty(\epsilon, \Theta) = \int_0^\epsilon \sqrt{\log\{1 + N_\infty(u, \Theta)\}}\, du$. *Suppose that* $J_\infty(\epsilon, \Theta) < \infty$ *is finite and that* $J_\infty(\epsilon, \Theta) \to 0$ *as* $\epsilon \to 0$. *Let* $M = \sup_{\theta\in\Theta} \|\theta\|_\infty$, $Z = 8LM$, *and define* $s_n(f_n) = \left\{\frac{1}{n}\sum_{i=1}^n E\left[\rho\left(\frac{|f_n(O_i)|}{Z}\right) \,\middle|\, \mathcal{F}_{i-1}\right]\right\}^{1/2}$, *where* $\rho(x) = e^x - x - 1$.

*If* $s_n(f_n) = O_P(\delta_n)$ *for some* $\delta_n \to 0$, *then* $M_n f_n = o_P(n^{-1/2})$.

**Theorem 3.5** (Equicontinuity of $M^{\mathbf{d}}_{2,n}$ and $M^{\mathbf{d}}_{3,n}$)**.** *Suppose Assumptions 3.2, 3.5, 3.6, and 3.7 hold. Then* $M^{\mathbf{d}}_{2,n} = o_P(n^{-1/2})$, $M^{\mathbf{d}}_{3,n} = o_P(n^{-1/2})$.

*Proof.* For $M^{\mathbf{d}}_{2,n}$, define $\Theta^{\mathbf{d}}_2 = \mathcal{M}^{\mathbf{d}} \times \mathcal{G}$, and, for $\theta_2 = (\mathbf{m}^{\mathbf{d}}, \bar{\mathbf{g}}) \in \Theta^{\mathbf{d}}_2$, define $\|\theta_2\|_\infty = \max\left\{\|\mathbf{m}^{\mathbf{d}}\|_\infty, \|\bar{\mathbf{g}}\|_\infty\right\}$.

Let $\widetilde{D}^{\mathbf{d}}(P_{\mathbf{Q},\bar{\mathbf{g}}}) = m^{\mathbf{d}}_1(\mathbf{Q}) + \sum_{k=1}^K D^{\mathbf{d}}_k(P_{\mathbf{Q},\bar{\mathbf{g}}})$ and $f^{\mathbf{d}}_{2,\theta_2}(O) = \widetilde{D}^{\mathbf{d}}(P_{\mathbf{Q},\bar{\mathbf{g}}})(O) - \widetilde{D}^{\mathbf{d}}(P_{\mathbf{Q}_\infty,\bar{\mathbf{g}}})(O)$. Thus, $M^{\mathbf{d}}_{2,n} = \frac{1}{n}\sum_{i=1}^n \left[f^{\mathbf{d}}_{2,\theta^{\mathbf{d}}_{2,n}}(O_i) - P_{\mathbf{Q}_0,\mathbf{g}_i} f^{\mathbf{d}}_{2,\theta^{\mathbf{d}}_{2,n}}\right]$, which is a martingale process indexed by $\theta^{\mathbf{d}}_{2,n} = \left(\mathbf{m}^{\mathbf{d}}(\mathbf{Q}^*_n), \bar{\mathbf{g}}_n\right)$. A direct calculation of the longitudinal EIC under Assumption 3.2 yields $\left\|f^{\mathbf{d}}_{2,\theta_2} - f^{\mathbf{d}}_{2,\theta'_2}\right\|_\infty \leq L_2 \left\|\theta_2 - \theta'_2\right\|_\infty$, where $L_2 = \frac{K}{\zeta} + \frac{K^2}{\zeta^2}$.

For $M^{\mathbf{d}}_{3,n}$, define $\Theta^{\mathbf{d}}_3 = \mathcal{G}$ and $f^{\mathbf{d}}_{3,\bar{\mathbf{g}}}(O) = D^{\mathbf{d}}(P_{\mathbf{Q}_\infty,\bar{\mathbf{g}}})(O)$. We have that

$$M^{\mathbf{d}}_{3,n} = \frac{1}{n}\sum_{i=1}^n \left[\{f^{\mathbf{d}}_{3,\bar{\mathbf{g}}_n} - f^{\mathbf{d}}_{3,\bar{\mathbf{g}}_\infty}\}(O_i) - P_{\mathbf{Q}_0,\mathbf{g}_i}\{f^{\mathbf{d}}_{3,\bar{\mathbf{g}}_n} - f^{\mathbf{d}}_{3,\bar{\mathbf{g}}_\infty}\}\right],$$

where $\left\|f^{\mathbf{d}}_{3,\bar{\mathbf{g}}} - f^{\mathbf{d}}_{3,\bar{\mathbf{g}}'}\right\|_\infty \leq L_3 \left\|\bar{\mathbf{g}} - \bar{\mathbf{g}}'\right\|_\infty$, where $L_3 = \frac{K(K+1)}{2\zeta^2}$.

Moreover, since $N_\infty(u, \Theta^{\mathbf{d}}_2) \leq N_\infty(u, \mathcal{M}^{\mathbf{d}}) N_\infty(u, \mathcal{G})$, and $N_\infty(u, \Theta^{\mathbf{d}}_3) = N_\infty(u, \mathcal{G})$, Assumption 3.5 implies the required entropy conditions. Since both outcome regressions and treatment probabilities are bounded by 1, we have $c_2 = \frac{8K}{\zeta} + \frac{8K^2}{\zeta^2}$ and $c_3 = \frac{4K(K+1)}{\zeta^2}$.

Finally, Assumptions 3.6 and 3.7 provide the required convergence in Lemma 3.1. Applying the general martingale equicontinuity result separately to the two processes yields $M^{\mathbf{d}}_{2,n} = o_P(n^{-1/2})$ and $M^{\mathbf{d}}_{3,n} = o_P(n^{-1/2})$.

□

Combining Theorems 3.3, 3.4, and 3.5 yields the asymptotic normality of ADL-LTMLE 3.1: $\sqrt{n}\left\{\psi^{\mathbf{d}}(\mathbf{Q}^*_n) - \psi^{\mathbf{d}}(\mathbf{Q}_0)\right\} \xrightarrow{d} N(0, \sigma^2_{\mathbf{d}})$.

**Semiparametric efficiency.** The semiparametric efficiency analysis of the ADL-LTMLE builds directly on the ADL-TMLE framework of [92], which establishes efficiency through local asymptotic normality and the convolution theorem [82].

Write $q(o) = q_1(l_1)\prod_{k=1}^{K} q_{k+1}(l_{k+1} \mid a_k, h_k)$ and $g_i(o) = \prod_{k=1}^{K} g_{k,i}(a_k \mid h_k)$. Then the longitudinal adaptive likelihood has the general factorization $p^n_{\mathbf{Q},\mathbf{g}_{1:n}}(o_1, \ldots, o_n) = \prod_{i=1}^{n} q(o_i)g_i(o_i)$. Define a submodel $\{\mathbf{Q}^h_\epsilon : \epsilon\}$ through $\mathbf{Q}_0$ by

$$dQ^h_{1,\epsilon}(l_1) = \{1 + \epsilon h_1(l_1)\}dQ_{0,1}(l_1)$$

and

$$dQ^h_{k+1,\epsilon}(l_{k+1} \mid a_k, h_k) = \{1 + \epsilon h_{k+1}(l_{k+1}, a_k, h_k)\}\, dQ_{0,k+1}(l_{k+1} \mid a_k, h_k)$$

for $k \in [K]$.

Let $P^n_{\mathbf{Q}^h_\epsilon,\mathbf{g}_{1:n}}$ denote the joint law of $(O_1, \ldots, O_n)$ under $\mathbf{Q}^h_\epsilon$ and $\mathbf{g}_{1:n}$. Its density is factorized by

$$\begin{aligned} &p^n_{\mathbf{Q}^h_\epsilon,\mathbf{g}_{1:n}}(o_1, \ldots, o_n) \\ &\quad = \prod_{i=1}^{n}\left[q^h_{1,\epsilon}(l_{1,i})\prod_{k=1}^{K} g_{k,i}(a_{k,i} \mid h_{k,i}, \mathcal{F}_{i-1})q^h_{k+1,\epsilon}(l_{k+1,i} \mid a_{k,i}, h_{k,i})\right]. \end{aligned}$$

**Proposition 3.1** (Local Asymptotic Normality)**.** *Under Assumption 3.4, for every bounded score direction $h$ and fixed $t \in \mathbb{R}$,*

$$\log \frac{dP^n_{\mathbf{Q}^h_{t/\sqrt{n}},\mathbf{g}_{1:n}}}{dP^n_{\mathbf{Q}_0,\mathbf{g}_{1:n}}} = t\Delta_{n,h} - \frac{t^2}{2}P_{\mathbf{Q}_0,\bar{\mathbf{g}}_\infty}h^2 + o_P(1),$$

*where $\Delta_{n,h} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} h(O_i) \xrightarrow{d} N\left(0, P_{\mathbf{Q}_0,\bar{\mathbf{g}}_\infty}h^2\right)$.*

*Proof of Proposition 3.1.* For participant $i$, write $h_{1,i} = h_1(L_{1,i})$ and, for $k \in [K]$, $h_{k+1,i} = h_{k+1}(L_{k+1,i}, A_{k,i}, H_{k,i})$. Therefore, $h(O_i) = \sum_{k=1}^{K+1} h_{k,i}$.

The log-likelihood ratio can be written as

$$\log \frac{dP^n_{\mathbf{Q}^h_{t/\sqrt{n}},\mathbf{g}_{1:n}}}{dP^n_{\mathbf{Q}_0,\mathbf{g}_{1:n}}} = \sum_{i=1}^{n}\sum_{k=1}^{K+1}\log\left(1 + \frac{t}{\sqrt{n}}h_{k,i}\right).$$

Applying Taylor expansion, we have

$$\log \frac{dP^n_{\mathbf{Q}^h_{t/\sqrt{n}},\mathbf{g}_{1:n}}}{dP^n_{\mathbf{Q}_0,\mathbf{g}_{1:n}}} = \frac{t}{\sqrt{n}}\sum_{i=1}^{n}\sum_{k=1}^{K+1} h_{k,i} - \frac{t^2}{2n}\sum_{i=1}^{n}\sum_{k=1}^{K+1} h^2_{k,i} + o(1).$$

Define $P_{\mathbf{Q}_0,\mathbf{g}_i}f = E\{f(O_i) \mid \mathcal{F}_{i-1}\}$. Since the longitudinal score components are conditionally orthogonal, we have $P_{\mathbf{Q}_0,\mathbf{g}_i}h = 0$ and $P_{\mathbf{Q}_0,\mathbf{g}_i}(h_k h_s) = 0$ for $k \neq s$. Consequently, $P_{\mathbf{Q}_0,\mathbf{g}_i}h^2 = \sum_{k=1}^{K+1} P_{\mathbf{Q}_0,\mathbf{g}_i}h^2_k$.

Define $Z_{i,h} = \sum_{k=1}^{K+1} h^2_{k,i} - P_{\mathbf{Q}_0,\mathbf{g}_i}\left(\sum_{k=1}^{K+1} h^2_k\right)$. The variables $Z_{i,h}$ form a bounded martingale-difference sequence. Therefore, $\frac{1}{n}\sum_{i=1}^{n} Z_{i,h} = o_P(1)$.

Under Assumption 3.4, $\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{K+1} h_{k,i}^2 = P_{\mathbf{Q}_0,\bar{\mathbf{g}}_n}h^2 + o_P(1) \xrightarrow{p} P_{\mathbf{Q}_0,\bar{\mathbf{g}}_\infty}h^2$. Since the conditional Lindeberg condition holds with $h$ bounded, the martingale CLT gives $\Delta_{n,h} \xrightarrow{d} N\left(0, P_{\mathbf{Q}_0,\bar{\mathbf{g}}_\infty}h^2\right)$, which proves the proposition. □

For the submodel $\{\mathbf{Q}_\epsilon^h : \epsilon\}$ defined above, let $\dot{\psi}^{\mathbf{d}}_{\mathbf{Q}_0}(h) := \frac{d}{d\epsilon}\psi^{\mathbf{d}}(\mathbf{Q}_\epsilon^h)\big|_{\epsilon=0}$ denote the pathwise derivative of $\psi^{\mathbf{d}}$ at $\mathbf{Q}_0$ in the score direction $h$. Under the fixed longitudinal reference law $P_{\mathbf{Q}_0,\bar{\mathbf{g}}_\infty}$, the pathwise derivative satisfies $\dot{\psi}^{\mathbf{d}}_{\mathbf{Q}_0}(h) = P_{\mathbf{Q}_0,\bar{\mathbf{g}}_\infty}\left(D^{\mathbf{d}}(P_{\mathbf{Q}_0,\bar{\mathbf{g}}_\infty})h\right)$, for any $h \in \mathcal{H}(\mathbf{Q}_0, \bar{\mathbf{g}}_\infty)$. Thus, $D^{\mathbf{d}}(P_{\mathbf{Q}_0,\bar{\mathbf{g}}_\infty})$ is the canonical gradient in the limiting reference experiment $P_{\mathbf{Q}_0,\bar{\mathbf{g}}_\infty}$. The corresponding semiparametric efficiency bound is therefore $P_{\mathbf{Q}_0,\bar{\mathbf{g}}_\infty}\left[\left(D^{\mathbf{d}}(P_{\mathbf{Q}_0,\bar{\mathbf{g}}_\infty})\right)^2\right]$.

When $\mathbf{Q}_\infty = \mathbf{Q}_0$, $\psi^{\mathbf{d}}(\mathbf{Q}_n^*) - \psi^{\mathbf{d}}(\mathbf{Q}_0)$ has $\frac{1}{n}\sum_{i=1}^{n}\left[D^{\mathbf{d}}(P_{\mathbf{Q}_0,\bar{\mathbf{g}}_\infty})(O_i) - P_{\mathbf{Q}_0,\mathbf{g}_i}D^{\mathbf{d}}(P_{\mathbf{Q}_0,\bar{\mathbf{g}}_\infty})\right]$ as the leading term. Since $P_{\mathbf{Q}_0,\mathbf{g}_i}D^{\mathbf{d}}(P_{\mathbf{Q}_0,\bar{\mathbf{g}}_\infty}) = 0$ and Assumption 3.4 hold, we have

$$\frac{1}{n}\sum_{i=1}^{n} Var_{P_{\mathbf{Q}_0,\mathbf{g}_i}} D^{\mathbf{d}}(P_{\mathbf{Q}_0,\bar{\mathbf{g}}_\infty}) = P_{\mathbf{Q}_0,\bar{\mathbf{g}}_n}\left[\left(D^{\mathbf{d}}(P_{\mathbf{Q}_0,\bar{\mathbf{g}}_\infty})\right)^2\right] \xrightarrow{p} P_{\mathbf{Q}_0,\bar{\mathbf{g}}_\infty}\left[\left(D^{\mathbf{d}}(P_{\mathbf{Q}_0,\bar{\mathbf{g}}_\infty})\right)^2\right].$$

Thus, the limiting variance of ADL-LTMLE equals the squared norm of the canonical gradient under the limiting reference law.

The regularity arguments developed for the ADL-TMLE in [92] directly apply to the present longitudinal experiment using the corresponding longitudinal stage-wise score. By the convolution theorem [82], the asymptotic distribution of any regular estimator is the convolution of a normal distribution $N\left(0, P_{\mathbf{Q}_0,\bar{\mathbf{g}}_\infty}\left[\left\{D^{\mathbf{d}}(P_{\mathbf{Q}_0,\bar{\mathbf{g}}_\infty})\right\}^2\right]\right)$ with an independent noise distribution. Since the ADL-LTMLE is regular and attains this Gaussian limit without additional noise, it is semiparametric efficient.

### 3.8.2 Data-Generating Mechanisms

**Two-stage setting ($K = 2$).** For the two-stage setting, the observed data are $O = (L_1, A_1, L_2, A_2, Y)$, where $L_1, L_2 \in \{0, 1\}$ and $Y$ is continuous. The baseline covariate follows $L_1 \sim$ Bernoulli(0.5). Conditional on $(L_1, A_1)$, the intermediate covariate $L_2$ satisfies $P(L_2 = 1 \mid L_1 = 0, A_1 = 0) = 0.2$, $P(L_2 = 1 \mid L_1 = 0, A_1 = 1) = 0.8$, $P(L_2 = 1 \mid L_1 = 1, A_1 = 0) = 0.7$, and $P(L_2 = 1 \mid L_1 = 1, A_1 = 1) = 0.3$.

The outcome is generated as $Y \mid (L_1, A_1, L_2, A_2) \sim \mathcal{N}(\mu(L_1, A_1, L_2, A_2),\ \sigma^2(L_1, A_1, A_2))$, with $\mu(L_1, A_1, L_2, A_2) = 8 + 2L_1 + 1.5L_2 + (4 + 2(1 - L_1))A_1 + (2 + 3L_1)A_2 - 1.5A_1A_2 + 1.5L_2A_2$. The standard deviation $\sigma(L_1, A_1, A_2)$ is set to 0.5 when $(L_1, A_1, A_2) = (0, 0, 0)$, to 4 when $(L_1, A_1, A_2) = (0, 0, 1)$ or $(1, 1, 1)$, to 1 when $(L_1, A_1, A_2) = (1, 0, 0)$, and to 1 otherwise.

**Three-stage setting ($K = 3$).** For the three-stage setting, the observed data are $O = (L_1, A_1, L_2, A_2, L_3, A_3, Y)$, where $L_1, L_2, L_3 \in \{0, 1\}$ and $Y$ is continuous. The baseline covariate follows $L_1 \sim$ Bernoulli(0.5). Conditional on $(L_1, A_1)$, $P(L_2 = 1 \mid$

$L_1, A_1) = \min\{0.9, \max(0.1,\ 0.3 + 0.1L_1 + 0.2A_1)\}$. Given $(L_1, A_1, L_2, A_2)$, $P(L_3 = 1 \mid L_1, A_1, L_2, A_2) = \min\{0.9, \max(0.1,\ 0.3 + 0.1L_2 + 0.1A_2 + 0.05A_1)\}$.

The outcome is generated as $Y \mid (L_1, A_1, L_2, A_2, L_3, A_3) \sim \mathcal{N}(\mu, \sigma^2)$, with $\mu = 10 + 2L_1 + 2L_2 + 2L_3 + 3A_1 + 2A_2 + A_3 - A_1A_2 - 0.5A_2A_3$, and the standard deviation $\sigma$ equal to 4 when $(A_1, A_2, A_3) = (0, 0, 1)$, equal to 2 when $(A_1, A_2, A_3) = (0, 1, 1)$, and equal to 1 otherwise.

### 3.8.3 Additional Simulation Results

We also implemented one-step estimators under the simulation scenarios in Section 3.6, as analogues of the ADL-LTMLE and AD-LTMLE estimators. Specifically, they are defined by: $\hat{\psi}^{\tau}_{n,\text{one-step}} = \psi^{\tau}_{n,\text{init}}(\mathbf{Q}_n) + \frac{1}{n}\sum_{i=1}^{n} D^{\tau}(P_{\mathbf{Q}_n,\bar{\mathbf{g}}_n})(O_i)$ and $\hat{\psi}^{\tau,\dagger}_{n,\text{one-step}} = \psi^{\tau}_{n,\text{init}}(\mathbf{Q}_n) + \frac{1}{n}\sum_{i=1}^{n} D^{\tau}(P_{\mathbf{Q}_n,\mathbf{g}_i})(O_i)$, respectively, where $\mathbf{Q}_n$ denotes the initial estimate of $\mathbf{Q}$. The two estimators coincide under a fixed design. The results are reported in Tables 3.3 and Tables 3.4 for $K = 2$ and $K = 3$, respectively, which show performance similar to that of the corresponding LTMLE-based estimators.

Table 3.3: Simulation results for $K = 2$ using one-step estimators. Performance of one-step estimators under different longitudinal designs. We report the empirical variance of the estimates of $\psi^1$ ($\text{Var}_1$) and $\psi^2$ ($\text{Var}_2$), the weighted sum of variances using normalized user-specified weight $w$ (TarVar), the empirical 95% coverage probabilities ($\text{Cov}_1$, $\text{Cov}_2$), and the corresponding oracle coverage probabilities ($\text{OCov}_1$, $\text{OCov}_2$) across 500 Monte Carlo runs.

| $w$ | Design | Estimator | TarVar | $\text{Var}_1$ | $\text{Var}_2$ | $\text{Cov}_1$(%) | $\text{Cov}_2$(%) | $\text{OCov}_1$(%) | $\text{OCov}_2$(%) |
|---|---|---|---|---|---|---|---|---|---|
| (1/2, 1/2) | Non-adaptive | $\hat{\psi}^{\tau}_{n,\text{one-step}}$ | 0.029 | 0.040 | 0.018 | 95.0 | 97.0 | 95.4 | 95.6 |
| | Adaptive | $\hat{\psi}^{\tau,\dagger}_{n,\text{one-step}}$ | 0.022 | 0.030 | 0.013 | 92.8 | 94.4 | 94.8 | 94.4 |
| | Adaptive | $\hat{\psi}^{\tau}_{n,\text{one-step}}$ | 0.019 | 0.026 | 0.012 | 93.0 | 94.6 | 94.2 | 95.4 |
| | Oracle | $\hat{\psi}^{\tau}_{n,\text{one-step}}$ | 0.017 | 0.024 | 0.011 | 94.2 | 94.6 | 95.0 | 95.0 |
| (1/3, 2/3) | Non-adaptive | $\hat{\psi}^{\tau}_{n,\text{one-step}}$ | 0.026 | 0.040 | 0.018 | 95.0 | 97.0 | 95.4 | 95.6 |
| | Adaptive | $\hat{\psi}^{\tau,\dagger}_{n,\text{one-step}}$ | 0.019 | 0.032 | 0.013 | 91.6 | 95.8 | 94.4 | 96.6 |
| | Adaptive | $\hat{\psi}^{\tau}_{n,\text{one-step}}$ | 0.017 | 0.028 | 0.012 | 92.6 | 95.6 | 95.0 | 96.0 |
| | Oracle | $\hat{\psi}^{\tau}_{n,\text{one-step}}$ | 0.015 | 0.024 | 0.010 | 94.8 | 95.6 | 95.4 | 95.6 |
| (2/3, 1/3) | Non-adaptive | $\hat{\psi}^{\tau}_{n,\text{one-step}}$ | 0.033 | 0.040 | 0.018 | 95.0 | 97.0 | 95.4 | 95.6 |
| | Adaptive | $\hat{\psi}^{\tau,\dagger}_{n,\text{one-step}}$ | 0.025 | 0.030 | 0.014 | 92.2 | 94.4 | 95.4 | 95.2 |
| | Adaptive | $\hat{\psi}^{\tau}_{n,\text{one-step}}$ | 0.021 | 0.026 | 0.012 | 93.2 | 94.8 | 95.4 | 95.6 |
| | Oracle | $\hat{\psi}^{\tau}_{n,\text{one-step}}$ | 0.019 | 0.023 | 0.012 | 94.6 | 94.6 | 95.2 | 94.8 |

Next, we conduct sensitivity analyses for the ADL-LTMLE approach. We vary the total and cohort-wise sample size, and report results using both the correctly specified strata-specific mean estimator and misspecified working outcome-regression models. Results across Tables 3.5-3.8 show that TarVar decreases substantially as the total sample size increases for both $K = 2$ and $K = 3$. While misspecified outcome regressions increase variance,

Table 3.4: Simulation results for $K = 3$ (one-step estimator). Performance of one-step estimators under different longitudinal designs. We report the weighted sum of variances using normalized user-specified weight $w$ (TarVar), the empirical variance of the estimates of $\psi^1$ ($\text{Var}_1$), $\psi^2$ ($\text{Var}_2$), and $\psi^3$ ($\text{Var}_3$), the empirical 95% coverage probabilities ($\text{Cov}_1$, $\text{Cov}_2$, $\text{Cov}_3$), and the corresponding oracle coverage probabilities ($\text{OCov}_1$, $\text{OCov}_2$, $\text{OCov}_3$) across 500 Monte Carlo runs.

| $w$ | Design | Estimator | TarVar | $\text{Var}_1$ | $\text{Var}_2$ | $\text{Var}_3$ | $\text{Cov}_1$(%) | $\text{Cov}_2$(%) | $\text{Cov}_3$(%) | $\text{OCov}_1$(%) | $\text{OCov}_2$(%) | $\text{OCov}_3$(%) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| (1/3, 1/3, 1/3) | Non-adaptive | $\hat{\psi}^{\tau}_{n,\text{one-step}}$ | 0.056 | 0.026 | 0.079 | 0.063 | 94.2 | 94.6 | 96.2 | 94.0 | 94.2 | 95.4 |
| | Adaptive | $\hat{\psi}^{\tau,\dagger}_{n,\text{one-step}}$ | 0.031 | 0.022 | 0.039 | 0.033 | 95.2 | 94.6 | 94.6 | 95.4 | 94.6 | 93.6 |
| | Adaptive | $\hat{\psi}^{\tau}_{n,\text{one-step}}$ | 0.026 | 0.020 | 0.031 | 0.026 | 95.2 | 94.4 | 95.2 | 95.2 | 95.0 | 95.4 |
| | Oracle | $\hat{\psi}^{\tau}_{n,\text{one-step}}$ | 0.022 | 0.020 | 0.025 | 0.020 | 95.0 | 94.8 | 96.4 | 94.8 | 94.0 | 94.6 |
| (1/6, 1/3, 1/2) | Non-adaptive | $\hat{\psi}^{\tau}_{n,\text{one-step}}$ | 0.062 | 0.026 | 0.079 | 0.063 | 94.2 | 94.6 | 96.2 | 94.0 | 94.2 | 95.4 |
| | Adaptive | $\hat{\psi}^{\tau,\dagger}_{n,\text{one-step}}$ | 0.032 | 0.027 | 0.038 | 0.029 | 94.6 | 95.2 | 94.4 | 94.6 | 95.0 | 94.6 |
| | Adaptive | $\hat{\psi}^{\tau}_{n,\text{one-step}}$ | 0.024 | 0.024 | 0.028 | 0.022 | 94.4 | 95.6 | 95.2 | 94.8 | 95.0 | 94.8 |
| | Oracle | $\hat{\psi}^{\tau}_{n,\text{one-step}}$ | 0.021 | 0.026 | 0.024 | 0.017 | 95.8 | 95.8 | 97.2 | 95.4 | 94.4 | 96.2 |
| (1/2, 1/3, 1/6) | Non-adaptive | $\hat{\psi}^{\tau}_{n,\text{one-step}}$ | 0.050 | 0.026 | 0.079 | 0.063 | 94.2 | 94.6 | 96.2 | 94.0 | 94.2 | 95.4 |
| | Adaptive | $\hat{\psi}^{\tau,\dagger}_{n,\text{one-step}}$ | 0.029 | 0.019 | 0.039 | 0.038 | 95.0 | 93.4 | 93.2 | 95.0 | 93.2 | 93.8 |
| | Adaptive | $\hat{\psi}^{\tau}_{n,\text{one-step}}$ | 0.024 | 0.017 | 0.031 | 0.029 | 94.8 | 94.2 | 94.6 | 95.4 | 94.4 | 95.4 |
| | Oracle | $\hat{\psi}^{\tau}_{n,\text{one-step}}$ | 0.021 | 0.016 | 0.026 | 0.025 | 96.2 | 95.6 | 95.2 | 95.0 | 95.0 | 94.4 |

coverage remains near nominal level across different sample sizes, supporting the robustness of ADL-LTMLE to model misspecification.

Table 3.5: Complete results of ADL-LTMLE using the correct strata-specific mean estimates ($K = 2$). The settings vary total sample size $n$ and cohort size $b$.

| $w$ | Setting | TarVar | $\text{Var}_1$ | $\text{Var}_2$ | $\text{Cov}_1$(%) | $\text{Cov}_2$(%) | $\text{OCov}_1$(%) | $\text{OCov}_2$(%) |
|---|---|---|---|---|---|---|---|---|
| (1/2, 1/2) | $n = 1000,\ b = 200$ | 0.036 | 0.049 | 0.024 | 94.6 | 95.4 | 95.4 | 96.0 |
| (1/2, 1/2) | $n = 2000,\ b = 200$ | 0.017 | 0.022 | 0.011 | 94.8 | 95.8 | 94.4 | 95.8 |
| (1/2, 1/2) | $n = 2000,\ b = 400$ | 0.019 | 0.026 | 0.012 | 93.0 | 94.6 | 94.2 | 95.4 |
| (1/2, 1/2) | $n = 4000,\ b = 800$ | 0.008 | 0.011 | 0.006 | 95.8 | 96.4 | 95.0 | 96.0 |
| (1/3, 2/3) | $n = 1000,\ b = 200$ | 0.033 | 0.051 | 0.023 | 93.6 | 94.8 | 94.2 | 95.2 |
| (1/3, 2/3) | $n = 2000,\ b = 200$ | 0.015 | 0.023 | 0.011 | 95.2 | 95.0 | 94.2 | 95.0 |
| (1/3, 2/3) | $n = 2000,\ b = 400$ | 0.017 | 0.028 | 0.012 | 92.6 | 95.6 | 95.0 | 96.0 |
| (1/3, 2/3) | $n = 4000,\ b = 800$ | 0.007 | 0.012 | 0.005 | 96.6 | 94.2 | 96.0 | 94.0 |
| (2/3, 1/3) | $n = 1000,\ b = 200$ | 0.039 | 0.046 | 0.025 | 94.6 | 94.8 | 95.4 | 95.2 |
| (2/3, 1/3) | $n = 2000,\ b = 200$ | 0.018 | 0.021 | 0.012 | 95.0 | 95.6 | 94.6 | 95.8 |
| (2/3, 1/3) | $n = 2000,\ b = 400$ | 0.021 | 0.026 | 0.012 | 93.2 | 94.8 | 95.4 | 95.6 |
| (2/3, 1/3) | $n = 4000,\ b = 800$ | 0.009 | 0.011 | 0.006 | 96.6 | 97.2 | 95.6 | 97.4 |

Table 3.6: Complete results of ADL-LTMLE using the misspecified working outcome regression $Y \sim L_1 + A_1 + L_2 + A_2$, instead of the correct outcome model ($K = 2$). The settings vary total sample size $n$ and cohort size $b$.

| $w$ | Setting | TarVar | $\text{Var}_1$ | $\text{Var}_2$ | $\text{Cov}_1(\%)$ | $\text{Cov}_2(\%)$ | $\text{OCov}_1(\%)$ | $\text{OCov}_2(\%)$ |
|---|---|---|---|---|---|---|---|---|
| (1/2, 1/2) | $n = 1000,\ b = 200$ | 0.041 | 0.049 | 0.033 | 94.4 | 93.4 | 95.0 | 94.2 |
| (1/2, 1/2) | $n = 2000,\ b = 200$ | 0.019 | 0.022 | 0.017 | 94.8 | 95.6 | 94.6 | 95.4 |
| (1/2, 1/2) | $n = 2000,\ b = 400$ | 0.021 | 0.026 | 0.017 | 92.8 | 95.0 | 94.6 | 95.6 |
| (1/2, 1/2) | $n = 4000,\ b = 800$ | 0.010 | 0.011 | 0.008 | 95.2 | 95.0 | 95.0 | 96.0 |
| (1/3, 2/3) | $n = 1000,\ b = 200$ | 0.038 | 0.052 | 0.031 | 93.8 | 93.6 | 95.0 | 94.2 |
| (1/3, 2/3) | $n = 2000,\ b = 200$ | 0.018 | 0.023 | 0.015 | 94.6 | 94.6 | 94.4 | 94.8 |
| (1/3, 2/3) | $n = 2000,\ b = 400$ | 0.019 | 0.028 | 0.014 | 92.8 | 95.0 | 95.4 | 95.0 |
| (1/3, 2/3) | $n = 4000,\ b = 800$ | 0.009 | 0.012 | 0.007 | 96.0 | 94.8 | 95.6 | 95.6 |
| (2/3, 1/3) | $n = 1000,\ b = 200$ | 0.043 | 0.047 | 0.037 | 94.6 | 94.0 | 94.8 | 94.6 |
| (2/3, 1/3) | $n = 2000,\ b = 200$ | 0.020 | 0.021 | 0.019 | 94.4 | 94.8 | 94.2 | 95.8 |
| (2/3, 1/3) | $n = 2000,\ b = 400$ | 0.024 | 0.026 | 0.018 | 94.0 | 94.2 | 95.4 | 95.2 |
| (2/3, 1/3) | $n = 4000,\ b = 800$ | 0.010 | 0.011 | 0.009 | 96.4 | 95.4 | 95.2 | 96.2 |

Table 3.7: Complete results of ADL-LTMLE using the correct strata-specific mean estimates ($K = 3$). The settings vary total sample size $n$ and cohort size $b$.

| $w$ | Setting | TarVar | $\text{Var}_1$ | $\text{Var}_2$ | $\text{Var}_3$ | $\text{Cov}_1(\%)$ | $\text{Cov}_2(\%)$ | $\text{Cov}_3(\%)$ | $\text{OCov}_1(\%)$ | $\text{OCov}_2(\%)$ | $\text{OCov}_3(\%)$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| (1/3, 1/3, 1/3) | $n = 1000,\ b = 200$ | 0.052 | 0.042 | 0.061 | 0.055 | 93.4 | 94.8 | 92.4 | 95.0 | 95.8 | 95.2 |
| (1/3, 1/3, 1/3) | $n = 2000,\ b = 200$ | 0.025 | 0.020 | 0.029 | 0.025 | 94.2 | 95.0 | 93.6 | 95.0 | 95.4 | 95.2 |
| (1/3, 1/3, 1/3) | $n = 2000,\ b = 400$ | 0.026 | 0.020 | 0.031 | 0.026 | 95.2 | 94.4 | 95.2 | 95.2 | 95.0 | 95.4 |
| (1/3, 1/3, 1/3) | $n = 4000,\ b = 800$ | 0.013 | 0.010 | 0.015 | 0.012 | 95.0 | 95.2 | 94.6 | 95.4 | 95.0 | 94.6 |
| (1/6, 1/3, 1/2) | $n = 1000,\ b = 200$ | 0.052 | 0.049 | 0.060 | 0.048 | 93.4 | 92.8 | 92.2 | 94.6 | 93.2 | 93.8 |
| (1/6, 1/3, 1/2) | $n = 2000,\ b = 200$ | 0.024 | 0.024 | 0.027 | 0.021 | 93.4 | 94.8 | 94.2 | 93.4 | 94.8 | 94.6 |
| (1/6, 1/3, 1/2) | $n = 2000,\ b = 400$ | 0.024 | 0.024 | 0.028 | 0.022 | 94.4 | 95.6 | 95.2 | 94.8 | 95.0 | 94.8 |
| (1/6, 1/3, 1/2) | $n = 4000,\ b = 800$ | 0.012 | 0.012 | 0.014 | 0.011 | 94.6 | 97.0 | 95.8 | 95.0 | 96.8 | 96.2 |
| (1/2, 1/3, 1/6) | $n = 1000,\ b = 200$ | 0.048 | 0.035 | 0.062 | 0.059 | 95.8 | 96.4 | 93.4 | 95.8 | 96.4 | 94.8 |
| (1/2, 1/3, 1/6) | $n = 2000,\ b = 200$ | 0.023 | 0.017 | 0.029 | 0.029 | 94.8 | 94.8 | 94.6 | 95.2 | 94.2 | 95.8 |
| (1/2, 1/3, 1/6) | $n = 2000,\ b = 400$ | 0.024 | 0.017 | 0.031 | 0.029 | 94.8 | 94.2 | 94.6 | 95.4 | 94.4 | 95.4 |
| (1/2, 1/3, 1/6) | $n = 4000,\ b = 800$ | 0.012 | 0.009 | 0.015 | 0.014 | 94.8 | 96.4 | 94.8 | 95.4 | 95.8 | 95.0 |

Table 3.8: Complete results of ADL-LTMLE using the misspecified working outcome regression $Y \sim L_1 + A_1 + A_2 + A_3$, instead of the correct outcome model ($K = 3$). The settings vary total sample size $n$ and cohort size $b$.

| $w$ | Setting | TarVar | $\text{Var}_1$ | $\text{Var}_2$ | $\text{Var}_3$ | $\text{Cov}_1(\%)$ | $\text{Cov}_2(\%)$ | $\text{Cov}_3(\%)$ | $\text{OCov}_1(\%)$ | $\text{OCov}_2(\%)$ | $\text{OCov}_3(\%)$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| (1/3, 1/3, 1/3) | $n = 1000,\ b = 200$ | 0.065 | 0.048 | 0.069 | 0.078 | 94.0 | 95.6 | 94.6 | 94.8 | 96.0 | 95.8 |
| (1/3, 1/3, 1/3) | $n = 2000,\ b = 200$ | 0.032 | 0.024 | 0.032 | 0.040 | 94.6 | 94.8 | 94.2 | 95.4 | 95.4 | 94.6 |
| (1/3, 1/3, 1/3) | $n = 2000,\ b = 400$ | 0.034 | 0.025 | 0.036 | 0.041 | 94.2 | 94.8 | 93.0 | 94.4 | 95.4 | 94.0 |
| (1/3, 1/3, 1/3) | $n = 4000,\ b = 800$ | 0.017 | 0.013 | 0.018 | 0.019 | 93.0 | 94.2 | 95.8 | 95.6 | 95.4 | 96.2 |
| (1/6, 1/3, 1/2) | $n = 1000,\ b = 200$ | 0.066 | 0.057 | 0.068 | 0.067 | 95.6 | 95.2 | 95.0 | 95.6 | 95.6 | 94.6 |
| (1/6, 1/3, 1/2) | $n = 2000,\ b = 200$ | 0.032 | 0.029 | 0.032 | 0.033 | 93.8 | 94.6 | 94.4 | 94.6 | 95.0 | 94.8 |
| (1/6, 1/3, 1/2) | $n = 2000,\ b = 400$ | 0.034 | 0.031 | 0.035 | 0.034 | 93.0 | 94.8 | 93.6 | 94.2 | 95.4 | 94.0 |
| (1/6, 1/3, 1/2) | $n = 4000,\ b = 800$ | 0.017 | 0.015 | 0.017 | 0.017 | 93.0 | 95.4 | 96.2 | 94.0 | 96.2 | 96.2 |
| (1/2, 1/3, 1/6) | $n = 1000,\ b = 200$ | 0.056 | 0.039 | 0.068 | 0.084 | 96.0 | 96.8 | 94.6 | 96.0 | 96.2 | 95.0 |
| (1/2, 1/3, 1/6) | $n = 2000,\ b = 200$ | 0.028 | 0.019 | 0.031 | 0.046 | 95.0 | 95.6 | 94.4 | 94.8 | 95.0 | 94.4 |
| (1/2, 1/3, 1/6) | $n = 2000,\ b = 400$ | 0.031 | 0.021 | 0.037 | 0.047 | 94.4 | 93.8 | 93.6 | 95.2 | 94.2 | 93.8 |
| (1/2, 1/3, 1/6) | $n = 4000,\ b = 800$ | 0.015 | 0.011 | 0.018 | 0.022 | 93.6 | 95.6 | 94.6 | 93.8 | 95.4 | 95.0 |

# Bibliography


[1] A. C. Atkinson, A. Biswas, and L. Pronzato. *Covariate-balanced Response-adaptive Designs for Clinical Trials with Continuous Responses that Target Allocation Probabilities*. Technical Report NI11042-DAE. Cambridge, UK: Isaac Newton Institute for Mathematical Sciences, 2011.

[2] Heejung Bang and James M Robins. "Doubly robust estimation in missing data and causal inference models". In: *Biometrics* 61.4 (2005), pp. 962–973.

[3] David Benkeser and Mark van der Laan. "The Highly Adaptive Lasso Estimator". In: *2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*. 2016, pp. 689–696. DOI: 10.1109/DSAA.2016.93.

[4] Aurélien Bibaut et al. "Post-contextual-bandit inference". In: *Advances in Neural Information Processing Systems* 34 (2021), pp. 28548–28559.

[5] Aurélien F Bibaut and Mark J van der Laan. "Fast rates for empirical risk minimization over càdlàg functions with bounded sectional variation norm". In: *arXiv preprint arXiv:1907.09244* (2019).

[6] Atanu Biswas and Anup Dewanji. "A randomized longitudinal play-the-winner design for repeated binary data". In: *Australian & New Zealand Journal of Statistics* 46.4 (2004), pp. 675–684.

[7] Djallel Bouneffouf, Irina Rish, and Charu Aggarwal. "Survey on applications of multi-armed and contextual bandits". In: *2020 IEEE congress on evolutionary computation (CEC)*. IEEE. 2020, pp. 1–8.

[8] Leo Breiman. "Random forests". In: *Machine Learning* 45 (2001), pp. 5–32.

[9] Bruce M Brown. "Martingale central limit theorems". In: *The Annals of Mathematical Statistics* 42 (1971), pp. 59–66.

[10] Sébastien Bubeck, Nicolo Cesa-Bianchi, et al. "Regret analysis of stochastic and nonstochastic multi-armed bandit problems". In: *Foundations and Trends in Machine Learning* 5.1 (2012), pp. 1–122.

[11] Marc Buyse et al. "The validation of surrogate endpoints in meta-analyses of randomized experiments". In: *Biostatistics* 1.1 (2000), pp. 49–67.

[12] Antoine Chambaz and Mark J van der Laan. "Inference in targeted group-sequential covariate-adjusted randomized clinical trials". In: *Scandinavian Journal of Statistics* 41.1 (2014), pp. 104–140.

[13] Antoine Chambaz and Mark J van der Laan. "Targeting the optimal design in randomized clinical trials with binary outcomes and no covariate: simulation study". In: *The International Journal of Biostatistics* 7.1 (2011).

[14] Antoine Chambaz, Wenjing Zheng, and Mark J van der Laan. "Targeted sequential design for targeted learning inference of the optimal treatment rule and its mean reward". In: *The Annals of Statistics* 45.6 (2017), pp. 2537–2564.

[15] Siu Hung Cheung et al. "Covariate-Adjusted Response-Adaptive Designs for Generalized Linear Models". In: *Journal of Statistical Planning and Inference* 149 (2014), pp. 152–161.

[16] Shein-Chung Chow and Mark Chang. "Adaptive Design Methods in Clinical Trials–A Review". In: *Orphanet Journal of Rare Diseases* 3.1 (2008), p. 11. DOI: 10.1186/1750-1172-3-11.

[17] Eugene B Cone et al. "Assessment of time-to-treatment initiation and survival in a cohort of patients with common cancers". In: *JAMA network open* 3.12 (2020), e2030072–e2030072.

[18] Michael J Daniels and Michael D Hughes. "Meta-analysis for the evaluation of potential surrogate markers". In: *Statistics in Medicine* 16.17 (1997), pp. 1965–1982.

[19] Weitao Duan, Shan Ba, and Chunzhe Zhang. "Online Experimentation with Surrogate Metrics: Guidelines and a Case Study". In: *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*. 2021, pp. 193–201.

[20] Michael R Elliott. "Surrogate Endpoints in Clinical Trials". In: *Annual Review of Statistics and its Application* 10 (2023), pp. 75–96.

[21] Laurence S Freedman, Barry I Graubard, and Arthur Schatzkin. "Statistical validation of intermediate endpoints for chronic diseases". In: *Statistics in Medicine* 11.2 (1992), pp. 167–178.

[22] Jingya Gao, Feifang Hu, and Wei Ma. "Response-Adaptive Randomization Procedure in Clinical Trials with Surrogate Endpoints". In: *Statistics in Medicine* 43.30 (2024), pp. 5911–5921.

[23] Sara A van de Geer. *Empirical Processes in M-Estimation*. Cambridge University Press, 2000.

[24] Elvin H. Geng et al. "Adaptive strategies for retention in care among persons living with HIV". In: *NEJM Evidence* 2.4 (2023), EVIDoa2200076.

[25] Rik Ghosh et al. "Optimal adaptive SMART designs with binary outcomes". In: *Biometrics* 80.4 (2024), ujae140.

[26] Peter B Gilbert and Michael G Hudgens. "Evaluating candidate principal surrogate endpoints". In: *Biometrics* 64.4 (2008), pp. 1146–1154.

[27] Richard D. Gill, Mark J. van der Laan, and Jon A. Wellner. "Inefficient estimators of the bivariate survival function for three models". en. In: *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques* 31.3 (1995), pp. 545–597.

[28] Vitor Hadad et al. "Confidence intervals for policy evaluation in adaptive experiments". In: *Proceedings of the National Academy of Sciences* 118.15 (2021), e2014602118.

[29] Peter Hall and Christopher C Heyde. *Martingale limit theory and its application*. Academic press, 2014.

[30] Ramon van Handel. "On the minimal penalty for Markov order estimation". In: *Probability Theory and Related Fields* 150 (2011), pp. 709–738.

[31] Jesse Y Hsu et al. "Surrogate markers for time-varying treatments and outcomes". In: *Clinical Trials* 12.4 (2015), pp. 309–316.

[32] Feifang Hu and William F Rosenberger. *The Theory of Response-Adaptive Randomization in Clinical Trials*. John Wiley & Sons, 2006.

[33] Feifang Hu and Li-Xin Zhang. "Asymptotic properties of doubly adaptive biased coin designs for multitreatment clinical trials". In: *The Annals of Statistics* 32.1 (2004), pp. 268–301.

[34] Jianhua Hu, Hongjian Zhu, and Feifang Hu. "A unified family of covariate-adjusted response-adaptive designs based on efficiency and ethics". In: *Journal of the American Statistical Association* 110.509 (2015), pp. 357–367.

[35] Yichun Hu and Nathan Kallus. "Dtr bandit: Learning to make response-adaptive decisions with low regret". In: *arXiv preprint arXiv:2005.02791* (2020).

[36] Tao Huang, Zhongqiang Liu, and Feifang Hu. "Longitudinal covariate-adjusted response-adaptive randomized designs". In: *Journal of Statistical Planning and Inference* 143.10 (2013), pp. 1816–1827.

[37] Tao Huang and Hongjian Zhu. *Longitudinal Covariate-Adjusted Response–Adaptive Randomization: Impact of Missing Data*. 2016.

[38] Xuelin Huang et al. "Using short-term response information to facilitate adaptive randomization for survival clinical trials". In: *Statistics in Medicine* 28.12 (2009), pp. 1680–1689.

[39] Edward H Kennedy. "Towards optimal doubly robust estimation of heterogeneous causal effects". In: *Electronic Journal of Statistics* 17.2 (2023), pp. 3008–3049.

[40] Michael R Kosorok and Eric B Laber. "Precision medicine". In: *Annual review of statistics and its application* 6.1 (2019), pp. 263–286.

[41] Michael R Kosorok et al. *Introduction to the theory and methods special issue on precision medicine and individualized policy discovery*. 2021.

[42] Mark J van der Laan. *A generally efficient targeted minimum loss based estimator. U.C. Berkeley Division of Biostatistics.* Working Paper 343. https://biostats.bepress.com/ucbbiostat/paper343. 2015.

[43] Mark J van der Laan. "Statistical inference for variable importance". In: *The International Journal of Biostatistics* 2.1 (2006).

[44] Mark J van der Laan and Susan Gruber. "Targeted minimum loss based estimation of causal effects of multiple time point interventions". In: *The international journal of biostatistics* 8.1 (2012).

[45] Mark J van der Laan and Alexander R Luedtke. "Targeted learning of the mean outcome under an optimal dynamic treatment rule". In: *Journal of Causal Inference* 3.1 (2015), pp. 61–95.

[46] Mark J van der Laan and Ivana Malenica. "Robust estimation of data-dependent causal effects based on observing a single time-series". In: *arXiv preprint arXiv:1809.00734* (2018).

[47] Mark J van der Laan, Eric C Polley, and Alan E Hubbard. "Super learner". In: *Statistical Applications in Genetics and Molecular Biology* 6.1 (2007).

[48] Mark J van der Laan and James M Robins. *Unified methods for censored longitudinal data and causality*. Springer, 2003.

[49] Mark J van der Laan and Sherri Rose. *Targeted Learning in Data Science*. Springer, 2018.

[50] Mark J van der Laan and Sherri Rose. *Targeted Learning: Causal Inference for Observational and Experimental Data*. Springer, 2011.

[51] Mark J van der Laan and Daniel Rubin. "Targeted maximum likelihood learning". In: *The International Journal of Biostatistics* 2.1 (2006).

[52] Mark J. van der Laan. "Higher Order Spline Highly Adaptive Lasso Estimators of Functional Parameters: Pointwise Asymptotic Normality and Uniform Convergence Rates". In: *arXiv preprint arXiv:2301.13354* (2023).

[53] Mark J. van der Laan. *The Construction and Analysis of Adaptive Group Sequential Designs. U.C. Berkeley Division of Biostatistics.* Working Paper 232. https://biostats.bepress.com/ucbbiostat/paper232. 2008.

[54] Nicholas Larsen et al. "Statistical challenges in online controlled experiments: A review of a/b testing methodology". In: *The American Statistician* 78.2 (2024), pp. 135–149.

[55] Tor Lattimore and Csaba Szepesvári. *Bandit Algorithms*. Cambridge University Press, 2020.

[56] Philip W Lavori and Ree Dawson. "A design for testing clinical strategies: biased adaptive within-subject randomization". In: *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 163.1 (2000), pp. 29–38.

[57] Huitan Lei et al. "A "SMART" design for building individualized treatment sequences". In: *Annual review of clinical psychology* 8.2012 (2012), pp. 21–48.

[58] DY Lin, TR Fleming, and V De Gruttola. "Estimating the proportion of treatment effect explained by a surrogate marker". In: *Statistics in Medicine* 16.13 (1997), pp. 1515–1527.

[59] Alexander R Luedtke and Mark J van der Laan. "Statistical inference for the mean outcome under a possibly non-unique optimal treatment strategy". In: *The Annals of Statistics* 44.2 (2016), pp. 713–742. DOI: 10.1214/15-AOS1384.

[60] Alexander R Luedtke and Mark J van der Laan. "Super-learning of an optimal dynamic treatment rule". In: *The International Journal of Biostatistics* 12.1 (2016), pp. 305–332.

[61] Ivana Malenica, Aurelien Bibaut, and Mark J van der Laan. "Adaptive Sequential Design for a Single Time-Series". In: *arXiv preprint arXiv:2102.00102* (2021).

[62] Ivana Malenica et al. "Adaptive sequential surveillance with network and temporal dependence". In: *Biometrics* 80.1 (Jan. 2024), ujad007.

[63] Thomas M McDonald et al. "Impatient bandits: Optimizing recommendations for the long-term without delay". In: *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 2023, pp. 1687–1697.

[64] Lina M. Montoya et al. "The optimal dynamic treatment rule superlearner: considerations, performance, and application to criminal justice interventions". In: *The International Journal of Biostatistics* 19.1 (2021), pp. 217–238.

[65] Susan A Murphy. "An experimental design for the development of adaptive treatment strategies". In: *Statistics in medicine* 24.10 (2005), pp. 1455–1481.

[66] Susan A Murphy. "Optimal dynamic treatment regimes". In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 65.2 (2003), pp. 331–355.

[67] Amy S Nowacki, Wenle Zhao, and Yuko Y Palesch. "A surrogate-primary replacement algorithm for response-adaptive randomization in stroke clinical trials". In: *Statistical Methods in Medical Research* 26.3 (2017), pp. 1078–1092.

[68] Judea Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, 2000.

[69] Ross L Prentice. "Surrogate endpoints in clinical trials: definition and operational criteria". In: *Statistics in Medicine* 8.4 (1989), pp. 431–440.

[70] Lee Richardson et al. "Pareto optimal proxy metrics". In: *arXiv preprint arXiv:2307.01000* (2023).

[71] Herbert Robbins. "Some aspects of the sequential design of experiments". In: *Bulletin of the American Mathematical Society* 58.5 (1952), pp. 527–535.

[72] David S Robertson et al. "Response-Adaptive Randomization in Clinical Trials: From Myths to Practical Considerations". In: *Statistical Science* 38.2 (2023), pp. 185–208.

[73] James Robins. "A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect". In: *Mathematical Modelling* 7.9-12 (1986), pp. 1393–1512.

[74] James M Robins. "Optimal Structural Nested Models for Optimal Sequential Decisions". In: *Proceedings of the Second Seattle Symposium in Biostatistics*. Springer. 2004, pp. 189–326.

[75] James M Robins and Sander Greenland. "Identifiability and exchangeability for direct and indirect effects". In: *Epidemiology* 3.2 (1992), pp. 143–155.

[76] William F Rosenberger and Oleksandr Sverdlov. "Handling Covariates in the Design of Clinical Trials". In: *Statistical Science* 23.3 (2008), pp. 404–419.

[77] William F Rosenberger, AN Vidyashankar, and Deepak K Agarwal. "Covariate-adjusted response-adaptive designs for binary response". In: *Journal of Biopharmaceutical Statistics* 11.4 (2001), pp. 227–236.

[78] Dan Rubin and Mark J van der Laan. "A general imputation methodology for nonparametric regression with censored data". In: (2005).

[79] David Simchi-Levi and Chonghuan Wang. "Multi-armed bandit experimental design: Online decision-making and adaptive inference". In: *International Conference on Artificial Intelligence and Statistics*. PMLR. 2023, pp. 3086–3097.

[80] Roy N. Tamura et al. "A case study of an adaptive clinical trial in the treatment of outpatients with depressive disorder". In: *Journal of the American Statistical Association* 89.427 (1994), pp. 768–776. DOI: 10.1080/01621459.1994.10476810.

[81] William R Thompson. "On the likelihood that one unknown probability exceeds another in view of the evidence of two samples". In: *Biometrika* 25.3-4 (1933), pp. 285–294.

[82] Aad W van der Vaart et al. *Weak convergence and empirical processes: With applications to statistics*. Springer, 1996.

[83] Tyler J VanderWeele. "Surrogate measures and consistent surrogates". In: *Biometrics* 69.3 (2013), pp. 561–565.

[84] Junyao Wang, Liwen Wu, and Abdus S Wahed. "Adaptive randomization in a two-stage sequential multiple assignment randomized trial". In: *Biostatistics* 23.4 (2022), pp. 1182–1199.

[85] Zeyi Wang et al. "Targeted maximum likelihood based estimation for longitudinal mediation analysis". In: *Journal of Causal Inference* 13.1 (2025), p. 20230013.

[86] LJ Wei and S Durham. "The randomized play-the-winner rule in medical trials". In: *Journal of the American Statistical Association* 73.364 (1978), pp. 840–843.

[87] Christopher J Weir and Rod S Taylor. "Informed decision-making: Statistical methodology for surrogacy evaluation and its role in licensing and reimbursement assessments". In: *Pharmaceutical Statistics* 21.4 (2022), pp. 740–756.

[88] Jeremy Yang et al. "Targeting for long-term outcomes". In: *Management Science* 70.6 (2024), pp. 3841–3855.

[89] Ruohan Zhan et al. "Off-policy evaluation via adaptive weighting with data from contextual bandits". In: *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*. 2021, pp. 2125–2135.

[90] Jianrong Zhang et al. "Optimal timing of cancer treatments: a call for emerging evidence from clinical trials and real-world studies: Clinical Studies". In: *British Journal of Cancer* 132.12 (2025), pp. 1085–1090.

[91] Kelly Zhang, Lucas Janson, and Susan Murphy. "Statistical inference with m-estimators on adaptively collected data". In: *Advances in Neural Information Processing Systems* 34 (2021), pp. 7460–7471.

[92] Wenxin Zhang and Mark van der Laan. "Efficient Statistical Estimation for Sequential Adaptive Experiments with Implications for Adaptive Designs". In: *arXiv preprint arXiv:2508.09135* (2025).

[93] Wenxin Zhang et al. "Evaluating and Utilizing Surrogate Outcomes in Covariate-Adjusted Response-Adaptive Designs". In: *arXiv preprint arXiv:2408.02667* (2024).

[94] Li-Xin Zhang et al. "Asymptotic properties of covariate-adjusted response-adaptive designs". In: *The Annals of Statistics* 35.3 (2007), pp. 1166–1182.

[95] Bin Zhao et al. "Antiretroviral therapy initiation within 7 and 8–30 days post-HIV diagnosis demonstrates similar benefits in resource-limited settings". In: *Aids* 36.12 (2022), pp. 1741–1743.

[96] Wanying Zhao et al. "Incorporating covariates information in adaptive clinical trials for precision medicine". In: *Pharmaceutical Statistics* 21.1 (2022), pp. 176–195.

[97] Wenjing Zheng and Mark J. van der Laan. "Cross-Validated Targeted Minimum-Loss-Based Estimation". In: *Targeted Learning: Causal Inference for Observational and Experimental Data*. New York, NY: Springer New York, 2011, pp. 459–474.

[98] Hongjian Zhu. "Covariate-adjusted response adaptive designs incorporating covariates with and without treatment interactions". In: *Canadian Journal of Statistics* 43.4 (2015), pp. 534–553.