ABSTRACT

ZHU, LIANGYU. Statistical Methods for Optimal Dose Finding. (Under the direction of Wenbin Lu and Rui Song.)

Statistical methods are becoming increasingly popular for optimizing drug doses in clinical trials. In a typical dose-finding trial, a single optimal dose is decided at the end of the trial and is recommended to all future patients. However, patients might respond differently to the same dose of a drug due to differences in their physical conditions, genetic factors, environmental factors and medication history. Taking patient heterogeneity into consideration when making dose decisions is essential for achieving better treatment results. Traditionally, personalized treatment finding process requires repeating clinical visits of the patient and frequent adjustments of the dosage. Thus the patient is constantly exposed to the risk of underdosing and overdosing during the process. Data driven methods for finding optimal personalized dosage have the potential to shorten the process and lower the risk for the patient. Existing statistical methods for finding personalized treatments are mostly restricted to a finite number of treatment options. In this dissertation, we study the statistical methods for finding the optimal personalized treatment when the treatment options are continuous. The problem is studied under the single-stage setting and the mobile health setting.

In Chapter 2, we study the statistical methods for finding optimal personalized doses under the single-stage setting. We review the existing statistical methods for personalized treatment finding. A kernel-assisted learning method is then proposed for estimating the optimal personalized dosage. Theoretical results and simulation studies are provided for the proposed method. This method is then applied to a wafarin dataset.

In Chapter 3, we study the methodologies for providing personalized dose suggestions using the data collected by health monitoring mobile applications. We are interested in evaluating the causal effect of the time-varying treatments and providing dose suggestions which minimizes the risk of adverse events in a short time period in the future. We review the existing methods for estimating treatment effects with longitudinal data and discuss the limitations of these methods when applied to mobile health data. We extend the definition of the lagged treatment effect by Boruvka et al. (2018) to continuous doses and provide a kernel based estimation method for estimating this lagged treatment effect. We then define a weighted advantage for the doses and estimate the optimal dose which optimizes this weighted advantage. Theoretical results and simulation studies are presented. The proposed method is then applied to a type 1 diabetes dataset collected by

health monitoring applications.

Statistical Methods for Optimal Dose Finding

by
Liangyu Zhu

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Statistics

Raleigh, North Carolina

2020

APPROVED BY:

Marie Davidian	Michael Kosorok
Wenbin Lu Co-chair of Advisory Committee	Rui Song Co-chair of Advisory Committee

DEDICATION

To my parents and to Alex.

ACKNOWLEDGEMENTS

I would like to thank my advisors Dr. Lu and Dr. Song for their patient guidance for my research projects. They expanded my knowledge of statistics research and provided me with inspirations and tools for conducting research on my own. I would like to thank Dr. Davidian and Dr. Kosorok for their valuable feedbacks and suggestions on my projects. I would like to thank all the professors, staff and friends from the Department of Statistics for providing all the help and support during this journey. Finally, I would like to thank my family for their endless love.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER

1

INTRODUCTION

The goal of dose-finding is to come up with safe and efficient drug administration in humans given a certain medical condition. Typical dose-finding trials happen at early phases of clinical experiments where different doses of a new drug are evaluated. An ideal dose-finding study is conducted by a double-blind randomized trial where each patient is randomly assigned a dose among a few safe dose levels for a candidate drug (Chevret 2006). At the end of the trial, a single dose leading to the best average response is determined as a recommendation for future patients. Here, the response is defined as an outcome of interest, it can include measurements of the therapeutic effect of the treatment and the negative symptoms that appear under the clinical or the laboratory settings due to the toxicity of the treatment. Therefore, dose-finding is the art of maximizing the therapeutic effect of the drug while managing the toxicity of the drug at an acceptable level.

However, the toxic effects and the therapeutic effects of the doses usually vary substantially among patients. Patients might respond differently to the same dose of a drug due to differences in their physical conditions, genetic factors, environmental factors and medication histories. The recommended single dose from traditional dose-finding trials might not be the optimal dosage for each individual patient. Taking patients' difference into consideration when making dose decisions is essential for achieving better treatment results. Traditionally, personalized treatment finding process requires repeating clinical

visits of the patient and frequent adjustments of the dosage. Thus the patient is constantly exposed to the risk of underdosing and overdosing during the process. Data driven methods for finding the optimal personalized dosage have the potential to tremendously shorten the process and lower the risk for the patient.

1.1 Motivation

Statistical methods are becoming increasingly popular for optimizing drug doses in clinical trials. However, existing methods for finding the optimal personalized treatments are mostly restricted to the scenario where there are a finite number of treatment options. In particular, people are interested in finding individualized treatment rules (ITR), which output a treatment option within a finite number of available treatments based on patient level information. Such treatment rules can thus be used to guide treatment decisions aiming to maximize the expected clinical outcome of interest, also known as the expected reward or value. An optimal treatment rule is defined to be the one that maximizes the value in the population among a class of treatment rules. Various statistical learning methods have been proposed to infer optimal individualized treatment rules using data from randomized trials or observational studies. Existing methods include model-based approaches, such as Q-learning (Watkins and Dayan 1992; Zhao et al. 2009; Qian and Murphy 2011; Schulte et al. 2014) and A-learning (Murphy 2003; Robins 2004; Henderson et al. 2010; Schulte et al. 2014), and direct value search methods by maximizing a nonparametric estimator of the value function (e.g. Zhang et al. 2012a,b; Zhao et al. 2012).

The above methods, however, are not directly applicable when the number of treatment levels is large. Let us illustrate with warfarin, which is an anticoagulant drug commonly used for the prevention of thrombosis and thromboembolism. Establishing the appropriate dosage for warfarin is known to be a difficult problem because the optimal dosage can vary by a factor of 10 among patients, from 10mg to 100mg per week (Consortium 2009). Incorrect dosages contribute largely to the adverse effects of warfarin usage. Underdosing will fail to alleviate symptoms in patients and overdosing will lead to catastrophic bleeding. In this case, an individualized dose rule (IDR), where a dose level is suggested within a continuous safe dose range according to each individual's physical conditions, would be better at tailoring to patient heterogeneity in drug response. Traditionally, the individualized warfarin dosage for a single patient is decided by repeated clinical visits. A safe initial dosage is given to the patient. The dosage is then adjusted every few weeks while the adverse events are closely monitored by the clinician. During the dose-finding

process, which lasts from weeks to months, the patient is constantly under the risk of overdosing or underdosing. Nowadays, there are organizations gathering an abundant amount of data associated with individual warfarin therapies, including the medical history, the demographic information, the genetic factors, and the therapeutic dosage of the patients. Data-driven methodologies for finding personalized dosages are made possible by the availability of these datasets. A data-driven dosage estimation method has the potential to greatly speed up the dosing process and exempt the patients from the risks of incorrect dosages. Therefore, in this dissertation, we study the statistical methods for providing personalized dosage suggestions which optimize the outcome of interest.

Data-driven treatment recommendations are also obtaining attention with the development of health-monitoring mobile applications. The emergence of electronic technologies such as fetal heart rate monitors, portable hemoglobin meters, self-powered pulse oximeters and continuous glucose monitors allows low-cost monitoring and detection of diseases. Physical conditions of the patients can thus be recorded in real-time. Mobile communication with smart phone devices also improve the process of healthcare by facilitating monitoring and alerting for adverse events and enhancing the interaction between the patients and the doctors (Silva et al. 2015). Important findings have been contributed by mobile health apps for diseases such as cardiology (Martínez-Pérez et al. 2013b; Bisio et al. 2015), diabetes(Arnhold et al. 2014; Sieverdes et al. 2013; Kirwan et al. 2013), obesity (Lopes et al. 2011; Zhu et al. 2010), and alcohol cessation (Carrà et al. 2016; Gustafson et al. 2011). Making treatment adjustments based on the real-time data collected by the mobile applications has the potential to lead to better treatment results and reduce unnecessary over-treatments. For example, diabetes is a chronic disease that requires continuing medical care to reduce the long-term risk caused by high blood glucose. The effectiveness of the medical process largely depends on the patient's self management in daily life, including frequent blood sugar checks, regular insulin use, diet management and consistent exercise (Hood et al. 2016). However, even if the patient successfully adheres to this complex treatment regime, calculating the correct insulin dosage can still be challenging. Hypoglycemia or hyperglycemia due to incorrect insulin dosages may lead to serious side effects in short term and chronic health issues in long term. Both the dosage and the timing of the premeal insulin will largely affect the effectiveness of the treatment. Mobile technologies have presented unique potential to support the self-management process for diabetes patients. Thousands of diabetes apps are available on different mobile operating systems (Martínez-Pérez et al. 2013a). Typical features include self-monitoring for insulin usage, blood glucose, food intake and exercise. Dose suggestions for rapid-reacting insulin are available in part of the applications, most of which are limited to dose suggestions

with simple mathematical operations using planned carbohydrate intake and measured blood glucose (Arnhold et al. 2014). However, the insulin dosage calculation varies largely among patients according to their personal insulin sensitivity factor and carbonhydrate factor (Huckvale et al. 2015). These simple dose rules fail to adapt to the variability among patients and might lead to incorrect dosage suggestions. In addition, the insulin absorption rate can be influenced by other factors including patients' physical activity level and the temperature of the environment. An ambient temperature of 30 Celsius degrees can speed up the insulin absorption by two to four times compared to an ambient temperature of 10 Celsius degrees (Zisser et al. 2008). Therefore, an individualized dose recommendation system based on the time-varying physical condition of the patient is needed for an accurate estimation of the optimal personalized dosage. However, analyzing mobile health data can be challenging because there are typically a large number of time points, time-varying treatments, and a non-definite time horizon (Luckett et al. 2019). In this dissertaton, we also study the statistical methodologies for providing individualized dose suggestions using mobile health data. A detailed outline of the dissertation is given below.

1.2 Outline

The structure of the dissertation is as follows.

In Chapter 2, we focus on statistical methodologies for optimal dose finding. We first review the existing statistical methods for finding optimal individualized treatment rules. The limitations of these methods when applied to continuous dose suggestions are discussed. We then present existing methodologies for finding optimal individualized dose rules (IDR), which output dose suggestions within a safe dose range based the information of the patients. These methodologies are either limited by strict model assumptions for the expected outcome, or have provided no statistical inference for the estimated optimal dose rules. Therefore, we propose a kernel-assisted learning (KAL) method which requires no model assumption for the expected outcome and is capable of providing statistical inference for the estimated dose rules. Theoretical results and simulation studies are presented for the proposed method. It is then applied to a warfarin dataset collected by Consortium (2009).

In Chapter 3, we focus on statistical methods for mobile health data. We are interested in evaluating the causal effect of the time-varying treatments and providing dose suggestions using the data collected by the mobile applications. We first review the existing methods for estimating treatment effects with longitudinal data and discuss

the limitations of these methods when applied to mobile health data. Then we extend the definition of lagged treatment effects by Luckett et al. (2019) to continuous doses under the mobile health setting. A kernel-assisted learning method based on structural nested models (Robins 1994, 2004) is proposed for estimating this lagged treatment effect. Next, we focus on providing dose recommendations based on the estimated lagged treatment effects. Existing treatment recommendation strategies under infinite time horizons typically aim at maximizing an accumulated long-term reward. However, for diseases such as high blood pressure or hyperglycemia, the mobile health interventions actually aim at monitoring adverse events in a short term (Haller et al. 2004; Heron and Smyth 2010). Maximizing the cumulative reward might not be the optimal criteria for providing treatment suggestions under this scenario. Therefore, we define a weighted advantage for the doses as a measurement of the treatment effect over a short period of time. We then estimate the optimal dose which maximizes this weighted advantage. Theoretical results and simulation studies are presented for this estimation method. Finally, we apply the proposed method to the Ohio type 1 diabetes dataset and estimate the optimal dosage for the rapid-reacting insulin doses. Results are presented and discussed at the end of Chapter 3.

1.3 Notations

In this dissertation, we use capital letters such as A, X, Y to denote random variables and lowercase letters such as a, x, y to denote specific values taken by the random variables. The probability mass function of a discrete random variable X is written as $p_X(\cdot)$. The probability density function of a random variable X is written as $f_X(\cdot)$. For any one-variable function $g(\cdot)$, the first, second and third order derivative of $g(\cdot)$ are written as $\dot{g}(\cdot)$, $\ddot{g}(\cdot)$, $\dddot{g}(\cdot)$. We use $I(\cdot)$ to denote the indicator function. For example $I(X > 0)$ is equal to 1 if $X > 0$ and is equal to 0 if $X \leq 0$. For a matrix Σ, $|\Sigma|$ is used to denote the determinant of the matrix. Under the mobile health setting where data are collected at multiple time points $t = 1, \dots T$. We use $\bar{X}_t$ to denote the history of a variable X up to time point t: $\bar{X}_t = \{X_1, \dots, X_t\}$. $\underline{X}_t$ denotes the the random variable from time t to T: $\underline{X}_t = \{X_t, X_{t+1}, \dots, X_T\}$. Let $\bar{X}$ denotes the complete history of variable X: $\bar{X} = \{X_1, \dots, X_T\}$. For any function $g(\cdot)$ of a random variable X, $\mathbb{P}_n g(X)$ is used to denote the empirical average of the function: $\mathbb{P}_n g(X) = \sum_{i=1}^{n} g(X_i)$.

CHAPTER

2

KERNEL ASSISTED LEARNING FOR SINGLE-STAGE PERSONALIZED DOSE FINDING

2.1 Introduction

In this chapter, we explore the statistical methodologies for single-stage personalized dose finding. Here, "single-stage" means that the treatment decision is made only once for each patient and the chosen treatment (or dosage) will be followed by the patient thereafter. There is plenty of literature on methodologies for finding individualized treatment rules (ITR), where one treatment is suggested among a finite number of treatment options based on patient level information. We define the optimal treatment rule as the one which maximizes the expected reward in the population among a class of treatment rules. Individualized treatment rules can thus provide guidance for treatment decisions by selecting the treatment which leads to the most desirable results. Existing methods for finding optimal treatment rules include model-based approaches, such as Q-learning (Watkins and Dayan 1992; Zhao et al. 2009; Qian and Murphy 2011) and A-learning (Murphy 2003; Robins 2004; Henderson et al. 2010), and direct value search methods

(Zhang et al. 2012a,b; Zhao et al. 2012).

The above methods, however, are not directly applicable when the number of possible treatment levels is large. Using continuous individualized dose rules (IDR), where a dose level is suggested within a safe dose range, can better adapt to patient heterogeneity in drug response. Extending the above methods for finding optimal individualized treatment rules to continuous treatment options is non-trivial. Several methods have been proposed for finding optimal individualized dose rules. One way of extending existing methods to the continuous dose case is to discretize the dose levels. Laber and Zhao (2015) proposed a tree-based method and turned the problem into a classification problem by dividing patients into subgroups and assigning a single dose to each subgroup. Chen et al. (2018) extended the outcome weighted learning method (Zhao et al. 2012) from binary treatment settings to ordinal treatment settings. However, in cases where patient responses are sensitive to dose changes, a discretized dose rule with a small number of levels will fail to provide dose recommendations leading to optimal clinical results. On the other hand, a discretized dose rule with a large number of levels may result in limited observations within each subgroup, and thus may be at risk of overfitting.

Alternatively, Rich et al. (2014) extended the Q-learning method by modeling the interactions between the dose level and covariates with both linear and quadratic terms in doses. However, such a parametric approach is sensitive to model misspecification and the estimated individualized dose rule might be far away from the true optimal dose rule. In addition, it cannot be guaranteed that the estimated optimal dose falls into the safe dose range. More recently, Chen et al. (2016) extended the outcome weighted learning method proposed by Zhao et al. (2012) and transformed the dose-finding problem into a weighted regression with individual rewards as weights. The optimal dose rule is then obtained by optimizing a non-convex loss function. This method is robust to model misspecification and has appealing computational properties, however, the associated statistical inference for the estimated dose rule is challenging to determine.

Therefore, we propose a kernel assisted learning method is proposed to infer the optimal individualized dose rule in a manner which enables statistical inference. The proposed method can be viewed as a direct value search method. Specifically, we first estimate the value function with a kernel based estimator. Then we search for the optimal individualized dose rule within a prespecified class of rules where the suggested doses always lie in the safe dose range. The proposed method is robust to model misspecification and is applicable to data from both randomized trials and observational studies. We establish the consistency and asymptotic normality of the estimated parameters in the obtained optimal dose rule. In particular, the asymptotic covariance of the estimators is

derived based on nontrivial calculations of the expectation of a U-statistic.

The remainder of the chapter is organized as follows. In Section 2.2, we present the problem setting and the notations for optimal dose finding. In Section 2.3, existing methods for finding the optimal ITR and the optimal IDR are reviewed. Our proposed method is presented in Section 2.4. The theoretical results of the estimated parameters are established in Section 2.5. In Section 2.6, we demonstrate the empirical performance of the proposed method via simulations. In Section 2.7, the proposed method is further illustrated with an application to a warfarin study. Some discussions and conclusions are given in Section 2.8. Proofs of the theoretical results are provided in Section 2.9.

2.2 Problem Setting

We first present the single-stage personalized dose-finding problem with a statistical framework. Assume that the observed data consist of n independent and identically distributed observations $\{(X_i, A_i, Y_i)\}_{i=1}^n$, where $X_i \in \mathcal{X}$ is a q-dimensional vector of covariates for the ith patient, $A_i \in \mathcal{A}$ is the dose assigned to the patient with $\mathcal{A}$ being the safe dose range (if $\mathcal{A}$ is a set of finite number of treatment options, then the problem transform into finding the optimal individualized treatment rule), and $Y_i \in \mathbb{R}$ is the observed outcome of interest after the treatment A_i is given to the patient. Without loss of generality, we assume that larger Y means better outcome. Let $\pi(X)$ denote an individualized dose rule, which is a deterministic mapping function from $\mathcal{X}$ to $\mathcal{A}$. To define the value function of an individualized dose rule, we use the potential outcome framework (Rubin 1978). Specifically, let $Y^*(a)$ be the potential outcome that would be observed when a dose level $a \in \mathcal{A}$ is given. Define the value function as the expected potential outcome in the population if everyone follows the dose rule π, i.e. $V(\pi) = E[Y^*\{\pi(X)\}]$. The optimal individualized dose rule is defined as $\pi^{opt} = \arg\max_{\pi \in \mathcal{G}} V(\pi)$, where $\mathcal{G}$ is a class of interested dose rules. For example, a linear class of dose rules can be written as: $\mathcal{G} = \{\pi_\beta(X) = \beta^T X : \beta \in \mathbb{R}^q\}$.

In order to estimate the value function from the observed data, we need to make the following three assumptions similar to those adopted in the causal inference literature (Robins 2004).

- $Y = \int_{\mathcal{A}} \delta(A = a) Y^*(a) da$, where $\delta(\cdot)$ is the Dirac delta function. This corresponds to the stable unit treatment value assumption (also known as the consistency assumption). This assumption assumes that the observed outcome is the same as the potential outcome had the dosage given to the patient be the actual dose.

This assumption implies that there is no interference among patients. In the case where the treatment options are discrete, this assumption can be written as: $Y = \sum_{a \in \mathcal{A}} I(A = a) Y^*(a)$.

- The potential outcomes $\{Y^*(a) : a \in \mathcal{A}\}$ are conditionally independent of A given X, which is also known as the no unmeasured confounders assumption. This assumption can be naturally satisfied by the design of a randomized dose trial. However, it cannot be validated in an observational study.

- There exists a $c > 0$ such that $f_{A|X}(A = a|X = x) \geq c$ for all $a \in \mathcal{A}, x \in \mathcal{X}$, where $f_{A|X}(a|x)$ is the conditional density of $A = a$ given $X = x$. In the discrete treatment case, the assumption can be written as: $p_{A|X}(A = a|X = x) > 0$ for all $a \in \mathcal{A}, x \in \mathcal{X}$. This assumption ensures that $V(\pi)$ can be estimated using the observed data for all $\pi : \mathcal{X} \to \mathcal{A}$. This is also known as the positivity assumption.

Under these assumptions, we can show that $V(\pi)$ can be estimated with the observed data :

$$\begin{aligned} V(\pi) &= E[Y^*(\pi(X))] \\ &= E_X[E\{Y^*(\pi(X))|X\}] \\ &= E_X[E\{Y^*(\pi(X))|A = \pi(X), X\}] \\ &= E_X[E\{Y|A = \pi(X), X\}]. \end{aligned}$$

The second equation above is based on the basic property of conditional densities. The third equation above is valid because of the no unmeasured confounder assumption. The fourth equation is based on the consistency assumption. The positivity assumption ensures that the right side of the last equation can be estimated empirically.

2.3 Literature Review

2.3.1 Methods for Finding Discrete Optimal Treatment Rules

Regression Based Method

We first consider the special case where there are only two treatment options, $\mathcal{A} = \{0, 1\}$. One straightforward approach is the Q-learning method (Watkins and Dayan 1992; Sutton et al. 1998; Zhao et al. 2009; Schulte et al. 2014), which first posits a parametric regression model for $\mu(A, X) = E(Y|A, X)$, say $\mu(A, X; \beta)$, where β are the related parameters. If

this model is correctly specified, then there exists some $\tilde{\beta}$ where $\mu(A, X) = \mu(A, X; \tilde{\beta})$. The optimal treatment rule would be $\pi^{opt}(X) = \arg\max_{\pi} \mu(\pi(X), X) = I\{\mu(1, X; \tilde{\beta}) > \mu(0, X; \tilde{\beta}))\}$. Therefore, we can estimate the optimal dose rule by obtaining an estimate of the parameters from the data, $\hat{\beta}$. Then $\hat{\pi}^{opt}(X) = I\{\mu(1, X; \hat{\beta}) > \mu(0, X; \hat{\beta})\}$. The estimation of the parameters can be achieved through standard ordinary least squares (OLS) or weighted least squares (WLS). An estimate of the maximum value function $V(\pi^{opt}) = E\{Y^*(\pi^{opt})\}$ would be:

$$\frac{1}{n}\sum_{i=1}^{n}\Big[\mu(1, X_i, \hat{\beta})\hat{\pi}^{opt}(X_i) + \mu(0, X_i, \hat{\beta})\{1 - \hat{\pi}^{opt}(X_i)\}\Big].$$

Extension of this method to more than two treatments is straightforward. However, an obvious drawback of this method is that the estimated treatment rule may be far away from the optimal rule when the posited model is not correct. The estimated maximum value function might not converge to the true maximum value function in such cases.

Notice that the optimal treatment rule in this case $\pi^{opt}(X) = I\{\mu(1, X) > \mu(0, X))\}$ only depends on the contrast, which is the difference of the expected reward between the treatments given the patient's information, $C(X) = \mu(1, X) - \mu(0, X)$. The A-learning method (Blatt et al. 2004; Robins 2004; Moodie et al. 2009; Henderson et al. 2010; Wallace and Moodie 2015) utilizes this idea by positing a model for the contrast function, say $C(X; \psi)$. The original μ-function can be written as $\mu(A, X) = h(X) + A \times C(X; \psi)$, where the form of $h(X) = \mu(0, X)$ does not need to be specified. The optimal treatment rule would thus be $\pi^{opt} = I\{C(X) > 0\})$. Let $\tilde{p}(X) = p(A = 1|X)$ be the propensity score of receiving treatment $A = 1$. Robins (2004) shows that the parameters ψ can be estimated consistently by the following estimation equation:

$$\sum_{i=1}^{n}\lambda(X_i)\Big\{A_i - \tilde{p}(X_i)\Big\} \times \Big\{Y_i - A_i C(X_i; \psi) - \theta(X_i)\Big\} = 0. \tag{2.1}$$

where $\lambda(X_i)$ is an arbitrary function with the same dimension as ψ and an arbitrary function $\theta(X)$. This method is also referred to as g-estimation. Robins (2004) also shows that, if the model $C(X; \psi)$ is correct and $\text{var}(Y|X)$ is constant, the optimal choice of $\lambda(X; \psi)$ will be $\partial C(A, X; \psi)/\partial\psi$, and the optimal choice of $\theta(X)$ will be $h(X)$.

In order to apply Equation (2.1) to obtain the estimated the parameters ψ, we can posit models for $\theta(X)$, say $h(X; \beta)$. If the data are not from a randomized trial, we also need to posit model for the propensity score, say $\tilde{p}(X; \gamma)$. Robins (2004) also shows that the g-estimation method has double robustness, meaning that Equation (2.1) yields

a consistent estimator of ψ as long as one of these two models is correctly specified. After the estimated parameter $\hat{\psi}$ is obtained, the estimated optimal dose rule would be $\hat{\pi}^{opt}(X) = I\{C(X;\hat{\psi}) > 0\}$.

Compared to the Q-learning method, the A-learning method is more robust to model misspecification because the exact form of the expected reward function does not need to be specified. Extensions to more than two-treatments (see Robins 2004; Moodie et al. 2007) are possible with more complicated formulations. However, the performance of the method still largely depends on the correct specification of the model for the contrast function $C(X)$. Schulte et al. (2014) also show that when the models for Q-learning and A-learning are both correctly specified, the A-learning might yield relatively inefficient inference for ψ without an optimal choice of λ. Furthermore, extensions to the continuous dose finding problem are not feasible because the contrast function cannot be easily defined with an infinite number of treatment options.

Inverse Probability Weighted Estimation Method

In certain cases, the treatment rule can be defined by a simplified class of rules involving only a subset of the covariates (Zhang et al. 2012b). For example, if $\mu(A, X; \beta) = \beta_1 X_1 + A(\beta_2 X_2 + \beta_3 X_3)$, then the optimal treatment rule would be $I\{X_2 + (\beta_3/\beta_2)X_3 > 0\}$ or $I\{X_2 + (\beta_3/\beta_2)X_3 < 0]\}$ depending on the sign of β_2. Therefore, the optimal treatment can be defined by the parameter space $\{\eta : \eta = \beta_2/\beta_3\}$. Let the class of treatment rules of the form $\pi_\eta(X) = \pi(X, \eta) = I(X_2 + \eta X_3 > 0)$ or $I(X_2 + \eta X_3 < 0)$ be $\mathcal{G}_\eta$. Zhang et al. (2012b) proposed a more robust method by directly searching for the optimal treatment rule within the class $\mathcal{G}_\eta$ without assuming any models for $\mu(A, X)$. The goal thus becomes to find $\eta^{opt} = \arg\max_\eta E\{Y^*(\pi_\eta)\}$. This can be achieved by first obtaining a nonparametric approximation of $E\{Y^*(\pi_\eta)\}$, then find the $\hat{\eta}$ which maximizes the approximated value function.

However, for a specific treatment rule π_η, the potential outcome $Y^*(\pi_\eta(x))$ is probably not observed for some $x \in \mathcal{X}$. Thus, estimating the value function can be regarded as a missing data problem, where $\{Y^*(\pi_\eta)(X), X\}$ is the full dataset and $\{I(A = \pi_\eta(X)), YI(A = \pi_\eta(X)), X\}$ is the observed dataset. Zhang et al. (2012b) proposed to estimate the value function with an inverse probability weighted estimator (IPWE). Assume that the data come from a randomized trial where the propensity score $\tilde{p}(X) = p(A = 1|X)$ is known. (In observational studies where the propensity score is unknown, it can be estimated by a parametric model, say $\tilde{p}(X;\gamma)$.) Let $\tilde{p}_c(X;\eta) = \tilde{p}(X)\pi(X,\eta) + \{1 - \tilde{p}(X)\}\{1 - \pi(X,\eta)\}$ be the probability of observing the outcome given by treatment rule $\pi(X, \eta)$. For a fixed

η, the inverse probability weighted estimator of the value function is:

$$\text{IPWE}(\eta) = \frac{1}{n}\sum_{i=1}^{n}\frac{Y_i I\{A_i = \pi_\eta(X)\}}{\tilde{p}_c(X_i;\eta)} = \frac{1}{n}\sum_{i=1}^{n}\frac{Y_i I\{A_i = \pi_\eta(X)\}}{\tilde{p}(X_i)^{A_i}\{1-\tilde{p}(X_i)\}^{1-A_i}}. \tag{2.2}$$

This estimator is shown to be consistent for $E\{Y^*(\pi_\eta)\}$ when $\tilde{p}(X)$ is known or $\tilde{p}(X;\gamma)$ is a correctly specified model for $\tilde{p}(X)$ (Cao et al. 2009) . Cao et al. (2009) also proposed another estimator of the value function with improved robustness. It is known as the augmented inverse probability weighted estimator (AIPWE):

$$\text{AIPWE}(\eta) = \frac{1}{n}\sum_{i=1}^{n}\Big\{\frac{Y_i I\{A_i = \pi_\eta(X_i)\}}{\tilde{p}_c(X_i;\eta)} - \frac{I\{A_i = \pi_\eta(X_i)\} - \tilde{p}_c(X_i;\eta)}{\tilde{p}_c(X_i;\eta)} m(X_i;\eta,\hat{\beta})\Big\}, \tag{2.3}$$

where,

$$m(X;\eta,\beta) = \mu(1,X;\beta)\pi(X,\eta) + \mu(0,X;\beta)\Big\{1-\pi(X,\eta)\Big\}$$

is an approximation for $E\{Y^*(\pi_\eta)\}$; $\mu(A,X;\beta)$ is a working model for $E(Y|A,X)$ with parameters β; $\hat{\beta}$ can be obtained using regular regression methods. The AIPWE is consistent for $E\{Y^*(\pi_\eta)\}$ as long as $\tilde{p}(X)$ is known, or a correct model is specified for $\tilde{p}(X)$, or $\mu(A,X;\beta)$ is correctly specified. If the propensity score is known or can be approximated with a correctly specified model, the estimation efficiency is improved when the working model $\mu(A,X;\beta)$ is also correctly specified (Zhang et al. 2012b). Then the estimated optimal treatment rule is $\hat{\pi}_\eta^{opt} = \pi(X,\hat{\eta}^{opt})$, where $\hat{\eta}^{opt}$ maximizes the corresponding estimator of the value function: $\hat{\eta}^{opt} = \arg\max_\eta \text{IPWE}(\eta)$ or $\hat{\eta}^{opt} = \arg\max_\eta \text{AIPWE}(\eta)$.

This type of methods, where the optimal treatment is estimated by directly searching within a prespecified class of treatment rules to optimize a certain metric, are known as direct methods. Zhang et al. (2012b) showed that the proposed direct method can achieve comparable performance to the regression-based methods with the addition advantage of robustness to model misspecification. One limitation of this method is that maximization of the approximated value function might be complicated or computationally expensive due to the nonsmoothness. Zhang et al. (2012b) resolved the optimization problem using classification and regression tree (CART). Laber and Zhao (2015) further extended the method by searching the optimal treatment rule directly within a nonparametric class of rules that can be represented by decision trees.

Outcome Weighted Learning

Another direct method is the outcome weighted learning (O-learning) method proposed by Zhao et al. (2012). The optimal treatment rule is regarded as a classification problem where the goal is to classify the patients into groups by their optimal treatments. Instead of maximizing an approximate of the value function, the outcome weighted learning method finds the optimal rule by directly minimizing a loss function which can be interpreted as a classification error. Assume that the data is from a randomized trial where $\tilde{p}(X) = p(A = 1|X)$ is a constant $\tilde{p}$, and assume that there are only two treatment options $\mathcal{A} = \{0, 1\}$. The value function can thus be written as:

$$V(\pi) = E\{Y^*(\pi)\} = E\Big\{\frac{YI(A = \pi(X))}{p(A|X)}\Big\} = E\Big\{\frac{YI(A = \pi(X))}{\tilde{p}A + (1 - \tilde{p})(1 - A)}\Big\}. \tag{2.4}$$

Maximizing Equation (2.4) is equivalent to minimizing:

$$E(Y|A = 1) + E(Y|A = -1) - V(\pi) = E\Big\{\frac{YI(A \neq \pi(X))}{\tilde{p}A + (1 - \tilde{p})(1 - A)}\Big\}. \tag{2.5}$$

If we regard $\pi(X)$ as a classification of the patients into two groups corresponding to the two treatment and weigh each misclassification event by $Y/\{\tilde{p}A + (1 - \tilde{p})(1 - A)\}$, the right side of Equation (2.5) will be an expectation of the classification error and can thus be estimated by the empirical classification error:

$$\frac{1}{n}\sum_{i=1}^{n}\frac{Y_i}{\tilde{p}A_i + (1 - \tilde{p})(1 - A_i)}I\{A_i \neq \pi(X_i)\}.$$

Notice that with this method, we would prefer a classification which is consistent with the assigned dose where the observed outcome is large. When the observed outcome is small, the assigned treatment is less likely to be recommended by the treatment rule. Similar to the robust method proposed by Zhang et al. (2012b), we search for the optimal treatment rule within the class of treatment rules $\mathcal{G}_\eta$ in the form of $\pi_\eta(X) = \pi(X; \eta) = I\{g(X; \eta) > 0\}$, where $g(X; \eta)$ is a prespecified function of X indexed by parameter η. The optimization problem thus becomes:

$$\hat{\eta}^{opt} = \arg\min_{\eta}\frac{1}{n}\sum_{i=1}^{n}\frac{Y_i}{\tilde{p}A_i + (1 - \tilde{p})(1 - A_i)}I\Big\{A_i \neq I(g(X_i; \eta) > 0)\Big\}. \tag{2.6}$$

However, the nonsmoothness of the loss function (the function we aim to minimize in Equation (2.6)) adds complexity to the optimization process. Zhao et al. (2012) resolved

this issue by substituting the indicator function $I\{g(X;\eta) > 0\}$ with a hinge loss function $\{1 - Ag(X;\eta)\}^+$ where $x^+ = \max(x, 0)$ (see Cortes and Vapnik 1995). A penalty term is added for the complexity of the treatment rule to avoid overfitting. Then,

$$\hat{\eta}^{opt} = \arg\min_{\eta} \frac{1}{n} \sum_{i=1}^{n} \frac{Y_i}{\tilde{p}A_i + (1-\tilde{p})(1-A_i)} \{1 - A_i g(X_i;\eta)\}^+ + \lambda_n \|\pi_\eta\|^2, \tag{2.7}$$

where $\|\pi\|$ is some seminorm for π (For example, when $\pi_\eta(X) = I\{X^T\eta > 0\}$, then $\|\pi_\eta\|$ can be the Euclidean norm of η), and λ_n is a tuning parameter which controls the strength of the penalty on the complexity of the treatment rule.

Both Zhang et al. (2012b)'s method and Zhao et al. (2012)'s method are based on finding an optimal treatment rule within a class of rules by optimizing an objective function. Zhang et al. (2012a) further extended these methods by providing a framework which unifies both of these methods into a weighted classification problem where the class of treatment rules does not need to be prespecified. However, this classification perspective is not feasible when the treatment options are continuous. Extensions of these methods to continuous dose finding are nontrivial.

2.3.2 Methods for Finding Optimal Dose Rules

The methods discussed in Section 2.3.1 provide different approaches to the problem of optimal treatment finding. However, when the number of possible treatment levels is large, the application of these methods would be problematic. For example, the dosage for the drug warfarin (as mentioned in Chapter 1) typically range from 10 mg to 100 mg weekly. Classifying the doses into 91 different levels and regarding this as a discrete treatment finding problem would not be ideal because each level would be only assigned to a small number of patients. Estimation of the value function might thus have large variance. The information in the data would not be used efficiently either. Regarding the treatment options as continuous doses is more appropriate in such cases. A randomized dose trial, where each patient is randomly assigned a dose within a safe dose range, is needed for finding optimal treatment rule with continuous doses. Such rules are also known as individualized dose rules (IDR). In the following discussion, we assume that $\mathcal{A}$ is an interval of the safe dose range. Without loss of generality, we assume that $\mathcal{A} = [0, 1]$.

Discretized Dose Rules

One approach to find an optimal treatment rule for continuous doses is to discretize the feasible treatment options to a small number of dose levels. Laber and Zhao (2015)

extended the AIPWE method of Zhang et al. (2012b) by first discretizing the treatment levels into a finite number of treatments, and then transforming the optimal treatment rule finding problem into a clustering problem among the patients. Similar to the method proposed by Zhang et al. (2012b), Laber and Zhao (2015) proposed to directly search for the optimal rule within a class of treatment rules, which is the class of treatment rules that can be represented by decision trees in this case. The optimal rule is estimated by maximizing an approximation of the value function. Within the context of a clustering problem, the objective function to maximize (the approximated value function) can be regarded as the purity measure for evaluation of the quality of the clustering (see Larson 2010). The optimal treatment rule within the class can thus be found by establishing a decision tree, where the algorithm recursively partition the covariate space into rectangular sets based on maximizing the total purity measure (see Breiman 1984; Hastie et al. 2009, for details).

Methods like IPWE and AIPWE are not directly applicable for estimating the value function, because the assigned doses to each patient are not restricted to the discretized dose levels. The probability of observing the outcome of the dose levels suggested by the corresponding rule would be 0. Laber and Zhao (2015) proposes a smoothed version of the AIPWE by replacing the indicator function $I\{a = \pi(x)\}$ in Equation (2.3) with a kernel smoother: $v_{\pi,h}(a|x) = h^{-1}k\Big(g(a) - g(\pi(x))\}/h\Big)\dot{g}(a)$, where $k(\cdot)$ is a symmetric density function, h is the bandwidth, $g(\cdot)$ is a prespecified one-to-one function from the treatment space to $\mathbb{R}$ and $\dot{g}(\cdot)$ is the derivative of the function $g(\cdot)$. One simple example of the kernel smoother would be $v_{\pi,h}(a|x) = (2h)^{-1}\dot{g}(a)I\Big\{|g(a) - g(\pi(x))| \leq h\Big\}$. The treatment rule $\pi(x)$ can thus be approximated by a class of distributions over $\mathcal{A}$ that has the mass around $\pi(x)$. A smoothed version of the value function $V(\pi) = E\{Y^*(\pi)\}$ can be written as $E\Big\{Yv_{\pi,h}(A|X)/f_{A|X}(A|X)\Big\}$. Here, the $f_{A|X}(a|x)$ is the conditional density of A given X. A smoothed version of the AIPWE estimator in Equation (2.3) would be:

$$\frac{1}{n}\sum_{i=1}^{n}\Big[\frac{\big\{Y_i - m(X_i;\beta)\big\}v_{\pi,h}(A_i|X_i)}{f_{A|X}(A_i|X_i)} + m(X_i;\beta)\Big], \tag{2.8}$$

where $m(X;\beta)$ is a parametric model for $\mu(A,X)$ with parameters β. Maximizing Equation (2.8) is equivalent to maximizing:

$$\hat{C}(\pi) = n^{-1}\Big[\frac{\{Y_i - m(X_i;\beta)\}v_{\pi,h}(A_i|X_i)}{f_{A|X}(A_i|X_i)}\Big].$$

The purity measure defined for the splitting of the nodes is based on $\hat{C}(\pi)$. Suppose

that for a subset of the covariate space $r \subset \mathcal{X}$, $\pi_{r,a,a'}$ denotes the rule that assigns treatment a for all $x \in r$ and treatment a' for all $x \in r^c$, where a, a' are within the discretized treatment space. For a node defined with a subset of the covariate space $\tilde{r} \subset \mathcal{X}$ and another subset $r \subset \mathcal{X}$, the purity measure of partitioning $\tilde{r}$ into $\tilde{r} \cap r$ and $\tilde{r} \cup r^c$ is defined as:

$$\mathcal{P}(\tilde{r}, r) = \left[\mathbb{P}_n \frac{\{Y - m(X, \hat{\beta})\} I(X \in \tilde{r}) v_{\pi_{r,a,a'},h}\{A|X\}}{f_{A|X}(\pi_{r,a,a'}(X)|X)}\right] \left[\mathbb{P}_n \frac{I(X \in \tilde{r}) v_{\pi_{r,a,a'},h}\{A|X\}}{f_{A|X}(\pi_{r,a,a'}(X)|X)}\right]^{-1}. \quad (2.9)$$

The algorithm starts from the root node defined by the whole parameter space. Each time it finds the split of a node where the purity measure is maximized. The algorithm ends when the maximal increase of purity measure is below a predefined threshold. The rule that corresponds to the partition of the final tree is the estimated optimal treatment rule.

This tree-based method provides an approach to extension of an existing method to finding optimal dose rules. However, when the number of discretized intervals is small, the suggested dose rule may fail to provide a precise estimation of the optimal dosage for the patients. As the number of discretized intervals increase, the variance of the purity measure as defined in Equation (2.9) will increase tremendously because there are a limited number of observations within each interval. Therefore, continuous dose rules would better adapt to the complicated mechanisms of the drugs and the heterogeneous needs of the patients.

Regression based method

Another natural approach is to establish regression models for $\mu(A, X) = E(Y|A, X)$ just as when the treatment options are discrete. When the dose level is continuous, it is essential to consider a nonlinear relationship between the dose and the outcome, because overdosing often leads to toxicity while underdosing might result in inefficiency of the medicine. Rich et al. (2014) applied this idea and used a linear regression model including both quadratic terms and interaction terms between the covariates and the treatment. A general form of the model can be written as follows (Laber and Zhao 2015):

$$\mu(A, X; \rho, \theta, \gamma) = X^T\rho + X^T\theta(A - X^T\gamma)^2. \quad (2.10)$$

where ρ, θ, γ are the parameters to be estimated. If $X^T\theta$ is positive, the optimal dose rule is $X^T\gamma$ and can thus be estimated by $\hat{\pi}^{opt}(X) = X^T\hat{\gamma}$. The parameter γ can be easily

interpreted as the effect of the covariates on the optimal IDR. It can be further extended to LASSO regression by including penalties of L_2 norms of the parameters in the loss function while estimating the parameters (Chen et al. 2016). Other variations of the regression based method include support vector regression (SVR; Vapnik 2013; Smola and Schölkopf 2004) and regression trees (Breiman 1984).

The estimated optimal dose rule by this method depends on the correctness of the model specified for $\mu(A, X)$. When the model is misspecified, the suggested dose rule may be far away from the true optimal dose rule. If $X^T\theta$ is not always positive, then the suggested optimal dose rule will fall on the edges of the dose range, meaning that the suggested dose is either the maximum feasible dose or no dose at all, which is not ideal in practice. Even if $X^T\hat{\theta}$ is positive, the optimal dose estimated by $X^T\hat{\gamma}$ may not fall into the safe dose range and thus not applicable in practice.

Outcome Weighted Learning

To avoid model misspecification, Chen et al. (2016) proposed a direct method by extending the outcome weighted learning method proposed by Zhao et al. (2012). Direct application of O-learning method to the continuous dose finding problem is not feasible because the probability of a certain dose a given covariates $x \in \mathcal{X}$, denoted as $p_{A|X}(a|x)$ in Equation (2.4), would be 0. Few subjects would satisfy $A_i = \pi(X_i)$, which will make the approximation of the value function unstable.

Chen et al. (2016) first extend Equation (2.4) by the replacing the $I\{A = \pi(X)\}$ with an indicator function $I\{A \in (\pi(X) - \phi, \pi(X) + \phi)\}$, where ϕ is a bandwidth. They show that the $\tilde{V}_\phi(\pi)$ defined below converges to the value function $V(\pi)$ when ϕ converge to 0:

$$\tilde{V}_\phi(\pi) = \frac{1}{2\phi} E\Big[\frac{Y I\{A \in (\pi(X) - \phi, \pi(X) + \phi)\}}{f_{A|X}(A|X)}\Big], \tag{2.11}$$

where $f_{A|X}$ here denotes the conditional density of A given X. As a result, the π^{opt}, which maximizes $\tilde{V}_\phi(\pi)$ and equivalently minimizes

$$E\Big[\frac{Y I\{|A - \pi(X)| > \phi\}}{2\phi f_{A|X}(A|X)}\Big],$$

will approximately maximize $V(\pi)$. Since the nonsmoothness induced by the indicator function makes it difficult for optimization, the author used a continuous loss function to

approximate the indicator function:

$$l_\phi\Big(A-\pi(X)\Big)=\min\Big(\frac{|A-\pi(X)|}{\phi},1\Big). \tag{2.12}$$

The function l_ϕ can be written as the difference of two convex functions and thus the difference of convex functions (DC) algorithm (see Le Thi Hoai and Tao 1997) can be used for the optimization of the loss function. The function to be minimized thus becomes:

$$\mathcal{R}_\phi(\pi)=E\Big[\frac{Yl_\phi\big(A-\pi(X)\big)}{\phi f_{A|X}(A|X)}\Big], \tag{2.13}$$

which can be estimated consistently with the empirical expectation:

$$\hat{\mathcal{R}}_\phi(\pi)=\mathbb{P}_n\frac{Yl_{\phi_n}[A-\pi(X)]}{\phi_n f_{A|X}(A|X)}, \tag{2.14}$$

where ϕ_n is some constant that goes to 0 as n goes to infinity. By including a term for penalizing the complexity of $\pi(X)$, the optimization problem is formalized as follows:

$$\pi^{opt}=\arg\min_\pi\Big\{\mathbb{P}_n\frac{Yl_{\phi_n}[A-\pi(X)]}{\phi_n f_{A|X}(A|X)}+\lambda_n\|\pi\|^2\Big\}, \tag{2.15}$$

where λ_n is the tuning parameter which controls the complexity of the dose rule to avoid overfitting.

The optimal dose rule which minimizes the loss function is searched directly within a class of rules. The author considers two classes of dose rules. The first class is the linear rules which can be written as $\pi(X)=X^Tw+b$. This method is also known as the linear outcome weighted learning (L-O-Learning). The second class is the non-linear rules defined in a reproducing kernel Hilbert space (RKHS) (see Vapnik 2013; Smola and Schölkopf 2004). An RKHS H_K associated with a symmetric continuous kernel function $K(\cdot,\cdot):\mathcal{X}\times\mathcal{X}\to\mathbb{R}$ is the complete span of all functions $K(\cdot,X),X\in\mathcal{X}$. Then there exists a function $\Phi(\cdot)$ that $\Phi(X_i)^T\Phi(X_j)=K(X_i,X_j)$. The dose rules from H_K can thus be represented by $\pi(X)=w^T\Phi(X)+b$ (Vapnik 2013). The non-linear outcome weighted learning is therefore also known as kernel-based outcome weighted learning (K-O-Learning).

This method provides a highly flexible model for the optimal dose rule without defining the relationship between the dose and the expected reward and is thus robust to model misspecification. However, no statistical inference has been provided for the estimated optimal dose. The suggested dose might be far from the actual optimal dose. Therefore,

we propose a kernel assisted learning method in Section 2.4 and present the statistical inference with this proposed method. This proposed method is also a direct method which imposes no assumption on the form of the conditional expectaion of the outcome $E(Y|A, X)$.

2.4 Kernel Assisted Learning

2.4.1 Method

In order to estimate the optimal IDR, we first estimate $V(\pi)$ with a kernel based estimator and then estimate π^{opt} by directly maximizing the estimated value function $\hat{V}(\pi)$. We search for the optimal individualized dose rule within a class of dose rules of the form: $\pi_\beta(x) = \pi(x;\beta) \in \mathcal{G}$, where $\mathcal{G} = \{g(\beta^T x), \beta \in \mathbb{R}^q\}$, and $g : \mathbb{R} \to \mathcal{A}$ is a predefined link function to ensure that the suggested dosage is within the safe dose range. Let $\beta^* = \arg\max_\beta V(\pi_\beta)$, then the optimal IDR within $\mathcal{G}$ is:

$$\pi_\beta^{opt}(X) = \pi(X;\beta^*) = \arg\max_{\pi\in\mathcal{G}} V(\pi).$$

If the true optimal IDR: $\pi^{opt} = \arg\max_\pi V(\pi) \in \mathcal{G}$, then $\pi_\beta^{opt}(X) = \pi^{opt}$. To see this, we illustrate with a toy example. If the true model for $E(Y|A,X)$ takes the form: $E(Y|A,X) = \tilde{\mu}(X) + Q\{A - g(\tilde{\beta}^T X)\}H(X)$, where $\tilde{\mu}(X)$ is an unspecified baseline function, $H(X)$ is a non-negative function and $Q(\cdot)$ is a unimodal function which is maximized at 0, then $E(Y \mid A, X)$ is maximized at dose level $A = g(\tilde{\beta}^T x)$ for patients with covariates $X = x$. Thus, the true optimal individualized dose rule is:

$$\begin{aligned}
\pi^{opt}(X) &= \arg\max_\pi V(\pi) \\
&= \arg\max_\pi E_X\Big[E\big\{Y \mid A = \pi(X), X\big\}\Big] \\
&= \arg\max_\pi E_X\Big[\tilde{\mu}(X) + Q\big\{\pi(X) - g(\tilde{\beta}^T X)\big\}H(X)\Big] \\
&= \arg\max_\pi E_X\Big[Q\big\{\pi(X) - g(\tilde{\beta}^T X)\big\}H(X)\Big] \\
&= g(\tilde{\beta}^T X) \in \mathcal{G}.
\end{aligned}$$

The last equation above is true because $Q\{\pi(X) - g(\tilde{\beta}^T X)\}H(X)$ is maximized at $g(\tilde{\beta}^T X)$ for each $X \in \mathcal{X}$. If a unique maximizer of $V(\pi_\beta)$ exists, then

$$\beta^* = \arg\max_\beta V(\pi_\beta) = \arg\max_\beta E_X\Big[Q\big\{g(\beta^T X) - g(\tilde{\beta}^T X)\big\}H(X)\Big] = \tilde{\beta}.$$

Therefore, $\pi_\beta^{opt} = g(\beta^{*T} X) = \pi^{opt}$. Notice that if $\pi^{opt} = \arg\max_\pi V(\pi) \notin \mathcal{G}$, then $\pi_\beta^{opt} \neq \pi^{opt}$. However, π_β^{opt} is still of interest as long as the form of $\mathcal{G}$ is flexible enough, because it maximizes the value function among this set of treatment rules. Therefore, we estimate β^* using $\hat{\beta} = \arg\max \hat{V}(\pi_\beta)$, and the optimal IDR within $\mathcal{G}$ can be estimated with $\hat{\pi}_\beta^{opt} = \pi(X;\hat{\beta})$. Notice that we do not need any model assumption on the form of

the conditional expectation $E(Y|A,X)$ to apply this method.

Next, we propose a kernel based estimator for the value function. Let

$$M(\beta) = V\big(\pi_\beta\big) = \int_{x\in\mathcal{X}} m\Big(x, g(\beta^T x)\Big) f_X(x) dx,$$

where $m(x,a) = E(Y \mid X = x, A = a)$ and $f_X(x)$ is the marginal density of X. Thus, $\beta^* = \arg\max_\beta M(\beta)$. The function $m\Big(x, g(\beta^T x)\Big)$ is estimated using the Nadaraya-Watson method given:

$$\hat{m}\Big(x, g(\beta^T x)\Big) = \frac{\frac{1}{n}\sum_{i=1}^n Y_i \frac{1}{h_x^q} K_q(\frac{x-X_i}{h_x}) \frac{1}{h_a} K\big\{\frac{g(\beta^T x)-A_i}{h_a}\big\}}{\frac{1}{n}\sum_{i=1}^n \frac{1}{h_x^q} K_q(\frac{x-X_i}{h_x}) \frac{1}{h_a} K\big\{\frac{g(\beta^T x)-A_i}{h_a}\big\}}, \tag{2.16}$$

where $K(\cdot)$ is a univariate kernel function and $K_q(\cdot)$ is a q dimensional kernel function. Here, h_x and h_a are bandwidths that go to 0 as $n \to \infty$. Note that for simplicity of notification, we use the same bandwidth for all dimensions of X here. In practice, we can use different bandwidths for different dimensions of X to increase the efficiency of the estimation. Moreover, the marginal density of X is estimated by $\hat{f}_X(x) = (1/n)\sum_{i=1}^n K_q\{(x - X_i)/h_x\}/h_x^q$. The estimated value function can thus be written as:

$$M_n(\beta) = \int_x \frac{\frac{1}{n}\sum_{i=1}^n Y_i \frac{1}{h_x^q} K_q(\frac{x-X_i}{h_x}) \frac{1}{h_a} K\big\{\frac{g(\beta^T X)-A_i}{h_a}\big\}}{\frac{1}{n}\sum_{i=1}^n \frac{1}{h_x^q} K_q(\frac{x-X_i}{h_x}) \frac{1}{h_a} K\big\{\frac{g(\beta^T X)-A_i}{h_a}\big\}} \Big\{\frac{1}{n}\sum_{i=1}^n \frac{1}{h_x^q} K_q(\frac{x-X_i}{h_x})\Big\} dx.$$

Then β^* is estimated with $\hat{\beta}_n = \arg\max_{\beta\in\Theta} M_n\big(\beta\big)$, where Θ is a compact subset of $\mathbb{R}^q$ containing β^*.

2.4.2 Computational Details

To implement the proposed method, the R package optimr() is used for optimization of the objective function (2.16). The integral in $M_n(\beta)$ is estimated by taking the average of N_g random grid points in the covariate space. In our implementation, we chose $N_g = 3000$. In order to find the global maximizer of $M_n(\beta)$, we start optimization from q different initial points $\{(1,0,...,0)^T, (0,1,0,...,0)^T, ..., (0,...,0,1)^T\}$ and choose the one that leads to the maximum objective function value. Denote the maximizer as $\hat{\beta}_n$. When there is only one continuous covariate included, following the theoretical rate of the bandwidth parameters, the bandwidths are chosen as $h_x = C_x \text{sd}(X) n^{-1/4.5}$, $h_a = C_a \text{sd}(A) n^{-1/4.5}$, where C_x and C_a are constants arbitrarily taken between 0.5 and 3.5. When there are multiple continuous covariates, different bandwidths are used for each dimension of X.

Choosing the constants in the bandwidths is thus nontrivial. For the simulation setting 5, we arbitrarily choose the rate of the bandwidths as $n^{-1/8}$ and then use 5-fold cross validation to choose the constants C_x, C_a which minimize the mean squared error.

When the covariates consist of both continuous variables and categorical variables, the categorical variables are stratified for estimation of the value function. In other words, the kernel estimation is conducted within each subgroups classified by the categorical variables. Specifically, assume that $X = (X_1^T, X_2^T)^T \in \mathcal{X}$, where X_1 is a q_1 dimensional vector of continuous variables and $X_2 \in \mathcal{D}$ is a q_2 dimensional vector of categorical variables. The form of $M_n(\beta)$ then becomes:

$$M_n(\beta) = \sum_{x_2 \in \mathcal{D}} \int_{x_1} \frac{\frac{1}{n}\sum_{i=1}^n Y_i \tilde{K}(x, X_i) \frac{1}{h_a} K\{\frac{g(\beta^T X) - A_i}{h_a}\}}{\frac{1}{n}\sum_{i=1}^n \tilde{K}(x, X_i) \frac{1}{h_a} K\{\frac{g(\beta^T X) - A_i}{h_a}\}} \{\frac{1}{n}\sum_{i=1}^n \tilde{K}(x, X_i)\} dx_1,$$

where $x = (x_1^T, x_2^T)^T$, $X_i = (X_{i1}^T, X_{i2}^T)^T$, $\tilde{K}(x, X_i) = (1/h_x^{q_1}) K_{q_1}\{(x_1 - X_{i1})/h_x\} I(X_{i2} = x_2)$.

2.5 Theoretical Results

In this section, we establish the asymptotic properties of $\hat{\beta}_n$. To prove these results, we need to make the following assumptions. In the following expressions, $\dot{f}(x)$, $\ddot{f}(x)$ and $\dddot{f}(x)$ denote the first, second and third derivatives of the function f with respect to x; $\kappa_{0,2} = \int_{\mathbb{R}} K^2(v)dv$ and $\dot{\kappa}_{0,2} = \int_{\mathbb{R}} \dot{K}^2(v)dv$.

Assumption 2.1. *1 Assume that $\mathcal{V}$ is the set of $v \in \mathbb{R}$ where $\dot{K}(v)$, $\ddot{K}(v)$ and $\dddot{K}(v)$ exist. For $h \to 0^+$ as $n \to \infty$ and constants l,u such that $l < 0 < u$, $\int_{[l/h,u/h]\cap\mathcal{V}} K(v)dv = 1 - O(h^2)$, $\int_{[l/h,u/h]\cap\mathcal{V}} vK(v)dv = O(h)$, $\int_{[l/h,u/h]\cap\mathcal{V}} \dot{K}(v)dv = O(h^3)$, $\int_{[l/h,u/h]\cap\mathcal{V}} -v\dot{K}(v)dv = 1 - O(h^2)$, $\int_{[l/h,u/h]\cap\mathcal{V}} K^2(v)dv = \kappa_{0,2} - O(h^2)$, $\int_{[l/h,u/h]\cap\mathcal{V}} \dot{K}^2(v)dv = \dot{\kappa}_{0,2} - O(h^2)$, $\int_{[l/h,u/h]\cap\mathcal{V}} \ddot{K}(v)dv = O(h^4)$, $\int_{[l/h,u/h]\cap\mathcal{V}} v\ddot{K}(v)dv = O(h^3)$, $\int_{[l/h,u/h]\cap\mathcal{V}} v^2\ddot{K}(v)dv = 2 - O(h^2)$, $\int_{[l/h,u/h]\cap\mathcal{V}} \dddot{K}(v)dv = O(h^3)$, $\int_{[l/h,u/h]\cap\mathcal{V}} v\dddot{K}(v)dv = O(h^2)$, $\int_{[l/h,u/h]\cap\mathcal{V}} v^2\dddot{K}(v)dv = O(h)$, $\int_{[l/h,u/h]\cap\mathcal{V}} v^3\dddot{K}(v)dv = O(1)$.*

Assumption 2.2. *The function $M(\beta) = V(\pi_\beta)$ has a unique maximizer β^*.*

Assumption 2.3. *The function $m(x, a)$ is uniformly bounded. The joint density function of X and A, $f_{X,A}(x, a)$, is uniformly bounded away from 0. In addition, the first, second, third and fourth order derivatives of $m(X, A)$ and $f_{X,A}(X, A)$ with respect to X and A exist and are uniformly bounded almost everywhere.*

Assumption 2.4. *The covariate X has bounded first, second and third moments.*

Assumption 2.5. *The derivatives for function* $g(\cdot)$*:* $\dot{g}(\beta^T X), \ddot{g}(\beta^T X), \dddot{g}(\beta^T X)$ *exist and are bounded almost everywhere.*

Assumption 2.1 are regularity assumptions to ensure that the kernel estimators are unbiased. It is trivial to prove that it can be satisfied by most commonly used kernel functions such as the Gaussian kernel function, polynomial kernel, splines and all sufficiently smooth bounded kernel functions. Assumption 2.2 is an identifiability condition for β^*. Assumptions 2.3–2.5 ensure the existence of the limit of the expectation of $M_n(\beta)$ and the existence of the covariance of the limiting distribution. In the following two theorems, we establish the consistency and asymptotic normality of $\hat{\beta}_n$, respectively.

Theorem 2.1. *Under assumptions 2.1–2.3, for* h_x*,* h_a *satisfying* $nh_x^{2q}h_a^2 \to \infty$ *as* $n \to \infty$*, we have* $\sup_{\beta\in\Theta} |M_n(\beta) - M(\beta)|$ *converge in probability to* 0*, where* Θ *is a compact region containing* β^**. Thus,* $\hat{\beta}_n = \arg\max_{\beta\in\Theta} M_n(\beta)$ *converge in probability to* β^**.*

Theorem 2.2. *Under assumptions 2.1–2.5, for* h_x*,* h_a *satisfying* $nh_x^{2q}h_a^2 \to \infty$ *and* $nh_x^q h_a^3 \to \infty$ *as* $n \to \infty$*, we have*

$$(nh_x^q h_a^3)^{1/2}(\hat{\beta}_n - \beta^*) \to N\Big\{0, D(\beta^*)^{-1}\Sigma_S(\beta^*)D^{-1}(\beta^*)\Big\}$$

in distribution as $n \to \infty$*, where*

$$D(\beta^*) = \int_x \Big[m_{aa}\big\{x, g(\beta^T x)\big\}\dot{g}^2(\beta^T x) + m_a\big\{x, g(\beta^T x)\big\}\ddot{g}(\beta^T x)\Big] f_X(x) x x^T dx,$$
$$\Sigma_S(\beta^*) = \int_x \dot{g}^2(\beta^T x) x x^T \kappa_{0,2}\dot{\kappa}_{0,2} f_X^2(x) \frac{m_2\big\{x, g(x^T\beta)\big\} - m^2\big\{x, g(x^T\beta)\big\}}{f_{X,A}\big\{x, g(x^T\beta)\big\}} dx,$$

and $m_a(x,a) = \partial m(x,a)/\partial a$*,* $m_{aa}(x,a) = \partial^2 m(x,a)/\partial a^2$*,* $m_2(x,a) = E(Y^2 \mid X = x, A = a)$*. Here,* $\kappa_{0,2} = \int_{\mathbb{R}^q} K_q^2(v)dv$*,* $\dot{\kappa}_{0,2} = \int_{\mathbb{R}} \dot{K}^2(v)dv$*.*

Proofs of the above theorems are based on theory for kernel density estimators (Schuster et al. 1969) and M-estimation (Kosorok 2008). Details of proofs are given in the Section 2.9. Note that the convergence rate is slower then $n^{-1/2}$ due to the kernel estimation of the value function.

2.6 Simulation Studies

In this section, we conduct simulations to show the capability of our proposed method in identifying the optimal individualized dose rule. We first simulate some simple settings with

only one covariate. X is generated randomly from the standard normal distribution. A is generated from the uniform distribution on $[0, 1]$. We first generate X and A independently to mimic a randomized dose trial where a random dose from the safe dose range is assigned to each patient. The true optimal dose rule is generated as $\pi^{opt}(X) = g(\tilde{\beta}_0 + \tilde{\beta}_1 X)$, where $g(s) = 1/\{1 + \exp(-s)\}$. Y is generated from a normal distribution with mean $m(A, X)$ and standard deviation 0.5, where $m(A, X) = \tilde{\mu}(X) - 10\{A - \pi^{opt}(X)\}^2$. We use two different baseline functions for $\tilde{\mu}(X)$ and two different sets of $(\tilde{\beta}_0, \tilde{\beta}_1)$ as shown in Table 2.1. The sample sizes are $n = 400$ and $n = 800$ and each setting is replicated 500 times.

The average bias and the standard deviation of the estimated parameters from 500 simulations are summarized in the first half of Table 2.2. The estimated parameters were close to the true parameters. The third column shows the average of the standard errors estimated with the covariance function formula derived in Theorem 2.2 (see Section 2.9). 95% confidence intervals were calculated with the estimated standard errors. The coverage probabilities are shown in the table. From the result, we can see that the bias and standard deviation of the estimated parameters decreased with larger sample sizes. The coverage probabilities of the confidence intervals were close to 95%, supporting the convergence results given in Section 2.5.

We also study the performance of our method when the training data are from observational studies, where the doses given to the patients may depend on the covariates X. The simulation settings are the same as settings 1–4 except that A is generated from the distribution beta$\big\{2\exp(\tilde{\beta}_0 + \tilde{\beta}_1 X), 2\big\}$. The results are summarized in the second half of Table 2.2. The proposed method was still capable of giving good estimates of the parameters and the coverage of the confidence intervals were close to 95%. These simulation implies that the proposed method performs well with data from both randomized trials and observational studies.

Table 2.1: Summary of simulation settings

	No baseline $\tilde{\mu}(X) = 0$	With baseline $\tilde{\mu}(X) = 1 + 0.5\cos(2\pi X)$
$\tilde{\beta}_0 = 0, \tilde{\beta}_1 = 0.5$	Setting 1	Setting 3
$\tilde{\beta}_0 = 0, \tilde{\beta}_1 = 1$	Setting 2	Setting 4

Under settings 1–4, we compare our method with linear based O-learning (LOL) and

Table 2.2: Simulation results from 500 replicates for randomized trials and observational studies.

Randomized trials									
		$\tilde{\beta}_0$				$\tilde{\beta}_1$			
	n	Bias[1]	SD[1]	SE[1]	CP	Bias[1]	SD[1]	SE[1]	CP
S1	400	2.5	46.6	47.5	95.6	-17.3	53.5	54.5	92.8
	800	2.4	33.7	33.4	95.8	-19.5	37.3	38.5	90.2
S2	400	2.1	52.2	54.4	95.6	38.9	91.0	93.7	94.6
	800	1.5	39.1	38.1	93.8	33.0	63.0	65.9	95.8
S3	400	2.7	54.1	55.7	95.2	-20.4	64.5	64.1	90.8
	800	1.6	38.8	39.3	95.0	-18.9	43.7	45.4	92.0
S4	400	2.4	61.8	63.4	95.4	39.4	103.5	111.2	96.2
	800	-1.5	44.4	44.3	94.6	33.6	75.0	77.5	95.6
Observational studies									
		$\tilde{\beta}_0$				$\tilde{\beta}_1$			
	n	Bias[1]	SD[1]	SE[1]	CP	Bias[1]	SD[1]	SE[1]	CP
S1	400	13.9	80.5	82.4	96.0	32.4	97.7	102.1	94.6
	800	8.5	47.3	47.0	94.6	-19.6	56.7	58.2	92.2
S2	400	21.9	83.3	88.1	96.4	7.6	146.4	150.4	95.2
	800	17.8	63.4	60.9	93.0	0.6	94.0	103.2	98.2
S3	400	13.0	84.2	85.4	95.6	5.4	101.6	104.6	95.0
	800	7.4	61.2	58.6	93.4	-4.6	68.4	71.8	96.2
S4	400	21.6	91.3	97.1	96.2	5.0	165.4	169.3	95.8
	800	20.3	71.2	67.6	93.2	2.0	109.0	116.8	97.0

[1] Note: These columns are in 10^{-3} scale

[2] Note: SD refers to the standard deviation of the estimated parameters from 500 replicates, SE refers to the mean of the estimated standard errors calculated by our covariance function, CP refers to the coverage probability of the 95% confidence intervals calculated using the estimated standard errors.

[3] Note: The worst case Monte Carlo standard error for proportions is 1.3%.

Table 2.3: Value estimate $V(\hat{\pi})$ from 500 simulations in settings 1-4

		Randomized trials			
	n	DQ	LOL	KOL	KAL
S1	400	-38.1(1.9)	-7.7(7.0)	-16.5(8.2)	**-2.7(2.7)**
	800	-32.7(0.9)	-3.9(3.7)	-9.3(4.3)	**-1.9(1.8)**
S2	400	-33.9(1.3)	-18.1(10.5)	-31.7(12.5)	**-3.3(3.7)**
	800	-20.0(0.8)	-15.6(7.5)	-20.4(7.4)	**-1.9(1.8)**
S3	400	-41.6(11.4)	-8.5(14.1)	-17.2(14.2)	**-3.7(12.4)**
	800	-61.2(11.5)	-4.3(12.1)	-10.0(12.7)	**-2.4(11.7)**
S4	400	-52.5(11.9)	-21.3(17.7)	-33.3(17.4)	**-4.2(12.5)**
	800	-23.3(11.9)	-17.8(14.5)	-22.4(13.9)	**-2.4(11.7)**
		Observational studies			
	n	DQ	LOL	KOL	KAL
S1	400	-29.5(1.2)	**-7.4(6.3)**	-15.6(7.7)	-8.1(8.0)
	800	-24.4(0.9)	-5.5(4.3)	-10.3(5.3)	**-3.1(3.1)**
S2	400	-16.0(7.6)	-14.1(6.6)	-21.3(9.7)	**-8.2(8.3)**
	800	-32.0(1.1)	-12.8(4.7)	-12.2(4.5)	**-4.4(4.4)**
S3	400	-29.1(11.6)	**-8.1(13.5)**	-11.7(14.6)	-9.8(15.7)
	800	-34.2(11.8)	-6.2(12.4)	-11.2(13.3)	**-3.5(12.0)**
S4	400	-83.8(13.8)	-14.7(13.0)	-20.7(15.8)	**-10.0(15.8)**
	800	-34.1(11.3)	-13.5(12.1)	-11.2(12.3)	**-5.1(12.2)**

[1] Note: DQ refers to discretized Q-learning, LOL refers to linear O-learning, KOL refers to kernel based O-learning, KAL refers to kernel assisted learning.

[2] All columns are in 10^{-3} scale. For settings 3 and 4, the numbers in the table are the value estimate $V(\hat{\pi}) - 1$ for the purpose of comparison with the first two settings.

kernel based O-learning (KOL) proposed in Chen et al. (2016) and a discretized dose rule estimated using Q-learning. For discretized Q-learning, we divide the safe dose range into 10 equally spaced intervals: $\mathcal{A} = \mathcal{A}_1 \cup ... \cup \mathcal{A}_{10}$ and create an indicator variable for each of the dose intervals $I = (I_1, I_2, ..., I_{10})$, where $I_j = I(A \in \mathcal{A}_j)$, $j = 1, \ldots, 10$. The covariates included in the regression models are (X, X^2, I, IX, IX^2). To this end, an optimal dose range is selected for each individual and the middle point of the selected interval is suggested to the patient. The results from 500 replicates are summarized in Table 2.3. Each column is the average value function of the dose rule estimated by the corresponding method. The value function is evaluated at a testing dataset. The numbers in the parentheses are the standard deviation of the estimated value functions. From Table 2.3, we see that the proposed method performed the best under most settings. In the simulation for observational studies, O-learning performed the best when the sample size was small. However, the proposed method performed comparatively well and performed better as the sample size increased. The discretized Q-learning method did not provide a good dose suggestion in this case.

We then apply our method to a slightly more complicated setting with 3 covariates. X_1, X_2, X_3 are generated independently from a uniform$(-1, 1)$ and A follows a uniform$(0, 2)$ and is independent of X. Y is generated as follows:
Setting 5:

$$\tilde{\beta} = c(1, 0.5, 0.5, 0), \quad X = (1, X_1, X_2, X_3)^T, \pi^{opt}(X) = \tilde{\beta}^T X,$$
$$m(A, X) = 8 + 4X_1 - 2X_2 - 2X_3 - 25 \times \{A - \pi^{opt}(X)\}^2, \quad Y \sim N\{m(A, X), 1\}.$$

The rate of the bandwidths are chosen to be $n^{-1/8}$.

The results of setting 5 are summarized in Table 2.4. The average bias of the estimated parameters was small and decreased as the sample size increased. However, the estimated standard error was not stable with increased number of parameters. Thus, the coverage probabilities of the 95% confidence intervals were inaccurate for some of the parameters. According to Horowitz (2001), the bootstrap can provide refinements to the estimation of standard deviations for kernel density estimation. Therefore, we estimate the standard error using the bootstrap with 100 repetitions and calculate the bootstrap confidence interval. It appears that the bootstrap confidence intervals were more stable. The coverage of the bootstrap confidence intervals were close to 95%.

Under setting 5, we compare our method with linear O-learning, kernel based O-learning and the discretized Q-learning with covariates (X, X^2, I, IX, IX^2). The results from 500 replicates are summarized in Table 2.5. Each column is the average value

function of the dose rule estimated by the corresponding method. The value function is evaluated at a testing dataset. The numbers in the parentheses are the standard deviation of the estimated value functions. From Table 2.5, we see that linear O-learning performed comparatively well with our kernel assisted learning method (KAL). The discretized Q-learning method did not provide a good dose suggestion in this case.

Table 2.4: Average $\hat{\beta}_n$ from 500 replicates for setting 5

n	Parameter	Bias*	SD *	SE *	CP	Bootstrap CP
400	β_0	0.9	18.8	17.8	92.0	90.6
	β_1	-13.0	32.8	32.9	92.6	92.4
	β_2	-15.7	32.6	33.5	91.4	90.8
	β_3	0.2	13.5	9.08	81.4	91.4
800	β_0	0.0	15.2	13.1	91.4	90.4
	β_1	-7.2	26.5	24.4	89.9	91.0
	β_2	-6.0	24.8	24.3	93.6	93.6
	β_3	-1.1	12.1	7.9	81.4	92.4

[1] Note: * columns are in 10^{-3} scale

[2] Note: SD refers to the standard deviation of the estimated parameters from 500 replicates, SE refers to the mean of the estimated standard error calculated by our covariance function, CP refers to the coverage probability of the 95% confidence intervals calculated using the estimated standard errors, Bootstrap CP refers to the coverage probability of the 95% confidence intervals calculated using the bootstrap estimated standard errors.

[3] Note: The worst case Monte Carlo standard error for proportions is 1.8%.

Table 2.5: Value estimate $V(\hat{\pi})$ from 500 simulations in setting 5

	Discretized Q	LOL	KOL	KAL
N=400	5.66(0.32)	7.85(0.14)	5.77(0.15)	7.96(0.10)
N=800	5.76(0.27)	7.92(0.11)	5.82(0.14)	7.97(0.09)

[1] Note: Discretized Q refers to discretized Q-learning, LOL refers to linear O-learning, KOL refers to kernel based O-learning, KAL refers to kernel assisted learning.

2.7 Warfarin Data Analysis

Warfarin is a widely used anticoagulant for prevention of thrombosis and thromboembolism. Although highly efficacious, dosing for warfarin is known to be challenging because of the narrow therapeutic index and the large variability among patients (Johnson et al. 2011). Overdosing of warfarin leads to bleeding and underdosing diminishes the efficacy of the medication. The international normalized ratio (INR) measures the clotting tendency of the blood. An INR between 2–3 is considered to be safe and efficacious for patients.

Typically, the warfarin dosage is decided empirically: an initial dose is given based on the population average, and adjustments are made in the subsequent weeks while the INR of the patient is being tracked. A stable dose is decided in the end to achieve an INR of 2–3 (Johnson et al. 2011). The dosing process may take weeks to months, during which the patient is constantly at risk of bleeding or under-dosing. Therefore, a quantitative method for warfarin dosing will greatly decrease the time, cost and risks for patients.

The following analysis uses the warfarin dataset collected by Consortium (2009). In the original paper, a linear regression was used to predict the stable dose using clinical results and pharmacogenetic information, including age, weight, height, gender, race, two kinds of medications (Cytochrom P450 and Amiodarone), and two genotypes (CYP2C9 genotype and VKORC1 genotype). This prediction method is based on the assumption that the stable doses received by the patients are optimal. However, later studies showed that the suggested doses by the International Warfarin Pharmacogenetic Consortium were suboptimal for elderly people, implying that the optimal dose assumption might not be valid (Chen et al. 2016).

We apply the proposed method to this dataset to estimate the optimal individualized dose rule for warfarin. Instead of using only the data of the patients with stablized INR, we include all patients who received weekly doses between 6 mg to 95 mg. The medication information is missing for half of the observations and is therefore excluded from our analysis. Observations which are missing in the other variables are removed from the dataset, resulting in a total of 3567 patients. The outcome variable is defined as $Y_i = -(\text{INR}_i - 2.5)^2$ for the i th observation. Stratification of the categorical variables is needed for the kernel density estimation. In order to ensure that there are enough observations in each stratified group, we consider only categorical variables that are distributed comparatively even among different groups. In our analysis, we use three variables: height, gender and the indicator variable for VKORC1 of type AG. Before we apply the proposed the method, we normalize all the variables by $X_{i,j} = (X_{i,j} - \bar{X}_j)/\text{sd}(X_j)$, where $j = 1, 2, 3$, $i = 1, \ldots, n$. $\bar{X}_j = \sum_{i=1}^{n} X_{i,j}/n$ and $\text{sd}(X_j)$ is the standard deviation of

the j-th variable.

The estimation results are shown in Table 2.6. The p-value is obtained for each of the parameters. The result implied that the optimal dose for male is higher than the optimal dose for female given all the other variables are the same. It was also implied that the patients with genotype VKORC1=AG need higher doses than the patients with VKORC1$\neq$ AG. We use the same variables and compare our method with O-learning and the discretized Q-learning method. For the discretized Q-learning method, we also divide the dose range into 10 equally spaced intervals. The suggested doses by the three methods are shown in Fig. 2.2. The result shows a tendency of the discretized Q-learning to suggest extreme doses, which is not ideal in real application. This might be due to the fact that the higher dose intervals contain small numbers of observations, and thus the estimated models are dominated by a few subjects.

To evaluate the estimated dose rules of these methods, we randomly take two thirds of the data as training data and the rest of the data as testing data. The optimal individualized dose rule is estimated with the training data. The value function of the suggested individualized dose rule $V(\hat{\pi})$ is estimated with the average of the Nadaraya Watson estimator for $E\{Y \mid X, A = \hat{\pi}(X)\}$ in the testing dataset. The tuning parameters for the Nadaray-Watson estimators are taken as $h_x = 1.25\text{sd}(X)n_{\text{test}}^{-1/4.5}$ and $h_a = 1.75\text{sd}(A)n_{\text{test}}^{-1/4.5}$, where $n_{\text{test}} = 1189$ is the size of the testing dataset. The process is repeated 200 times. The distribution of the estimated value of the suggested dose is shown in Fig. 2.3. The suggested individualized dose rule with our proposed method lead to better expected outcomes in the population compared to the other methods. The performance of the discretized Q-learning method was not stable as shown in the result. However, this result was only based on the three variables selected, while in reality, the two medications (Cytochrome P450 enzyme and Amiodarone) and the genotype CYP2C9 are also of significant importance in warfarin dosing. The computation complexity of our proposed method restricted its capability of handling higher dimensional problems.

Table 2.6: Estimated $\hat{\beta}$ with warfarin data with kernel assisted learning

Variable	Estimated Parameter	SE	p-value
Intercept	-0.463	0.064	0.000
Height	-0.263	0.101	0.005
Gender	0.268	0.134	0.023
VKORC1.AG	-0.4682	0.094	0.000

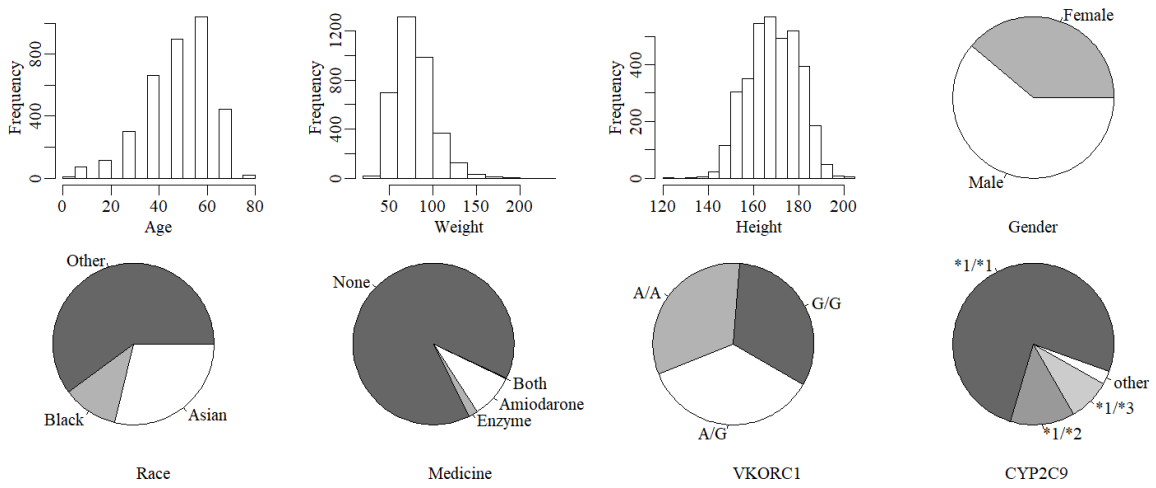

Figure 2.1: Distribution of the variables in the warfarin dataset.

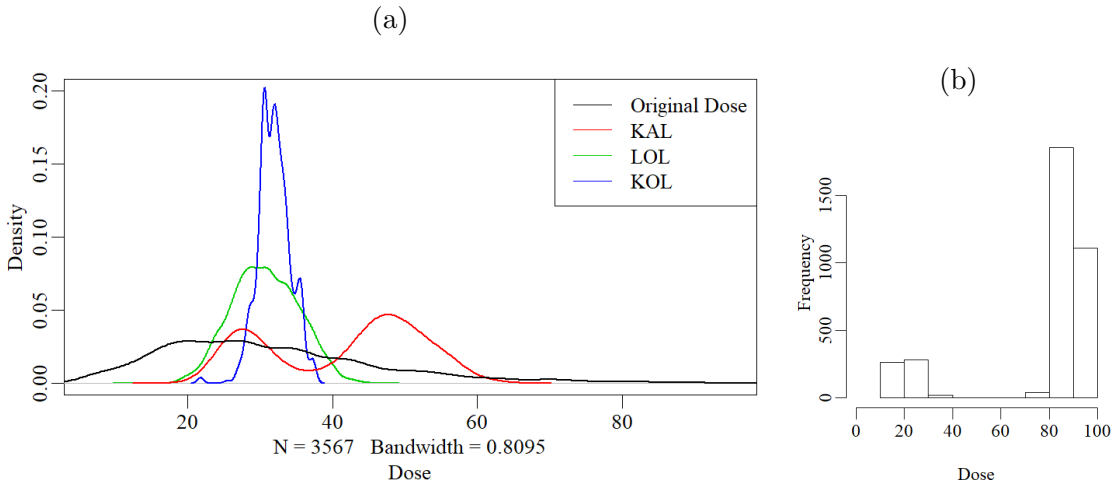

Figure 2.2: Empirical distribution of suggested doses of several methods for the walfarin dataset. In panel (a), the black line is the distribution of the original doses from the dataset. The green line denotes the result from linear O-learning. The blue line denotes the result from kernel based O-learning. The red line denotes the result from kernel assisted learning. Panel (b) is the histogram of the suggested doses using discretized Q-learning.

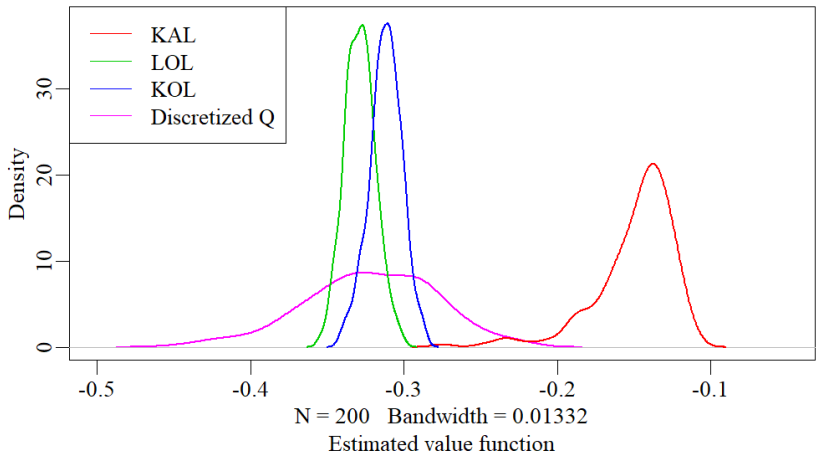

Figure 2.3: Empirical distribution of the estimated value function over 200 repetitions for the warfarin dataset. The green line denotes the result from linear O-learning. The blue line denotes the result from kernel based O-learning. The red line denotes the result from kernel assisted learning. The pink line denotes the result from discretized Q-learning.

2.8 Discussion and Conclusion

The proposed kernel assisted learning method for estimating the optimal individualized dose rule provides the possibility of conducting statistical inference with estimated dose rules, thus providing insights on the importance of the covariates in the dosing process. In our simulation settings, the proposed method was capable of identifying the optimal individualized dose rule when the optimal dose rule was inside the prespecified class of rules. In the warfarin dosing case, based on the three covariates selected, the suggested dose appeared to lead to better expected clinical result compared to the other methods.

The proposed method has several possible extensions. Notice that the form of the prespecified rule class can be extended to a link function with a nonlinear predictor $g\{\Psi(X)^T\beta\}$ where $\Psi(\cdot) = \{\Psi_1(\cdot), ..., \Psi_c(\cdot)\}^T$ are some prespecified basic spline functions and $\beta \in \mathbb{R}^c$. The accuracy of the approximated value function might also be improved by extending the multivariate kernel $K_q(x/h_x)/h_x$ to $|H|^{-1/2}K_q(H^{-1/2}x)$ (Duong and Hazelton 2005).

One weakness of the proposed method is that the accuracy of the estimated value function is sensitive to the choice of bandwidth. The kernel density estimator in the denominator of $M_n(\beta)$ might lead to large bias when the bandwidths are not properly chosen. As the dimension of X increases, the choice of the bandwidths is nontrivial. In this article, we used cross validation to choose C_x and C_a for setting 5 by minimizing

the mean squared error of the Nadaraya-Watson estimator. However, due to the complex form of $M_n(\beta)$, this method might not be optimal when the dimension of the covariates further increases. The criteria for choosing bandwidths needs to be studied further.

In the future, we are interested in variable selection when dealing with high dimensional data. Extensions to multi-stage dose finding problems is also of interest. Personalized Dose Finding is still a relatively new problem. With the complicated mechanisms of various diseases, there are many more problems to be tackled in this realm.

2.9 Proof and Technical Details

2.9.1 Proof of Theorem 2.1

We first prove the uniform convergence of $M_n(\beta)$ to $M(\beta)$. For simplicity of notation, let's define:

$m_x(x,a) = \partial m(x,a)/\partial x$, $m_{2x}(x,a) = \partial m_2(x,a)/\partial x$, $m_{2a}(x,a) = \partial m_2(x,a)/\partial a$. Similarly, $f_a(x,a) = \partial f_{X,A}(x,a)/\partial a$, $f_x(x,a) = \partial f_{X,A}(x,a)/\partial x$. We write $M_n(\beta)$ as

$$M_n(\beta) = \int_x \frac{A_n(x;\beta)}{B_n(x;\beta)} C_n(x)dx,$$

where

$$A_n(x;\beta) = \frac{1}{n}\sum_{i=1}^{n} Y_i \frac{1}{h_x^q} K_q(\frac{x-X_i}{h_x}) \frac{1}{h_a} K\big\{\frac{g(\beta^T x) - A_i}{h_a}\big\},$$
$$B_n(x;\beta) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{h_x^q} K_q(\frac{x-X_i}{h_x}) \frac{1}{h_a} K\big\{\frac{g(\beta^T x) - A_i}{h_a}\big\},$$
$$C_n(x) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{h_x^q} K_q(\frac{x-X_i}{h_x}).$$

Notice that $M(\beta)$ can be written as

$$M(\beta) = \int_x \frac{A(x;\beta)}{B(x;\beta)} C(\beta)dx,$$

where $A(x;\beta) = m\{x, g(\beta^T x)\} f_{X,A}\{x, g(\beta^T x)\}$, $B(x;\beta) = f_{X,A}\{x, g(\beta^T x)\}$ and $C(x) = f_X(x)$. Thus,

$$\sup_\beta \Big|M_n(\beta) - M(\beta)\Big| = \sup_\beta \Big| \int_x \Big\{ \frac{A_n(x;\beta)}{B_n(x;\beta)} C_n(x) - \frac{A(x;\beta)}{B(x;\beta)} C(x) \Big\} dx \Big|$$
$$\le \sup_\beta \Big| \int_x \Big\{ \frac{A_n(x;\beta)}{B_n(x;\beta)} - \frac{A(x;\beta)}{B(x;\beta)} \Big\} C_n(x)dx \Big| + \sup_\beta \Big| \int_x \frac{A(x;\beta)}{B(x;\beta)} \Big\{ C_n(x) - C(x) \Big\} dx \Big|$$
$$\le \sup_{a,x} \Big| \frac{A_n^*(x,a)}{B_n^*(x,a)} - \frac{A^*(x,a)}{B^*(x,a)} \Big| \Big\{ \int_x C_n(x)dx \Big\} + \sup_{a,x} |m(x,a)| \Big\{ \int_x \Big| C_n(x) - C(x) \Big| dx \Big\}$$

where

$$A_n^*(x,a) = \frac{1}{n}\sum_{i=1}^{n} Y_i \frac{1}{h_x^q} K_q(\frac{x-X_i}{h_x}) \frac{1}{h_a} K(\frac{a-A_i}{h_a}),$$
$$B_n^*(x,a) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{h_x^q} K_q(\frac{x-X_i}{h_x}) \frac{1}{h_a} K(\frac{a-A_i}{h_a}),$$

and $A^*(x,a) = m(x,a) f_{X,A}(x,a)$, $B^*(x,a) = f_{X,A}(x,a)$. It is trivial to prove that $\int_x C_n(x)dx = \int_x C(x)dx = 1$. Thus, under the boundedness of $m(x,a)$, we only need to show that:

$$\int_x |C_n(x) - C(x)|dx = o_p(1)$$
$$\sup_{a,x} \Big| \frac{A_n^*(x,a)}{B_n^*(x,a)} - \frac{A^*(x,a)}{B^*(x,a)} \Big| \to 0 = o_p(1).$$

To prove the first equation, notice that $\int_x |C_n(x) - C(x)|dx \le \int_x C_n(x)dx + \int_x C(x)dx = 2$. By the dominated convergence theorem, it suffices to show the uniform convergence of the kernel density estimate $C_n(x)$ to $C(x)$, which can be proved according to Schuster (1969).

For the second equation,

$$\begin{aligned}
&\sup_{a,x} \Big| \frac{A_n^*(x,a)}{B_n^*(x,a)} - \frac{A^*(x,a)}{B^*(x,a)} \Big| \\
&= \sup_{a,x} \Big| \frac{\big\{A_n^*(x,a) - A^*(x,a)\big\} B^*(x,a) - A^*(x,a)\big\{B_n^*(x,a) - B^*(x,a)\big\}}{B_n^*(x,a) B^*(x,a)} \Big| \\
&\le \sup_{x,a} \Big| \frac{A_n^*(x,a) - A^*(x,a)}{B_n^*(x,a)} \Big| + \sup_{x,a} \Big| \frac{\big\{B_n^*(x,a) - B^*(x,a)\big\} A^*(x,a)}{B^*(x,a) B_n^*(x,a)} \Big|.
\end{aligned}$$

Under the boundedness of $A^*(x,a)$ and the assumption that $f_{X,A}(x,a)$ is uniformly bounded away from 0, it suffices to show that $\sup_{a,x} \big|A_n^*(x,a) - A^*(x,a)\big| = o_p(1)$, and $\sup_{a,x} \big|B_n^*(x,a) - B^*(x,a)\big| = o_p(1)$.

To prove the uniform convergence of $A_n^*(x,a)$, notice that:

$$\sup_{a,x} \Big| A_n^*(x,a) - A^*(x,a) \Big| \le \sup_{a,x} \Big| A_n^*(x,a) - E\big\{A_n^*(x,a)\big\} \Big| + \sup_{a,x} \Big| E\big\{A_n^*(x,a)\big\} - A^*(x,a) \Big|. \tag{2.17}$$

We prove the convergence of the two parts on the right side separately. First we obtain,

$$
\begin{aligned}
&E\big\{A_n^*(x,a)\big\} = \\
&\int_{x_i,a_i}\int_{y_i}\big\{y_i\frac{1}{h_x^q}K_q(\frac{x-x_i}{h_x})\frac{1}{h_a}K(\frac{a-a_i}{h_a})\big\}f_{Y|X,A}(y_i|x_i,a_i)f_{X,A}(x_i,a_i)dy_idx_ida_i \\
&=\int_{x_i,a_i}\big\{m(x_i,a_i)\frac{1}{h_x^q}K_q(\frac{x-x_i}{h_x})\frac{1}{h_a}K(\frac{a-a_i}{h_a})\big\}f_{X,A}(x_i,a_i)dx_ida_i \\
&=\int_u\int_{v=(a-1)/h_a}^{a/h_a} m(x-uh_x,a-vh_a)K_q(u)K(v)f_{X,A}(x-uh_x,a-vh_a)dvdu \\
&=\int_u\int_{v=(a-1)/h_a}^{a/h_a}\big\{m\big(x,a\big)-uh_xm_x\big(x,a\big)-vh_am_a\big(x,a\big)+O(h_x^2)+O(h_xh_a)+O(h_a^2)\big\} \\
&K(u)K(v)\big\{f_{X,A}\big(x,a\big)-uh_xf_x\big(x,a\big)-vh_af_a\big(x,a\big)+O(h_x^2)+O(h_xh_a)+O(h_a^2)\big\}dvdu \\
&=m\big(x,a\big)f_{X,A}\big(x,a\big)+O(h_x^2)+O(h_a^2) \\
&=A^*(x,a)+O(h_x^2)+O(h_a^2),
\end{aligned}
$$

where the third equality is achieved by letting $u=(x-X)/h_x$ and $v=\{g(\beta^Tx)-A\}/h_a$. The fourth equation is from Taylor expansion and the fifth equation is based on Assumption (A1) that $\int_u uK_q(u)du=0$, $\int_{\{g(\beta'x)-1\}/h_a}^{g(\beta'x)/h_a}K(v)dv=1-O(h_a^2)$ and $\int_{\{g(\beta'x)-1\}/h_a}^{g(\beta'x)/h_a}vK(v)dv=O(h_a)$. By the assumption of uniform boundedness of the second order derivatives of $m(x,a)$ and $f_{X,A}(x,a)$, we have $\sup_{x,a}|E\big\{A_n^*(x,a)\big\}-A^*(x,a)|\to 0$.

Then we prove the convergence of the first part of Equation (2.17):

$$
\begin{aligned}
&\sup_{x,a}\Big|A_n^*(x,a)-E\big\{A_n^*(x,a)\big\}\Big| \\
&=\frac{1}{h_x^qh_a}\sup_{x,a}\Big|\int_{X,A}m(X,A)K_q(\frac{x-X}{h_x})K(\frac{a-A}{h_a})d\big\{F_n(X,A)-F(X,A)\big\}\Big| \\
&=\frac{1}{h_x^qh_a}\sup_{x,a}\Bigg|\Big[\big\{F_n(x,a)-F(x,a)\big\}m(X,A)K_q(\frac{x-X}{h_x})K(\frac{a-A}{h_a})\Big]_{-\infty}^{\infty} \\
&-\int_{X,A}\big\{F_n(x,a)-F(x,a)\big\}d\big\{m(X,A)K_q(\frac{x-X}{h_x})K(\frac{a-A}{h_a})\big\}\Bigg| \\
&\le\frac{C_3}{h_x^qh_a}\sup_{x,a}\Big|F_n(x,a)-F(x,a)\Big|,
\end{aligned}
$$

where C_3 is a constant, $F(X,A)$ is the cumulative joint distribution of X and A and $F_n(x,a)=\{\sum_{i=1}^n I(X_i\le x,A_i\le a)\}/n$. Here $X_i\le x$ means that each term of X_i is less than or equal to the corresponding term of x. The last inequality can be obtained by the boundedness of $m(x,a)$, $K_q(u)$ and $K(v)$. By Lemma 2.1 of Schuster (1969)

we know that there exists a universal constant C_4 such that for any $n > 0$, $\epsilon_n > 0$, $P_F\Big\{ \sup_{x,a} \big|F_n(x,a) - F(x,a)\big| > \epsilon \Big\} \leq C_4 \exp(-2n\epsilon^2)$. For n sufficiently large:

$$P\Big\{ \sup_{x,a} \big|A_n^*(x,a) - A^*(x,a)\big| > \epsilon \Big\} \leq P\Big\{ \sup_{x,a} \Big|A_n^*(x,a) - E\big\{A_n^*(x,a)\big\}\Big| > \frac{\epsilon}{2} \Big\}$$
$$\leq P\Big\{ \sup_{x,a} \big|F_n(x,a) - F(x,a)\big| > \frac{h_x^q h_a \epsilon}{2C_3} \Big\} \leq C_4 \exp\big(-2n \frac{h_x^{2q} h_a^2 \epsilon^2}{C_3^2} \big).$$

If $nh_x^{2q}h_a^2 \to \infty$ then $P\Big\{ \sup_{x,a} \big|A_n^*(x,a) - A^*(x,a)\big| \Big\} \to 0$. Thus the uniform convergence of $A_n^*(x,a)$ is proved. The uniform convergence of $B_n^*(x,a)$ can be proved similarly. Thus, we can obtain $\sup_{\beta\in\Theta} \big|M_n(x;\beta) - M(x;\beta)\big| \xrightarrow{p} 0$. By Theorem 2.10 of Kosorok (2008), we now obtain that $\hat{\beta}_n \xrightarrow{p} \beta^*$.

2.9.2 Proof of Theorem 2.2

Since $\hat{\beta}_n$ and β^* are maximizers of $M_n(\beta)$ and $M(\beta)$, they are solutions of $S_n(\beta) = 0$ and $S(\beta) = 0$, where, $S(\beta) = \partial M(\beta)/\partial\beta$ and $S_n(\beta) = \partial M_n(\beta)/\partial\beta$. By Taylor expansion, we have

$$0 = S_n(\hat{\beta}_n) = S_n(\beta^*) + D_n(\beta^*)(\hat{\beta}_n - \beta^*) + \frac{1}{2}(\hat{\beta}_n - \beta^*)^T \frac{\partial^2}{\partial\beta\partial\beta^T} S_n(\tilde{\beta})(\hat{\beta}_n - \beta^*)$$
$$= S_n(\beta^*) + \Big\{ D_n(\beta^*) + \frac{1}{2}(\hat{\beta}_n - \beta^*)^T \frac{\partial^2}{\partial\beta\partial\beta^T} S_n(\tilde{\beta}) \Big\}(\hat{\beta}_n - \beta^*),$$

where $\tilde{\beta}$ is on the line segment connecting $\hat{\beta}_n$ and β^*, $D_n(\beta) = \partial^2 M_n(\beta)/(\partial\beta\partial\beta^T)$, $D(\beta) = \partial^2 M(\beta)/(\partial\beta\partial\beta^T)$. To prove the weak convergence of $\hat{\beta}_n$, we can first prove that:

$$(nh_x^q h_a^3)^{1/2}\big\{ S_n(\beta^*) - S(\beta^*) \big\} \to N\big\{0, \Sigma_S(\beta^*)\big\}, \tag{2.18}$$

in distribution as $n \to \infty$,

$$D_n(\beta^*) - D(\beta^*) = o_p(1), \tag{2.19}$$

$$\frac{1}{2}(\hat{\beta}_n - \beta^*)^T \frac{\partial^2}{\partial\beta\partial\beta^T} S_n(\tilde{\beta}) = o_p(1). \tag{2.20}$$

Then we obtain:

$$(nh_x^q h_a^3)^{1/2}(\hat{\beta}_n - \beta^*) = -\Big\{ D_n(\beta^*) + \frac{1}{2}(\hat{\beta}_n - \beta^*)^T \frac{\partial^2}{\partial\beta\partial\beta^T} S_n(\beta^*) \Big\}(nh_x^q h_a^3)^{1/2} S_n(\beta^*)$$
$$\to N\Big\{0, D(\beta^*)^{-1}\Sigma_S(\beta^*)D(\beta^*)^{-1}\Big\}$$

in distribution as $n \to \infty$.

Proof of Equation (2.18)

To prove Equation 2.18, we first write $S_n(\beta)$ as:

$$S_n(\beta) = \frac{\partial M_n(\beta)}{\partial \beta} = \int_x \frac{\tilde{A}_n(x;\beta)B_n(x;\beta) - A_n(x;\beta)\tilde{B}_n(x;\beta)}{B_n(x;\beta)^2} C_n(x)dx,$$

where

$$\tilde{A}_n(x;\beta) = \frac{\partial}{\partial \beta} A_n(x;\beta) = \frac{1}{n}\sum_{i=1}^{n} Y_i \frac{1}{h_x^q} K_q(\frac{x - X_i}{h_x}) \frac{1}{h_a^2} \dot{K}\big\{\frac{g(\beta^T x) - A_i}{h_a}\big\} \dot{g}(\beta^T x)x,$$

$$\tilde{B}_n(x;\beta) = \frac{\partial}{\partial \beta} B_n(x;\beta) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{h_x^q} K_q(\frac{x - X_i}{h_x}) \frac{1}{h_a^2} \dot{K}\big\{\frac{g(\beta^T x) - A_i}{h_a}\big\} \dot{g}(\beta^T x)x.$$

Since $S_n(\beta)$ is of the integration form, to calculate the limiting distribution of S_n, we can first calculate the limiting distribution of the part inside the integral for a fixed x.

Let the parts inside the integral of $S_n(\beta)$ and $S(\beta)$ be:

$$G_n(x;\beta) = \frac{\tilde{A}_n(x;\beta)B_n(x;\beta) - A_n(x;\beta)\tilde{B}_n(x;\beta)}{B_n(x;\beta)^2} C_n(x),$$

$$G(x;\beta) = \frac{\tilde{A}(x;\beta)B(x;\beta) - A(x;\beta)\tilde{B}(x;\beta)}{B^2(x;\beta)} C(x),$$

where $\tilde{A}(x;\beta) = \Big[m\big\{x, g(\beta^T x)\big\} f_a\big\{x, g(\beta^T x)\big\} + m_a\big\{x, g(\beta^T x)\big\} f_{X,A}\big\{x, g(\beta^T x)\big\}\Big] \dot{g}(\beta^T x)x$ and $\tilde{B}(x;\beta) = f_a\{x, g\big(\beta^T x\big)\big\}$.

To prove the limit distribution of $G_n(\beta) - G(\beta)$, we need the following lemma:

Lemma 2.3. *If $\{A_n\}_{n=1}^{\infty}$ and $\{B_n\}_{n=1}^{\infty}$ are two sequences of random variables and $c_n(A_n - A) \to N(0, \Sigma_A)$ in distribution and $d_n(B_n - B) \to N(0, \Sigma_B)$ in distribution, where $c_n/d_n \to 0$ as $n \to \infty$. Then:*

$$c_n(A_n B_n - AB) = c_n(A_n - A)B + o_p(1).$$

Proof. Notice that: $c_n(A_n B_n - AB) = (c_n/d_n)A_n\{d_n(B_n - B)\} + c_n(A_n - A)B$, where $d_n(B_n - B)$ converges to a normal distribution, A_n converge in probability to A and $c_n/d_n \to 0$. Thus the first term is $o_p(1)$. Then we have $c_n(A_n B_n - AB) = c_n(A_n - A)B + o_p(1)$. □

Under the assumption of the boundedness of the first three derivatives of $m(x,a)$ and $f(x,a)$, we can prove that $E\big\{\tilde{A}_n(x;\beta)\big\} = \tilde{A}(x;\beta) + O(h_x^2 + h_a^2)$. Together with the law of large numbers, we obtain that $\tilde{A}_n(x;\beta) \xrightarrow{p} \tilde{A}(x;\beta)$. Since $\tilde{A}_n(x;\beta)$ is the sum of n i.i.d variables, with the central limit theorem we can obtain that $(nh_x^q h_a^3)^{1/2}\{\tilde{A}_n(x;\beta) - \tilde{A}(x;\beta)\}$ converges to a normal distribution if $nh_x^q h_a^3 \mathrm{Var}\{\tilde{A}_n(x;\beta)\}$ converges to a constant covariance matrix. Notice now that,

$$\begin{aligned}
&\mathrm{Var}\big\{\tilde{A}_n(x;\beta)\big\} = \frac{1}{n}\mathrm{Var}\Big[Y_i \frac{1}{h_x^q} K_q\big(\frac{x - X_i}{h_x}\big)\frac{1}{h_a^2}\dot{K}\big\{\frac{g(\beta^T x) - A_i}{h_a}\big\}\dot{g}(\beta^T x)x\Big] \\
&= \frac{1}{n}\dot{g}^2(\beta^T x)xx^T\left\{E\Big[Y_i^2 \frac{1}{h_x^{2q}} {K_q}^2\big(\frac{x - X_i}{h_x}\big)\frac{1}{h_a^4}\dot{K}^2\big\{\frac{g(\beta^T x) - A_i}{h_a}\big\}\Big] - E^2\big[\tilde{A}_n(x;\beta)\big]\right\} \\
&= \frac{1}{n}\dot{g}^2(\beta^T x)xx^T E\Big[Y_i^2 \frac{1}{h_x^{2q}} {K_q}^2\big(\frac{x - X_i}{h_x}\big)\frac{1}{h_a^4}\dot{K}^2\big\{\frac{g(\beta^T x) - A_i}{h_a}\big\}\Big] + O(\frac{1}{n}),
\end{aligned}$$

where the expectation in the last equation can be calculated similarly as before:

$$\begin{aligned}
&E\Big[Y_i^2 \frac{1}{h_x^{2q}} {K_q}^2\big(\frac{x - X_i}{h_x}\big)\frac{1}{h_a^4}\dot{K}^2\big\{\frac{g(\beta^T x) - A_i}{h_a}\big\}\Big] \\
&= \frac{1}{h_x^q h_a^3}\Big[m_2\big\{x, g(\beta^T x)\big\} f_{X,A}\big\{x, g(\beta^T x)\big\}\kappa_{0,2}\dot{\kappa}_{0,2} + O(h_x^2) + O(h_a^2) + O(h_x h_a)\Big].
\end{aligned}$$

Thus,

$$nh_x^q h_a^3 \mathrm{Var}\big\{\tilde{A}_n(x;\beta)\big\} = \dot{g}^2(\beta^T x)xx^T m_2\big\{x, g(\beta^T x)\big\} f_{X,A}\big\{x, g(\beta^T x)\big\}\kappa_{0,2}\dot{\kappa}_{0,2} + O(h_x^q h_a^3).$$

Therefore, for $nh_x^q h_a^3 \to \infty$ as $n \to \infty$, we have:

$$\begin{gathered}
(nh_x^q h_a^3)^{1/2}\big\{\tilde{A}_n(x;\beta) - \tilde{A}(x;\beta)\big\} \to \\
N\Big[0, g^2(\beta^T x)xx^T m_2\big\{x, g(\beta^T x)\big\} f_{X,A}\big\{x, g(\beta^T x)\big\}\kappa_{0,2}\dot{\kappa}_{0,2}\Big]
\end{gathered}$$

in distribution as $n \to \infty$.

Similarly, we can obtain that, as $n \to \infty$, $(nh_x^q h_a^3)^{1/2}\big\{\tilde{B}_n(x;\beta) - \tilde{B}(x;\beta)\big\}$ converges in distribution to $N\Big[0, \dot{g}^2(\beta^T x)xx^T f_{X,A}\big\{x, g(\beta^T x)\big\}\kappa_{0,2}\dot{\kappa}_{0,2}\Big]$, $(nh_x^q h_a^3)^{1/2}\big\{A_n(x;\beta) - A(x;\beta)\big\}$ converges in distribution to $N\Big[0, m_2\big\{x, g(\beta^T x)\big\} f_{X,A}\big\{x, g(\beta^T x)\big\}\kappa_{0,2}\tilde{\kappa}_{0,2}\Big]$, where $\tilde{\kappa}_{0,2} = \int K^2(s)ds$. $(nh_x^q h_a^3)^{1/2}\big\{B_n(x;\beta) - B(x;\beta)\big\}$ converges in distribution to $N\Big[0, f_{X,A}\big\{x, g(\beta^T x)\big\}\kappa_{0,2}\tilde{\kappa}_{0,2}\Big]$.

By Lemma 2.3 and the above convergence results, we obtain that:

$$\begin{aligned}
&(nh_x^q h_a^3)^{1/2}\big\{\tilde{A}_n(x;\beta)B_n(x;\beta) - \tilde{B}_n(x;\beta)A_n(x;\beta)\big\} \\
&= (nh_x^q h_a^3)^{1/2}\big\{\tilde{A}_n(x;\beta)B(x;\beta) - \tilde{B}_n(x;\beta)A(x;\beta)\big\} + o_p(1) \\
&= (nh_x^q h_a^3)^{1/2}\big\{\frac{1}{n}\sum_{i=1}^{n}\Phi_i(x;\beta)\big\} + o_p(1),
\end{aligned}$$

where

$$\Phi_i(x;\beta) = \big\{Y_i B(x;\beta) - A(x;\beta)\big\}\frac{1}{h_x^q}K_q(\frac{x-X_i}{h_x})\frac{1}{h_a^2}\dot{K}\big\{\frac{g(\beta^T x) - A_i}{h_a}\big\}\dot{g}(\beta^T x)x.$$

Similar to previous calculations, by the central limit theorem we can prove that $(nh_x^q h_a^3)^{1/2}$ $\{\sum_{i=1}^n \Phi_i(x;\beta)/n\}$ converge to a normal distribution, where the covariance of the asymptotic distribution is $\Sigma_\Phi(x;\beta)$:

$$\Sigma_\Phi(x;\beta) = \dot{g}^2(\beta^T x)xx^T\kappa_{0,2}\dot{\kappa}_{0,2}\Big[m_2\big\{x, g(x'\beta)\big\} - m^2\big\{x, g(x'\beta)\big\}\Big]f_{X,A}^3\big\{x, g(x'\beta)\big\}.$$

Notice that

$$(nh_x^q h_a^3)^{1/2}G_n(x;\beta) = (nh_x^q h_a^3)^{1/2}\Big\{\frac{1}{n}\sum_{i=1}^{n}\Phi_i(x;\beta)\Big\}\frac{C_n(x)}{B_n^2(x;\beta)} + o_p(1).$$

Together with $C_n(x) \xrightarrow{p} C(x)$, $B_n(x;\beta) \xrightarrow{p} B(x;\beta)$, and Slusky's theorem, we now obtain that:

$$(nh_x^q h_a^3)^{1/2}\{G_n(x;\beta) - G(x;\beta)\} \to N\big\{0, \Sigma_G(x;\beta)\big\},$$

where:

$$\Sigma_G(x;\beta) = \dot{g}^2(\beta^T x)xx^T\kappa_{0,2}\dot{\kappa}_{0,2}f_X^2(x)\frac{m_2\big\{x, g(x'\beta)\big\} - m^2\big\{x, g(x'\beta)\big\}}{f_{X,A}\big\{x, g(x'\beta)\big\}}.$$

Now let us calculate the covariance of $S_n(x;\beta)$. By the tightness of $G_n(x;\beta)$ and $G(x;\beta)$, $(nh_x^q h_a^3)^{1/2}\{G_n(x;\beta) - G(x;\beta)\}$ converges weakly to a Gaussian process $\mathcal{G}(x)$with mean 0

and covariance function $\Sigma_{\mathcal{G}}(\beta)$, where $\Sigma_{\mathcal{G}}(x_1, x_2; \beta)$ is the limit of

$$
\begin{aligned}
&\mathrm{Cov}\big\{(nh_x^q h_a^3)^{1/2} G_n(x_1;\beta), (nh_x^q h_a^3)^{1/2} G_n(x_2;\beta)\big\} \\
&= \frac{h_x^q h_a^3}{n} \sum_{i=1}^{n}\sum_{j=1}^{n} \mathrm{Cov}\big\{\Phi_i(x_1;\beta), \Phi_j(x_2;\beta)\big\} \frac{C(x_1)C(x_2)}{B^2(x_1;\beta)B^2(x_2;\beta)} + o_p(1) \\
&= \frac{h_x^q h_a^3}{n} \sum_{i=1}^{n} \mathrm{Cov}\big\{\Phi_i(x_1;\beta), \Phi_i(x_2;\beta)\big\} \frac{C(x_1)C(x_2)}{B^2(x_1;\beta)B^2(x_2;\beta)} + o_p(1) \\
&= h_x^q h_a^3 \mathrm{Cov}\big\{\Phi_1(x_1;\beta), \Phi_1(x_2;\beta)\big\} \frac{C(x_1)C(x_2)}{B^2(x_1;\beta)B^2(x_2;\beta)} + o_p(1) \\
&= T(x_1, x_2;\beta)\Big\{ \int_u K_q(u) K_q(u + \frac{x_2 - x_1}{h_x}) du \Big\} \Big[\int_v \dot{K}(v) \dot{K}\big\{v + \frac{g(\beta^T x_1) - g(\beta^T x_2)}{h_a}\big\} dv \Big] \\
&+ o_p(1),
\end{aligned}
$$

for $x_1, x_2 \in \mathcal{X}$, where

$$
\begin{aligned}
&T(x_1, x_2;\beta) = \\
&\Big[m_2\big\{x_1, g(\beta^T x_1)\big\} B(x_1;\beta) B(x_2;\beta) - m\big\{x_1, g(\beta^T x_1)\big\}\big\{A(x_2;\beta)B(x_1;\beta) + \\
&A(x_1;\beta)B(x_2;\beta)\big\} + A(x_1;\beta)A(x_2;\beta)\Big] \dot{g}(\beta^T x_1)\dot{g}(\beta^T x_2) x_1 x_2^T \frac{C(x_1)C(x_2)}{B^2(x_1;\beta)B^2(x_2;\beta)}.
\end{aligned}
$$

When $x_1 \neq x_2$, $\int_u K_q(u) K\{u + (x_2 - x_1)/h_x\} du$ and $\int_v \dot{K}(v)\dot{K}[v + \{g(\beta^T x_1) - g(\beta^T x_2)\}/h_a] dv$ will converge to 0 as $h_x, h_a \to 0$. Thus $\Sigma_{\mathcal{G}}(x_1, x_2;\beta) = 0$ for $x_1 \neq x_2$. Therefore,

$$
(nh_x^q h_a^3)^{1/2}\big\{S_n(\beta^*) - S(\beta^*)\big\} = \int_x (nh_x^q h_a^3)^{1/2}\big\{G_n(x;\beta^*) - G(x;\beta^*)\big\} dx \to N\big\{0, \Sigma_S(\beta^*)\big\},
$$

in distribution, where $\Sigma_S(\beta) = \int_{x_1}\int_{x_2} \Sigma_{\mathcal{G}}(x_1, x_2;\beta) dx_1 dx_2 = \int_x \Sigma_G(x;\beta) dx$.

Proof of Equation (2.19)

First, write $D_n(\beta)$ as:

$$
\begin{aligned}
D_n(\beta) &= \frac{\partial^2}{\partial\beta\partial\beta^T} M_n(\beta) \\
&= \int_x \Big\{ \frac{\frac{\partial^2}{\partial\beta\partial\beta^T} A_n(x;\beta)}{B_n(x;\beta)} - 2\frac{\frac{\partial}{\partial\beta} A_n(x;\beta) \frac{\partial}{\partial\beta} B_n(x;\beta)}{B_n^2(x;\beta)} - \frac{A_n(x;\beta)\frac{\partial^2}{\partial\beta\partial\beta^T} B_n(x;\beta)}{B_n(x;\beta)^2} \\
&+ 2\frac{A_n(x;\beta)\frac{\partial}{\partial\beta} B_n(x;\beta)\frac{\partial}{\partial\beta^T} B_n(x;\beta)}{B_n(x;\beta)^3} \Big\} C_n(x) dx.
\end{aligned}
$$

Similar to previous calculations, under the assumption of boundedness of the first three order derivatives of $m(x,a)$ and $f_{X,A}(x,a)$, we obtain that $\partial^2 A_n(x;\beta)/(\partial\beta\partial\beta^T)$ converges in probability to:

$$
\begin{aligned}
&2\Big[m_{aa}\big\{x,g(\beta^T x)\big\}f_{X,A}\big(x,g(\beta^T x)\big)+m_a\big\{x,g(\beta^T x)\big\}f_a\big\{x,g(\beta^T x)\big\}+\\
&m\big\{x,g(\beta^T x)\big\}f_{aa}\big\{x,g(\beta^T x)\big\}\Big]\dot{g}^2(\beta^T x)xx^T\\
&+\Big[m_a\big\{x;g(\beta^T x)\big\}f_{X,A}\big\{x;g(\beta^T x)\big\}+m\big\{x,g(\beta^T x)\big\}f_a\big\{x;g(\beta^T x)\big\}\Big]\ddot{g}(\beta^T x)xx^T,
\end{aligned}
$$

and $\partial^2 B_n(x;\beta)/(\partial\beta\partial\beta^T)$ converges in probability to

$$
2f_{aa}\big\{x,g(\beta^T x)\big\}\dot{g}^2(\beta^T x)xx^T+f_a\big\{x,g(\beta^T x)\big\}\ddot{g}(\beta^T x)xx^T.
$$

Together with the previous convergence results for $\tilde{A}_n(x;\beta)$, $\tilde{B}_n(x;\beta)$, $A_n(x;\beta)$, $B_n(x;\beta)$, $C_n(x)$, we obtain that $D_n(\beta)$ converge in probability to

$$
\int_x\Big[m_{aa}\big\{x,g(\beta^T x)\big\}\dot{g}^2(\beta^T x)+m_a\big\{x,g(\beta^T x)\big\}\ddot{g}(\beta^T x)\Big]f_X(x)xx^T dx=D(\beta).
$$

Proof of Equation (2.20)

For notation, let x_l be the l th component of the vector $\mathbf{x}$, and β_j and β_k are the j th and k th component of vector β, $j,k,l\in\{1,...d\}$. Let $S_{n,l}(\beta)$ be the l th component of vector $S_n(\beta)$. Since we have proved that $\hat{\beta}_n-\beta^*$ converge in probability to 0, to prove Equation (2.20) it suffices to show that: $\partial^2 S_{n,l}(\beta)/(\partial\beta\partial\beta^T)=O_p(1)$, By calculation, for $j,k,l\in\{1,...,d\}$:

$$
\begin{aligned}
&\frac{\partial^2}{\partial\beta_j\partial\beta_k}S_{n,l}(\beta)\\
&=\int_x\Bigg\{\frac{\frac{\partial^2}{\partial\beta_j\partial\beta_k}\tilde{A}_{n,l}(x;\beta)}{B_n(x;\beta)}-3\frac{\tilde{B}_{n,k}(x;\beta)\frac{\partial}{\partial\beta_j}\tilde{A}_{n,l}(x;\beta)}{B_n^2(x;\beta)}-3\frac{\tilde{A}_{n,l}(x;\beta)\frac{\partial}{\partial\beta_k}\tilde{B}_{n,j}(x;\beta)}{B_n^2(x;\beta)}\\
&-\frac{A_n(x;\beta)\frac{\partial^2}{\partial\beta_j\partial\beta_k}\tilde{B}_{n,l}(x;\beta)}{B_n^2}+6\frac{\tilde{A}_{n,l}(x;\beta)\tilde{B}_{n,j}(x;\beta)\tilde{B}_{n,k}(x;\beta)}{B_n^3}\\
&+2\frac{A_n(x;\beta)\big(\frac{\partial\tilde{B}_{n,l}}{\partial\beta_k}\tilde{B}_{n,j}+\frac{\partial\tilde{B}_{n,j}}{\partial\beta_k}\tilde{B}_{n,l}+\frac{\partial\tilde{B}_{n,l}}{\partial\beta_j}\tilde{B}_{n,k}\big)}{B_n^3}\\
&-6\frac{A_n(x;\beta)\tilde{B}_{n,l}(x;\beta)\tilde{B}_{n,j}(x;\beta)\tilde{B}_{n,k}(x;\beta)}{B_n^4(x;\beta)}\Bigg\}C_n(x)dx,
\end{aligned}
$$

where $\tilde{A}_{n,j}(x;\beta)$ is the j th component of vector $\tilde{A}_n(x;\beta)$ and $\tilde{B}_{n,j}(x;\beta)$ is the j th component of vector $\tilde{B}_n(x;\beta)$. With similar calculation as before, under the assumption that the first four orders of derivatives of $m(x,a)$ and $f_{X,A}(x,a)$ are bounded, we obtain that: $\partial^2 S_{n,l}(\beta)/(\partial\beta_j\partial\beta_k) = O_p(1)$. Thus $\partial^2 S_{n,l}(\beta)/(\partial\beta\partial\beta^T) = O_p(1)$.

2.9.3 Estimation of Covariance

From above, the covariance of the asymptotic distribution for $(nh_x^q h_a^3)^{1/2}(\hat{\beta}_n - \beta^*)$ is given by:

$$D(\beta^*)^{-1}\Sigma_S(\beta^*)D(\beta^*)^{-1}.$$

First, $D(\beta)$ can be estimated with $D_n(\beta)$. Then for the estimation of $\Sigma_S(\beta)$, notice that $\Sigma_S(\beta) = \int_{x_1}\int_{x_2}\Sigma_{\mathcal{G}}(x_1,x_2;\beta)dx_1dx_2 = \int_x \Sigma_G(x;\beta)dx$, and

$$(nh_x^q h_a^3)^{1/2}G_n(x;\beta) = (nh_x^q h_a^3)^{1/2}\Big\{\frac{1}{n}\sum_{i=1}^{n}\Phi_i(x;\beta)\Big\}\frac{C_n(x)}{B_n^2(x;\beta)} + o_p(1).$$

Therefore, $\Sigma_G(x;\beta) = \Sigma_\Phi(x;\beta)C_n^2(x)/B_n^4(x;\beta)$ where $\Sigma_\Phi(x;\beta)$ can be estimated empirically by the sample covariance of $\Phi_i(x;\beta)$. $\Phi_i(x;\beta)$ is approximated by plugging in the $A_n(x;\beta)$, $B_n(x;\beta)$, $C_n(x)$ for $A(x;\beta)$, $B(x;\beta)$ and $C(x)$. Finally, we plug in $\hat{\beta}_n$ for β^* to obtain the estimated covariance.

CHAPTER

3

CAUSAL EFFECT ESTIMATION AND OPTIMAL DOSE SUGGESTIONS IN MOBILE HEALTH

3.1 Introduction

With the development of mobile applications and portable health monitoring devices, a large amount of data from the patients are becoming available for the purpose of disease monitoring and healthcare interventions. Mobile applications for diseases such as obesity, high blood pressure, cardiology are gaining popularity by providing real-time tracking of the physical status of the patients and sending alerts for high risk events. For diseases such as diabetes, dose suggestions based on the collected data of the patient is also becoming a common feature for the mobile health applications. Healthcare interventions using mobile applications have the great potential to deliver low-cost healthcare services with more accuracy regardless of time and location. However, there is limited literature on evaluating the effect of the treatments under the mobile health setting. Statistical methodologies for recommending dosages using mobile health data are also of great interest to researchers. Analyzing mobile health data can be challenging because there are typically a large

number of time points, time-varying treatments, and a non-definite time horizon (Luckett et al. 2019).

In this chapter, we first study the statistical methodologies for evaluating the causal effect of mobile health interventions. Generalized estimating equations (GEE) are commonly used for studying dependence of an outcome variable on a set of covariates observed over time (Liang and Zeger 1986; Zhao and Prentice 1990; Liang et al. 1992; Schafer 2006). GEE enhance the efficiency of the generalized linear models by including into the estimation equations the correlations among repeated observations of a subject over time. Such approaches typically require a full working correlation model and will be computationally expensive as the time points get larger. Application of GEE in mobile health data has been limited to time-invariant treatments(Evans et al. 2012; Carrà et al. 2016). Liao et al. (2015) proposed the micro-randomized trial design for estimating the causal effect of just-in-time treatments under the mobile health setting (Klasnja et al. 2015; Liao et al. 2016; Dempsey et al. 2015). Liao et al. (2016) and Boruvka et al. (2018) defined the proximal and lagged treatment effects for time-varying treatments with data from micro-randomized trials. A centered and weighted estimation method based on inverse probability of treatment-estimators (Robins et al. 2000; Murphy et al. 2001) is then proposed for estimating these causal effects.

Providing personalized treatment suggestions based on mobile health data is also of great interest. Dynamic treatment regimes (DTR) have been proposed for providing sequential treatment suggestions based on longitudinal data from randomized trials or observational data (Murphy 2003; Moodie et al. 2007; Kosorok and Moodie 2015; Chakraborty and Moodie 2013). A dynamic treatment regime is a set of decision rules that decide which treatments to assign to the patients according to patients' time-varying measurements during the ongoing treatment process. An optimal DTR is the one that yields the most favourable expected mean outcome over a fixed period of time. Optimal dynamic treatment regimes are typically estimated by backward induction based on parametric models for the expected outcome (Q-learning) (Watkins and Dayan 1992; Sutton et al. 1998; Murphy 2005; Schulte et al. 2014). Robustness of these methods can be further enhanced by using semi-parametric models (Murphy 2003; Robins 2004; Moodie et al. 2007; Tang and Kosorok 2012; Schulte et al. 2014) or non-parametric models (Zhao et al. 2009). Zhao et al. (2015) avoid the risk of model misspecification by directly maximizing a nonparametric estimation of the cumulative reward among a predefined class of treatment regimes.

However, mobile health data usually have infinite time horizons. Sequential decision making process in infinite horizon can be modeled as a Markov decision process (Puterman

2014). Ertefaie and Strawderman (2018) defined the optimal DTR in the infinite horizon as the one which maximizes the expected cumulative discounted reward. To estimate the optimal DTR, a parametric model is first posited for the maximum expected cumulative discounted reward. Least square estimation equations are then constructed based on the Bellman equation (Sutton et al. 1998). The optimization is achieved through greedy gradient Q-learning (Maei et al. 2010). Luckett et al. (2019) proposed the V-learning method for finding the optimal DTR. They first posit a model for the expected cumulative discounted reward of a specific treatment regime. Then they search for the treatment regime which maximizes the estimated cumulative discounted reward function within a prespecified class of treatment regimes. However, both of these two methods are limited to discrete states(covariates) and treatments.

In this chapter, we extend Boruvka et al. (2018)'s definition of lagged effect to continuous treatments and use a kernel-based structural nested model for estimating causal effects of continuous treatments based on mobile health data. Then, the key interest is to find the treatment regime which optimize the outcome or minimize the risk of adverse events within a time period in the near future. We define a weighted advantage function as a measurement of the treatment effect over a time period in the near future. The optimal treatment at a certain time point is defined as treatment which maximizes this weighted advantage. The rest of the chapter is structured as follows. In Section 3.2, we formalize the problem in a statistical framework. Existing methods for evaluating treatment effects with longitudinal data are reviewed in Section 3.3. We then define the lag k treatment effect for continuous doses and present the proposed methodology for estimating this lagged effect in Section 3.4. A weighted advantage is also defined. The strategy for making dose suggestions is discussed based on the proposed framework. Theoretical results related to the proposed method are given in Section 3.5. Section 3.6 presents the simulation study. In Section 3.7, we apply the proposed method to the Ohio type 1 diabetes dataset. Discussions and conclusions are in Section 3.8. Finally, the detailed proof of the theoretical results and additional results for the simulation studies and the real data application are given in Section 3.9.

3.2 Problem Setting

We assume that for each individual, the measurements are taken at time points with fixed time intervals, $t = 1, ..., T$. Let $A_t \in \mathcal{A}$ denotes the treatment at decision time t, where $\mathcal{A}$ is a continuous interval of possible values of doses. $X_t \in \mathbb{R}^q$ are covariates measured at time t. $Y_t \in \mathbb{R}$ denotes the outcome measured at time t following the

decision A_{t-1}, $t > 1$. Without loss of generality, we assume that higher values of Y_t denote better outcomes. We assume that X_t and Y_t are observed simultaneously and A_t is a decision made after observing X_t and Y_t. Thus, the observed data for one subject are $\{(X_1, A_1),(Y_2, X_2, A_2), \ldots, (Y_T, X_T, A_T), (Y_{T+1}, X_{T+1})\}$. Assume that the data consist of n subjects. For the i-th subject, the observed data are denoted as $\{(X_1^i, A_1^i),(Y_2^i, X_2^i, A_2^i), \ldots, (Y_T^i, X_T^i, A_T^i), (Y_{T+1}^i, X_{T+1}^i)\}$, $i = 1, \ldots, n$. Let the overbar denotes the history of a random variable, $\bar{X}_t = (X_1, ..., X_t)$. All information accrued up to time t can be represented by $H_t = (\bar{X}_t, \bar{Y}_t, \bar{A}_{t-1})$. To illustrate this model with an example, in a type 1 diabetes study, A_t could be the rapid-reacting insulin dose at time t. Y_t could be the stability of the blood glucose between time $t-1$ and t. X_t could include the food intake, exercise and blood glucose levels.

To define the treatment effects, we adopt the potential outcome framework by Rubin (1974). The potential outcomes include treatments expressed as potential outcomes of past treatments. $X_t(\bar{a}_{t-1})$ and $Y_t(\bar{a}_{t-1})$ are the potential measurements of covariates and potential outcomes at time t had the sequence of treatments $\bar{a}_{t-1}$ been allocated to the patient, $\bar{a}_{t-1} \in \mathcal{A}^{t-1}$. $A_t(\bar{a}_{t-1})$ is defined as the potential treatment at t had the sequence of $\bar{a}_{t-1}$ be allocated. This notation implicitly assumes that the potential outcomes are not influenced by future treatments and the outcome of one subject is not affected by the treatments received by other subjects. The latter is also known as the stable unit treatment value assumption (SUTVA; see Rubin (1974)). For simplicity, we denote $A_2(A_1)$ by A_2, $A_t(\bar{A}_{t-1})$ by A_t. Then $H_t(\bar{A}_{t-1}) = \{X_1, A_1, Y_2(A_1), X_2(A_1), A_2(A_1), \ldots, Y_t(\bar{A}_{t-1}), X_t(\bar{A}_{t-1})\}$.

A dynamic treatment regime $\pi = (\pi_1, ..., \pi_T)$ is a set of rules that outputs a distribution of treatment options at each time point based on past history $\pi_t = \{f_{\pi,t}(a|h_t), a \in \mathcal{A}\}$; $f_{\pi,t}$ here denotes the conditional density of choosing treatment a given history h_t at time t. Let $\mathcal{H}_t$ be the space of all possible histories. A treatment regime is deterministic if $f_{\pi,t}(a|h_t) = \delta(a = g_t(h_t))$, for some $g_t : \mathcal{H}_t \to \mathcal{A}$, where $\delta(\cdot)$ is the Dirac delta function. Then for simplicity of notation, we write π_t as $\pi_t = g_t(h_t)$.

To use the observed data to estimate the treatment effect, we make the following assumptions(Robins 2004):

- Consistency: The potential outcomes had the treatments given to the patient equal to the observed treatment history $\bar{A}_T$ are equal to the observed data. More specifically, for $\bar{a}_{t-1} = \bar{A}_{t-1}$, $\bar{Y}_t(\bar{a}_{t-1}) = \bar{Y}_t$, $\bar{X}_t(\bar{a}_{t-1}) = \bar{X}_t$ and $\bar{A}_t(\bar{a}_{t-1}) = \bar{A}_t$ for $2 \le t \le T$, where the left sides of the equations are the potential outcomes and the right sides are the observed variables; At time $T+1$, $\bar{Y}_{T+1}(\bar{a}_T) = \bar{Y}_{T+1}$, and $\bar{X}_{T+1}(\bar{a}_T) = \bar{X}_{T+1}$ for $\bar{a}_{t-1} = \bar{A}_{t-1}$.

- Positivity: For any treatment $a \in \mathcal{A}$ and past history $h_t \in \mathcal{H}_t$, $f_{A|H}(A_t = a|H_t = h_t) > 0$, where $f_{A|H}(A_t|H_t)$ denotes the conditional distribution for treatment assignments from the dataset. In other words, all treatments $a \in \mathcal{A}$ can possibly be observed given h_t for any $h_t \in \mathcal{H}_t$. In randomized trials, this assumption can be ensured by the design of the study. However, in observational studies, the set of possible treatments may differ with different treatment history. In such cases, the expected outcomes of a treatment regime π cannot be estimated from the observed data when it has a non-zero probability of suggesting treatments where $f_{\pi,t}(a|h_t) > 0$ and $f_{A|H}(A_t = a|H_t = h_t) = 0$. Such treatment regimes are regarded as non-identifiable treatment regimes (Robins 2004). We could then limit our attention to identifiable treatment regimes by adding constraints to the suggested treatments. In practice, when $\mathcal{A}$ is unknown, this assumption would be hard to examine. In such scenarios, causal inference for treatments that are uncommon in the observed dataset should be made with caution.

- Sequential ignorability: The potentials outcomes $\{Y_{t+1}(\bar{a}_t), X_{t+1}(\bar{a}_t), A_{t+1}(\bar{a}_t), ..., Y_{T+1}(\bar{a}_T), X_{T+1}(\bar{a}_T)\}$ are independent of A_t conditional on H_t, for $t \leq T$ and $\bar{a}_t \in \mathcal{A}^T$. This assumption is also known as the no unmeasured confounder assumption. It is naturally satisfied in a sequentially randomized study, where treatments are randomized for each time point. In an observational study, this assumption cannot be verified and is often assumed.

3.3 Literature Review

In this section, we review the existing methods for evaluating treatment effects with longitudinal data. In particular, we are interested in methods that evaluate the treatment effects of time-varying treatments.

Generalized Estimating Equations

The most naive approach for estimating the effect of treatments on a response variable is using generalized linear models. Assuming that the covariates, treatments and outcomes at different time points (X_t, A_t, Y_{t+1}) are independent from each other (notice that in our setting, the outcome is at the next time point), then the generalized linear models can be used to estimate $\mu(X_t, A_t) = E(Y_{t+1}|X_t, A_t)$. The treatment can be regarded as a covariate and the effect of the treatment can be estimated by the estimated parameter corresponding to the treatment in the linear component. Let $\tilde{X}_t = (X_t^T, A_t)^T$. The marginal density of

Y_{t+1} is modeled as a distribution from the exponential family with conditional mean:

$$E(Y_{t+1}|\tilde{X}_t) = \mu(\tilde{X}_t;\beta) = g(\beta^T\tilde{X}_t).$$

where $g(\cdot)$ is a known function, β is a vector of parameters to be estimated. Under the assumption that all the observations are independent from each other (meaning that the observations of the same subject taken at different time points are also independent from each other), the parameters can be consistently estimated by the score equation:

$$U(\beta) = \mathbb{P}_n \frac{\partial\bar{\mu}(\tilde{X};\beta)}{\partial\beta}\{\bar{Y} - \bar{\mu}(\tilde{X};\beta)\} = 0,$$

where $\bar{\mu}(\tilde{X};\beta) = (\mu(\tilde{X}_1;\beta), \mu(\tilde{X}_2;\beta), ..., \mu(\tilde{X}_T;\beta))^T$, $\bar{Y} = (\bar{Y}_2, Y_3..., Y_{T+1})^T$. However, the independence assumption is likely to be unvalid with longitudinal data because the repeated observations from the same objects are usually correlated with each other. For example, the blood glucose of a person throughout the day should be changing continuously. If we measure the blood glucose every 10 minutes, two continuous measurements are likely to be close to each other. Failing to model the correlation among repeated observations of the same subject would lead to incorrect inference of the parameters, even if the proposed model is correct.

To take the correlation between repeated observations into account, Liang and Zeger (1986) proposed the generalized estimating equations. They used a working correlation matrix $R(\alpha)$ for the T observations of one subject, where α is a parameter to be estimated. For $i \in \{1, \ldots, n\}$, let var(Y_{t+1})= $a\{\mu(\tilde{X}_t)\}\phi$, where $a(\cdot)$ is a known function defined by the distribution from the exponential family and ϕ is the dispersion parameter that needs to be estimated as well. Let $A^i = \text{diag}\Big(a\{\mu(\tilde{X}_1^i)\}, a\{\mu(\tilde{X}_2^i)\}, \ldots, a\{\mu(\tilde{X}_T^i)\}\Big)$ be a $T \times T$ diagonal matrix and define $V^i = {A^i}^{1/2} R(\alpha) {A^i}^{1/2}\phi$. Notice that if $R(\alpha)$ is the true correlation matrix for $Y_{i,2}, ..., Y_{i,T+1}$, then V^i is the true covariance matrix for $Y_{i,2}, ..., Y_{i,T+1}$. The general estimation equation for the parameters then becomes:

$$U(\beta,\phi,\alpha) = \sum_{i=1}^{n} \frac{\partial\bar{\mu}(\tilde{X}^i;\alpha)}{\partial\beta} V^i(\alpha,\phi)^{-1}\{\bar{Y}^i - \bar{\mu}(\tilde{X}^i;\beta)\} = 0.$$

To estimate β, we can first obtain the moment estimation of α and ϕ: $\hat{\alpha}$ and $\hat{\beta}$. Then $\hat{\beta}$ can be obtained by solving the estimation equation by $U(\beta, \hat{\alpha}, \hat{\phi}) = 0$.

The weakness of the model is that it assumes the treatment effect to be constant across time. Furthermore, the model assumes that the treatments in the past do not have any impact on the future responses (for example, the treatment at time t do not influence

the response at time $t+2$), which is usually unvalid for dynamic treatments regimes. With a dynamic treatment regime, the treatment effect of the current decision will further influence the decisions and responses in the future. Therefore, alternative methods are needed to estimate the treatment effect for dynamic treatment regimes.

3.3.1 Marginal Mean Models

To evaluate the time-varying treatment effect of a dynamic treatment regime, it is essential to estimate the counterfactual responses (also referred to as the potential outcome), which are the responses that would have been observed had the patient followed a treatment regime of interest. In this section, we first consider a single response at the end of the medical process Y and binary treatment options $A_t \in \{0,1\}, t = 1, \ldots T$. Thus the observed data for one subject are $\{(X_1, A_1), (X_2, A_2), \ldots, (X_T, A_T), Y\}$ and $H_t = (\bar{X}_t, \bar{A}_{t-1})$. We are interested in estimating $E\{Y(\pi_1, \ldots, \pi_T)\}$ where $Y(\pi_1, \ldots, \pi_T)$ is the potential outcome had the treatments followed the treatment regime π. More specifically, when the treatments follow the treatment regime π, $A_1 = \pi_1(H_1)$, $A_2 = \pi_2\{H_2(A_1)\}$, $\ldots$, $A_T = \pi_T\{H_T(\bar{A}_{T-1})\}$.

Let S indicate certain subpopulation we are interested in. For simplicity of notation, we assume that S is included in X_1. Let π^{obs} be the actual treatment regime that generates the observed dataset, so $\pi_t^{obs}(a|x) = p_{A_t|X_t}(a|x)$ for $a \in \mathcal{A}$ and $x \in \mathbb{R}^q$. For a treatment regime π, define

$$W_{\pi,t}(H_t) = \Pi_{j=1}^{t} \frac{p_{\pi,j}(A_j|H_j)}{p_{\pi^{obs},j}(A_j|H_j)}.$$

which is the ratio of the probability of the treatment history $\{A_1, \ldots A_t\}$ under the treatment regime π versus the probability under the treatment regime π^{obs} given H_t. It can be shown that if a treatment regime π is identifiable from the observed data, then under the sequential ignorability assumption, $E[Y(\pi_1, \ldots, \pi_T)|S]$ can be estimated with:

$$E\Big[Y(\pi_1, \ldots, \pi_T)|S = s\Big] = E\Big[W_{\pi,T}(H_T)Y|S = s\Big]$$
$$= \sum_{\bar{a}_T \in \bar{\mathcal{A}}} \int_{\bar{x}_{T-1}} \int_y \Bigg[\Big\{ \Pi_{t=1}^{T} p_{\pi,t}(a_t|h_t) \Big\} y f_{Y|\bar{A}_T, \bar{X}_T}(y|\bar{a}_T, \bar{x}_T) \Big\{ \Pi_{t=2}^{T-1} f_{X_t|\bar{A}_{t-1}, \bar{X}_{t-1}}(x_t|\bar{a}_{t-1}, \bar{x}_{t-1}) \Big\}$$
$$f_{X_1|S}(x_1|s) \Bigg] dy d\bar{x}_{T-1} \text{(Murphy et al. 2001).}$$

The $\bar{\mathcal{A}}$ in the equation above denotes the space for $\bar{A} = \{A_1, \ldots, A_T\}$. This is also referred

to as the G-computation formula (Robins 1986). However, calculation of the items on the right side of the above formula needs the knowledge of the complete distributions of $f_{Y|\bar{A}_T,\bar{X}_T}$ and $f_{X_t|\bar{A}_{t-1},\bar{X}_{t-1}}$ for $t = 1, \ldots, T-1$, which is usually not available in real application. To obtain $E\{Y(\pi_1, \ldots, \pi_T)|S\}$ directly from the above equation is thus not feasible. Nonetheless, given a parametric model for $E[Y(\pi_1, \ldots, \pi_T)|S]$, say $\mu(S; \beta_\pi)$, a weighted version of the generalized linear regression estimation equation gives an unbiased estimation for the parameters (Robins 1999):

$$\mathbb{P}_n\Big[W_{\pi,T}(H_T)\frac{\partial\mu(S;\beta_\pi)}{\partial\beta_\pi}\{Y - \mu(S;\beta_\pi)\}\Big] = 0.$$

Estimation equations with improved efficiency are given in Murphy et al. (2001).

This method is also referred to as marginal mean models. It allows us to evaluate the effect of a specific dynamic treatment regime on a single outcome variable. However, this equation above requires a full model for the conditional expectation of the response given the treatment regime, while we are really more interested in the comparison of the treatment effects between different treatments. Murphy (2003) and Robins (2004) thus proposed methods that are more robust to model misspecification by directly modeling the difference between the expected outcomes given different treatment regimes.

Blip Effect Estimation Using Structural Nested Models

Robins (2004) defined a blip treatment effect. This treatment effect is also defined under the setting where the goal is to optimize a single response at the end of the medical process. Given two treatment regimes $\pi = (\pi_1, ..., \pi_T)$, $\pi^* = (\pi_1^*, \ldots, \pi_T^*)$ and a treatment history $\bar{a}_{t-1}$, let the potential response Y under the regime $\{\bar{a}_{t-1}, \pi^*, \underline{\pi}_t\}$ to be $Y(\bar{a}_{t-1}, \pi^*, \underline{\pi}_t)$. Here, $\{\bar{a}_{t-1}, \pi^*, \underline{\pi}_{t+1}\}$ denotes a regime which follows $\bar{a}_{t-1}$ till $t-1$, follows π^* at time t and then follows the regime π from time $t+1$ to T. π^* is usually taken as a baseline regime. Common choices for π^* can be $\pi^* \equiv 0$ or $\pi^* \equiv \pi^{obs}$. The blip effect is defined as:

$$\gamma_t^{\pi,\pi^*}(\bar{x}_t, \bar{a}_t) = E\{Y(\bar{a}_t, \underline{\pi}_{t+1}) - Y(\bar{a}_{t-1}, \pi_t^*, \underline{\pi}_{t+1})|\bar{X}_t = \bar{x}_t, \bar{A}_t = \bar{a}_t\}. \tag{3.1}$$

This blip function measures the effect of treatment a_t versus $\pi_t^*(\bar{x}_t, \bar{a}_{t-1})$ at time t. When $\pi^* = \pi$, $\gamma_t^{\pi,\pi}(\bar{x}_t, \bar{a}_{t-1})$ measures the effect of the treatment a_t versus $\pi_t(\bar{x}_t, \bar{a}_{t-1})$. To estimate this blip effect, we need to first propose a parametric model for $\gamma_t^{\pi,\pi}(\bar{x}_t, \bar{a}_t)$, say

$\gamma_t^\pi(\bar{x}_t, \bar{a}_t; \psi)$, which satisfies $\gamma_t^\pi(\bar{x}_t, \bar{a}_t; \psi) = 0$ if $a_t = \pi_t(\bar{x}_t, \bar{a}_{t-1})$. Let:

$$U_t^\pi(\psi) = Y - \sum_{j=t}^{T} \Big\{ \gamma_t^\pi(\bar{X}_t, \bar{A}_t; \psi) \Big\}.$$

The author showed that, under the sequential randomization assumption, an efficient estimation equation for the parameters ψ would be:

$$\mathbb{P}_n \sum_{t=1}^{T} \Big[U_t^\pi(\psi) - E\Big\{ U_t^\pi(\psi) | \bar{A}_{t-1}, \bar{X}_{t-1} \Big\} \Big] \Big[d_t(\bar{A}_t, \bar{X}_t) - E\Big\{ d_t(\bar{A}_t, \bar{X}_t) | \bar{A}_{t-1}, \bar{X}_t \Big\} \Big] = 0 \tag{3.2}$$

where $d_t(\bar{A}_t, \bar{X}_t)$ is a vector of functions with the same dimension as the parameter ψ, arbitrarily chosen by the researcher. Robins (2004) also showed that the above equation is most efficient when $d_t(\bar{A}_t, \bar{X}_t) = \partial U_t^\pi(\psi)/\partial\psi$. To obtain $E\{U_t(\psi)|\bar{A}_{t-1}, \bar{X}_{t-1}\}$ and $E\Big\{ d_t(\bar{A}_t, \bar{X}_t)|\bar{A}_{t-1}, \bar{X}_t \Big\}$, we need to know the conditional density $f_{A_t|\bar{A}_{t-1},\bar{X}_t}(a_t|\bar{a}_{t-1}, \bar{x}_t)$ and the conditional expectation $E\{Y|\bar{A}_{t-1}, \bar{X}_t\}$. In an observational study, $f_{A_t|\bar{A}_{t-1},\bar{X}_t}$ is usually unknown. Therefore, working models need to be posited for $f_{A_t|\bar{A}_{t-1},\bar{X}_t}(a_t|\bar{a}_{t-1}, \bar{x}_t)$ and $E\{Y|\bar{A}_{t-1} = \bar{a}_{t-1}, \bar{X}_t = \bar{x}_{t-1}\}$, say $\tilde{f}_t(\bar{a}_{t-1}, \bar{x}_{t-1}; \alpha)$ and $m(\bar{a}_{t-1}, \bar{x}_t; \xi)$. To apply the estimation method, we first need to obtain estimators for α and ξ: $\hat{\alpha}$ and $\hat{\xi}$. Then $\hat{\psi}(\hat{\alpha}, \hat{\xi})$ can be calculated by solving Equation (3.2). This method is also referred to as structural nested models. Robins (2004) proved that under the sequential ignorability assumption and a correctly specified model for $\gamma_t^\pi(\bar{x}_t, \bar{a}_t)$, the above estimation equation leads to consistent and asymptotic normal estimators for ψ if either $\tilde{f}_t(\bar{a}_{t-1}, \bar{x}_{t-1}; \alpha)$ or $m(\bar{a}_{t-1}, \bar{x}_t; \xi)$ is correctly specified. This property is also referred to as the double robustness of the estimation equations. For more details, please refer to Robins (2004).

Proximal Treatment Effect and Lagged Treatment Effect

Boruvka et al. (2018) further extended the above method to the mobile health setting where a response is observed at each time point $\{Y_2, \ldots, Y_{T+1}\}$. The treatments considered in this setting are alert messages sent by the applications to the patients to encourage certain behaviors or to prevent high-risk events. Examples include exercise reminders for obese patients, and mindfulness messages for stress management applications. The goal is then to examine the effect of these messages on these responses through out time. For example, for weight control applications, we would want to evaluate the effect of workout reminder messages on the amount of workout done by the application user. Data from micro-randomized trials (MPT; Klasnja et al. 2015; Liao et al. 2015; Dempsey et al. 2015)

can be used for this purpose. In micro-randomized trials, each participant is sequentially randomized to the treatments (receiving message or not receiving message at time t, $t = 1, \ldots, T$) for a large number of times. Measurements of the covariates of interest are taken intensively throughout the trial. This can be modeled under a binary treatment setting where $A_t = 1$ for a message sent and 0 for no messages sent. Let Y_t be the total amount of mobile-tracked movements from time $t-1$ to t. Boruvka et al. (2018) proposed a lag k treatment effect for treatments at time t on the future outcome Y_{t+k}:

$$E\Big\{Y_{t+k}(\bar{A}_{t-1}, 1, A_{t+1}^{a_t=1}, \ldots, A_{t+k-1}^{a_t=1}) - Y_{t+k}(\bar{A}_{t-1}, 0, A_{t+1}^{a_t=0}, \ldots, A_{t+k-1}^{a_t=0}) | S_t(\bar{A}_{t-1})\Big\}, \quad (3.3)$$

where $A_{t+1}^{a_t=a}$ denotes the potential treatment $A_{t+1}(\bar{A}_{t-1}, A_t = a)$, $A_{t+l}^{a_t=a}$ denotes $A_{t+l}(\bar{A}_{t-1}$, $A_t = a$, $A_{t+1}^{a_t=a}$, ..., $A_{t+l-1}^{a_t=a})$, for $l = 2, \ldots, k-1$; $S_t(\bar{A}_{t-1})$ is a summary statistic of the potential history $H_t(\bar{A}_{t-1})$. Using our notation from previous sections, we can also write it as:

$$E\Big\{Y_{t+k}(\bar{A}_{t-1}, 1, \pi_{t+1}^{obs}, \ldots, \pi_{t+k-1}^{obs}) - Y_{t+k}(\bar{A}_{t-1}, 0, \pi_{t+1}^{obs}, \ldots, \pi_{t+k-1}^{obs}) | S_t(\bar{A}_{t-1})\Big\}. \quad (3.4)$$

Notice that this definition of treatment effect is marginalized over all information of $H_t(\bar{A}_{t-1})$ that is not included in $S_t(\bar{A}_{t-1})$. When $S_t = H_t$, Equation (3.4) is a generalization of the blip treatment effect in Equation (3.1) when π^* is taken to be the constant regime $\pi^* \equiv 0$ and π is taken to be the actual treatment regime that generates the dataset : $\pi_t^{obs}(a_t|h_t) = p_{A_t|H_t}(A_t = a_t|H_t = h_t)$. To see this connection more clearly, let us define:

$$\gamma_{t,t+k}^{\pi,\pi^*}(h_t, \bar{a}_t) = E\{Y_{t+k}(\bar{A}_t, \underline{\pi}_{t+1}) - Y_{t+k}(\bar{A}_{t-1}, \pi_t^*, \underline{\pi}_{t+1}) | H_t(\bar{A}_{t-1}) = h_t, \bar{A}_t = \bar{a}_t\}$$

When $S_t = H_t$, under the consistency and the sequential ignorability assumption, this lag k treatment effect is equal to the blip effect with the form:

$$\gamma_{t,t+k}^{\pi,\pi^*}(h_t, \bar{a}_{t-1}, a_t = 1) =$$
$$E\Big\{Y_{t+k}(\bar{A}_t, \underline{\pi}_{t+1}^{obs}) - Y_{t+k}(\bar{A}_{t-1}, \pi_t^* = 0, \underline{\pi}_{t+1}^{obs}) | H_t(\bar{A}_{t-1}) = h_t, \bar{A}_t = \bar{a}_t, a_t = 1\Big\}.$$

By conditioning on S_t instead of the complete covariate history H_t, this definition allows the possibility of addressing specific scientific questions during micro-randomized trials by using only the summary information we are interested in. For example, if we are interested in the average treatment effect marginalized on all previous information, we can take $S_t = \emptyset$. If we interested in the average treatment effect for patients in different stages of the disease, we can take S_t as the variable indicating the stage of the disease. Notice,

for treatments A_l ($l \leq t-1$) not included in $S_t(\bar{A}_{t-1})$, the treatment effect defined in Equation (3.3) depends on the distribution of A_l. This treatment effect is causal only for the variables contained in $S_t(\bar{A}_{t-1})$.

Under the assumptions of consistency and sequential ignorability, it can also be shown that:

$$E\Big\{Y_{t+k}(\bar{A}_{t-1}, 1, A_{t+1}^{a_t=1}, \ldots, A_{t+k-1}^{a_t=1}) - Y_{t+k}(\bar{A}_{t-1}, 0, A_{t+1}^{a_t=0}, \ldots, A_{t+k-1}^{a_t=0}) | S_t(\bar{A}_{t-1})\Big\}$$
$$= E\Big[E\{Y_{t+k}|A_t = 1, H_t\} - E\{Y_{t+k}|A_t = 0, H_t\}|S_t\Big].$$

Boruvka et al. (2018) proposed a method for estimating this lag k treatment effect. First, a model is posited for the lag k treatment effect in the following form:

$$E\Big[E\{Y_{t+k}|A_t = 1, H_t\} - E\{Y_{t+k}|A_t = 0, H_t\}|S_t\Big] = g_{kt}(S_t)^T\beta_k,$$

where $g_{kt}(\cdot)$ is a vector of functions. Assuming that the treatments are sequentially randomized with a known probability $p_{A_t|H_t}(A_t = 1|H_t)$, the parameters can then be estimated with the following estimation equation:

$$0 = \sum_{t=1}^{T-k+1} \Big[Y_{t+k} - c_{kt}(H_t)^T\alpha_k - \Big\{A_t - \tilde{p}(1|S_t)\Big\}g_{kt}(S_t)^T\beta_k\Big] W_t \begin{bmatrix} c_{kt}(H_t) \\ \{A_t - p_{A_t|S_t}(1|S_t)\}g_{kt}(S_t) \end{bmatrix}, \tag{3.5}$$

where $W_t = \tilde{p}(A_t|S_t)/p_{A_t|H_t}(A_t|H_t)$ and $\tilde{p}(A_t|S_t)$ is an arbitrary probability mass function that depends on H_t only through S_t. $c_{kt}(H_t)^T\alpha_k$ is a working model for $E[W_tY_{t+k}|H_t]$, thus α_k are nuisance parameters. Equation (3.5) has a similar form as Equation (3.2). The use of the weighting allows the estimation equation to be unbiased when the treatment effect is conditional on S_t instead of H_t. This method provides us with a tool for evaluating treatment effects of mobile health interventions for binary treatments.

3.4 Causal Effect Estimation for Continuous Doses with Mobile Health Data

There is great potential for mobile applications to provide dose suggestions for diseases which require frequent medication administration, such as type 1 diabetes. In such cases, the treatment options lie in a continuous dose range. In this section, we extend the above definition of the lag k treatment effect to continuous dose settings $A_t \in \mathcal{A}$, $t = 1, \ldots, T$. We are interested in making causal inference of the treatments conditional on the past history $H_t \in \mathcal{H}_t$ and provide dose suggestions based on the treatment effects. We do not assume that our data come from micro-randomized trials because such trials would not be feasible for actual medications such as insulin doses. Therefore, the distribution of the treatments $A_t|H_t$ is assumed to be unknown.

3.4.1 Lag k Treatment Effect

Define the conditional lag k treatment effect of treatment a with respect to a specific treatment a_0 at time t as:

$$\begin{aligned}&\tau_{t,k}(a, a_0, H_t(\bar{A}_{t-1})) = \\ &E\Big\{Y_{t+k}(\bar{A}_{t-1}, a, A_{t+1}^{a_t=a}, \ldots, A_{t+k-1}^{a_t=a}) \quad - Y_{t+k}(\bar{A}_{t-1}, a_0, A_{t+1}^{a_t=a_0}, \ldots, A_{t+k-1}^{a_t=a_0})|H_t(\bar{A}_{t-1})\Big\}\end{aligned} \tag{3.6}$$

where $k \geq 1$ and a_0 is a baseline dosage. The expectation in equation (3.6) is taken over all the possible future treatments from time t to $t + k - 1$. Notice that when the reference treatment $a_0 = 0$ and $H_t = S_t$, this definition of lag k treatment effect is same as the treatment effect defined by Boruvka et al. (2018).

Under the assumptions of positivity, consistency and sequential ignorability, we can estimate the conditional lag k treatment effect with the observed data for any $a \in \mathcal{A}$ (See Section 3.9 for the proof):

$$\begin{aligned}&E\Big\{Y_{t+k}(\bar{A}_{t-1}, a, A_{t+1}^{a_t=a}, \ldots, A_{t+k-1}^{a_t=a}) - Y_{t+k}(\bar{A}_{t-1}, a_0, A_{t+1}^{a_t=a_0}, \ldots, A_{t+k-1}^{a_t=a_0})|H_t(\bar{A}_{t-1})\Big\} \\ &= E(Y_{t+k}|A_t = a, H_t) - E(Y_{t+k}|A_t = a_0, H_t).\end{aligned} \tag{3.7}$$

3.4.2 Lag K Weighted Advantage

In addition to evaluation of the treatment effect, we also aim to find the treatment regime which maximizes this outcome over a short period of time in the future. Assume that we are interested in finding the dosage at time t which maximizes the outcome from time $t+1$ to $t+K$, then the optimal treatment regime at time t is defined as:

$$\begin{aligned}
&\pi_t^{opt} = \\
&\arg \max_{\pi_t:\mathcal{H}_t\to\mathcal{P}(\mathcal{A})} \sum_{k=1}^{K} w_k E\Bigg\{ Y_{t+k}\Big(\bar{A}_{t-1}, a_t = \pi_t\{H_t(\bar{A}_{t-1})\}, A_{t+1}^{a_t=\pi_t\{H_t(\bar{A}_{t-1})\}}, \dots, \\
&A_{t+k-1}^{a_t=\pi_t\{H_t(\bar{A}_{t-1})\}}\Big) \Bigg| H_t(A_{t-1}) \Bigg\} \\
&= \arg \max_{\pi_t:\mathcal{H}_t\to\mathcal{P}(\mathcal{A})} \sum_{k=1}^{K} w_k \tau_{t,k}\Big(\pi_t\{H_t(\bar{A}_{t-1})\}, a_0, H_t(\bar{A}_{t-1})\Big),
\end{aligned}$$

where $w_1, \dots, w_K$ are predefined non-negative weights and $w_1 + \cdots + w_K = 1$. Thus, we define the lag K weighted advantage to be:

$$\begin{aligned}
&\tilde{\tau}_{t,K}(a, a_0, H_t(\bar{A}_{t-1})) = \sum_{k=1}^{K} w_k \tau_{t,k}(a, a_0, H_t(\bar{A}_{t-1})) \\
&= \sum_{k=1}^{K} w_k \Big\{ Y_{t+k}(\bar{A}_{t-1}, a, A_{t+1}^{a_t=a}, \dots, A_{t+k-1}^{a_t=a}) - \\
&Y_{t+k}(\bar{A}_{t-1}, a_0, A_{t+1}^{a_t=a_0}, \dots, A_{t+k-1}^{a_t=a_0}) | H_t(\bar{A}_{t-1}) \Big\}.
\end{aligned}$$

Then the optimal treatment regime can be written as:

$$\pi_t^{opt} = \arg \max_{\pi_t:\mathcal{H}_t\to\mathcal{P}(\mathcal{A})} \tilde{\tau}_{t,K}\Big(a = \pi_t(H_t(\bar{A}_{t-1})), a_0, H_t(\bar{A}_{t-1})\Big)$$

For example, if we have hourly data of diabetes patients and we want to minimize the amount of time the blood sugar being outside 80-140 mg/dL range within four hours after the dose injection, we could define Y_t as the percentage of time when the blood sugar is inside the optimal range at the t-th hour. Take $K = 4$ and $w_1 = w_2 = w_3 = w_4 = 0.25$. An optimal dose suggestion at time t would be the one which maximizes the lag K weighted advantage at time t. Therefore, we define the optimal treatment regime at time t to be:

$$\pi_t^{opt} = \arg \max_{\pi_t} \tilde{\tau}_{t,K}\Big(\pi_t\{H_t(\bar{A}_{t-1})\}, a_0, H_t(\bar{A}_{t-1})\Big),$$

where the expectation is taken over $H_t(\bar{A}_{t-1})$. Notice that the choice of a_0 does not affect the optimal treatment regime. In this article, we take $a_0 = 0$, which is a meaningful reference dosage. For simplicity of notation, we write $\tau_{t,k}\Big(a, 0, H_t(\bar{A}_{t-1})\Big)$ as $\tau_{t,k}\Big(a, H_t(\bar{A}_{t-1})\Big)$ the rest of this article.

3.4.3 Estimation Method

We use a kernel-based structural nested models to estimate the lag k treatment effect. We first assume a quadratic model for the lag k treatment effects. The following model assumes that the lag k treatment effects for $k = 1, \ldots, K$ depend on H_t only through S_t, where $S_t \in \mathcal{S}$ are some summary statistics of the past history.

$$\begin{aligned} &E(Y_{t+k}|A_t = a, H_t) - E(Y_{t+k}|A_t = 0, H_t) = \tau_{t,k}(a, a_0, H_t) \\ &= \tau_k(a, S_t; \alpha_k, \beta_k) = \alpha_k a^2 + \{\beta_k^T g_k(S_t)\}a, \end{aligned} \tag{3.8}$$

where g_k is a q dimensional function of S_t. Boruvka et al. (2018) showed that the models for the lag k effects for different k do not constrain one another. Notice that we assume that S_t for $t = 1, \ldots, T$ are from the same vector space $\mathcal{S}$ so that the parameters in this model do not vary across different t. The motivation for using a quadratic model is that both underdosing and overdosing might lead to unfavorable outcomes in practice. Let $\alpha = (\alpha_1, \ldots, \alpha_K)^T, \beta = (\beta_1, \ldots, \beta_K)^T$, and $w = (w_1, ..., w_K)^T$. Without loss of generality, we assume that $g_1(\cdot) = \vdots = g_K(\cdot)$. (If $g_1(\cdot), \ldots, g_K(\cdot)$ are not equal, there exists $g(\cdot)$ which includes all the functions from the terms of $\{g_1(\cdot), \ldots, g_K(\cdot)\}$. Thus we can substitute $g_1(\cdot), \ldots, g_K(\cdot)$ with $g(\cdot)$ and model (3.8) will still be correct.) Then the weighted lag K advantage is:

$$\begin{aligned} \tilde{\tau}_K(a, S_t; \alpha, \beta) &= \sum_{k=1}^{K} w_k \tau_k(a, S_t, \alpha_k, \beta_k) \\ &= \{w^T\alpha\}a^2 + \{w^T\beta\}^T g(S_t)a \\ &= \tilde{\alpha}_K a^2 + \tilde{\beta}_K^T g(S_t)a, \end{aligned}$$

where $\tilde{\alpha}_K = w^T\alpha$, $\tilde{\beta}_K = w^T\beta$. The lag K weighted advantage thus also follows a quadratic form. Notice that under the model above, $\tilde{\tau}_{t,K}\big(a, a_0, H_t(\bar{A}_{t-1})\big) = \tilde{\tau}_K(a, S_t; \alpha, \beta)$ also depends on H_t only through S_t. Thus the optimal treatment regime which maximizes the

lag K weighted advantage:

$$\begin{aligned}\pi_t^{opt} &= \arg\max_{\pi_t} \tilde{\tau}_{t,K}(a = \pi_t, a_0, H_t) \\ &= \arg\max_{\pi_t} \tilde{\tau}_K(a = \pi_t, S_t; \alpha, \beta)\end{aligned}$$

also depends on H_t only through S_t. When $\tilde{\alpha}_K < 0$, the optimal dose at time t would be a deterministic treatment regime: $\pi_t^{opt} = -\{\tilde{\beta}_K^T g(S_t)\}/2\tilde{\alpha}_K$. The parameter $-\tilde{\beta}_{K,j}/\tilde{\alpha}_K$ can be interpreted as the difference of the optimal dosage for patients with one unit difference in the j-th term of $g(S_t)$ while having all the other covariates the same, $j = 1, \ldots, q$ where q is the dimension of $g(S_t)$. When $\tilde{\alpha}_K \geq 0$, the optimal treatment falls on the edge of $\mathcal{A}$.

Now we present the method for estimating the lag k causal effect under the proposed model. Let $U_{t+k} = Y_{t+k} - \tau_k(A_t, a_0, H_t)$. Under the proposed model,

$$U_{t+k}(\alpha_k, \beta_k) = Y_{t+k} - \tau_k(A_t, S_t; \alpha_k, \beta_k) = Y_{t+k} - \alpha_k A_t^2 - \beta_k^T g_k(S_t) A_t.$$

According to Theorem 3.3 in Robins (2004), under the assumption of sequential randomization and consistency, we can obtain:

$$E\left[\Big\{d(A_t, H_t) - E(d(A_t, H_t)|H_t)\Big\} \times \Big\{U_{t+k} - E(U_{t+k}|H_t)\Big\}\right] = 0, \tag{3.9}$$

where $d(\cdot, \cdot)$ is an arbitrary function and $t \in \{1, \ldots, T + 1 - k\}$.

Assume that the data consist of n independent subjects $\{H_{T+1}^1, \ldots, H_{T+1}^n\}$, where $H_{T+1}^i = (X_1^i,\ Y_1^i, A_1^i,\ \ldots,\ X_{T+1}^i,\ Y_{T+1}^i)$, $i = 1, \ldots, n$. Then we can estimate α_k, β_k with the following equation:

$$\begin{aligned}0 = \mathbb{P}_n \sum_{t=1}^{T-k+1} &\Big\{d_{t+k}(A_t, H_t) - E\Big(d_{t+k}(A_t, H_t)|H_t\Big)\Big\}\times \\ &\Big\{U_{t+k}(\alpha_k, \beta_k) - E\big(U_{t+k}(\alpha_k, \beta_k)|H_t\big)\Big\},\end{aligned} \tag{3.10}$$

where

$$d_{t+k}(A_t, H_t) = -\frac{\partial U_{t+k}(\alpha_k, \beta_k)}{\partial(\alpha_k, \beta_k)} = \begin{pmatrix} A_t^2 \\ A_t g_k(S_t) \end{pmatrix}. \tag{3.11}$$

Since $d_{t+k}(\cdot)$ depends on H_t only through S_t, we can write it as $d_{t+k}(A_t, S_t)$. Notice that

$E(U_{t+k}|H_t) = E(Y_{t+k}|H_t) - \tau_k(A_t, S_t)$. To apply the estimation equation, we need to obtain $E\{d_{t+k}(A_t, S_t)|H_t\}$ and $E(Y_{t+k}|H_t)$. The traditional approach is to use regression models to estimate these conditional expectations. However, as t increases, the complexity of the model increases as well. Thus, there is a high risk of model misspecification for both of these conditional expectations when t gets large. Instead of using parametric models for these conditional expectations, we estimate these expectations using kernel density estimators. However, the high dimension of H_t can induce large variance to the nonparametric estimators as t gets larger. Therefore we show that, if the following assumption is satisfied:

$$A_t \perp Y_{t+k}(\bar{a}_{t+k-1})|S_t \text{ for } \bar{a}_{t+k-1} \in \mathcal{A}^{t+k-1}, \tag{3.12}$$

then we can obtain the following equation for an arbitrary function $d(\cdot) : \mathcal{A} \times \mathcal{S} \to \mathbb{R}^{q_k+1}$ where q_k is the dimension of $g_k(S_t)$, (see Section 3.9 for proof):

$$E\left[\Big\{d(A_t, S_t) - E(d(A_t, S_t)|S_t)\Big\} \times \Big\{U_{t+k} - E(U_{t+k}|S_t)\Big\}\right] = 0. \tag{3.13}$$

Therefore we can use the estimation equation below to estimate α_k, β_k:

$$0 = \mathbb{P}_n \sum_{t=1}^{T-k+1} \Big\{d_{t+k}(A_t, S_t) - E\Big(d_{t+k}(A_t, S_t)|S_t\Big)\Big\} \times \Big\{U_{t+k}(\alpha_k, \beta_k) - E\big(U_{t+k}(\alpha_k, \beta_k)|S_t\big)\Big\},$$

where $d_{t+k}(A_t, S_t)$ is taken to be the same as Equation (3.11). The advantage of this estimation equation is that the dimension of S_t does not increase with t. Therefore it is possible to use kernel estimations for $E(U_{t+k}|S_t)$ and $E(d(A_t, S_t)|S_t)$ and no model assumptions need to be imposed on $A_t|S_t$ and $Y_{t+k}|S_t$. The above equation can thus be written as:

$$\begin{aligned} 0 = \sum_{i=1}^{n} \sum_{t=1}^{T-k+1} & \begin{pmatrix} A_t^{i\,2} - E(A_t^{i\,2}|S_t^i) \\ \{A_t^i - E(A_t^i|S_t^i)\}g_k(S_t^i) \end{pmatrix} \Big\{Y_{t+k}^i - E(Y_{t+k}^i|S_t^i) \\ & - \begin{pmatrix} A_t^{i\,2} - E(A_t^{i\,2}|S_t^i) \\ \{A_t^i - E(A_t^i|S_t^i)\}g_k(S_t^i) \end{pmatrix}^T \begin{pmatrix} \alpha_k \\ \beta_k \end{pmatrix} \Big\}. \end{aligned}$$

Let $B_t(s) = E(A_t^2|S_t = s)$, $C_t(s) = E(A_t|S_t = s)$, $D_{t,k}(s) = E(Y_{t+k}|S_t = s)$. $B_t(s)$, $C_t(s)$, $D_{t,k}(s)$ are estimated with the kernel estimators: $\hat{B}_t(s) = \sum_{i=1}^{n} A_t^{i\,2} K_\Lambda(s -$

$S_t^i)/\{\sum_{i=1}^n K_\Lambda(s-S_t^i)\}$, $\hat{C}_t(s) = \sum_{i=1}^n A_t^i K_\Lambda(s-S_t^i)/\{\sum_{i=1}^n K_\Lambda(s-S_t^i)\}$, $\hat{D}_{t,k}(s) = \sum_{i=1}^n Y_{t+k}^i K_\Lambda(s-S_t^i)/\{\sum_{i=1}^n K_\Lambda(s-S_t^i)\}$, where $K(\cdot)$ is a multivariate kernel function and $K_\Lambda(u) = |\Lambda|^{-1/2}K(\Lambda^{-1/2}u)$, Λ is a symmetric and positive definite bandwidth matrix. From the estimation equation above, we can derive the estimated parameters:

$$
\begin{aligned}
&\begin{pmatrix}\hat{\alpha}_k\\ \hat{\beta}_k\end{pmatrix} = \left[\sum_{i=1}^n\sum_{t=1}^{T-k+1}\begin{pmatrix}A_t^{i2}-\hat{B}_t(S_t^i)\\ \{A_t^i-\hat{C}_t(S_t^i)\}g_k(S_t^i)\end{pmatrix}^{\otimes 2}\right]^{-1}\\
&\left[\sum_{i=1}^n\sum_{t=1}^{T-k+1}\begin{pmatrix}A_t^{i2}-\hat{B}_t(S_t^i)\\ \{A_t^i-\hat{C}_t(S_t^i)\}g_k(S_t^i)\end{pmatrix}\Big\{Y_{t+k}^i-\hat{D}_{t,k}(S_t^i)\Big\}\right]\\
&=\left[\sum_{i=1}^n\sum_{t=1}^{T-k+1}\begin{pmatrix}A_t^{i2}-\frac{\sum_{j=1}^n A_t^{j\,2}K_\Lambda(S_t^j-S_t^i)}{\sum_{j=1}^n K_\Lambda(S_t^j-S_t^i)}\\ (A_t^i-\frac{\sum_{j=1}^n A_t^j K_\Lambda(S_t^j-S_t^i)}{\sum_{j=1}^n K_\Lambda(S_t^j-S_t^i)})f_k(S_t^i)\end{pmatrix}^{\otimes 2}\right]^{-1}\\
&\left[\sum_{i=1}^n\sum_{t=1}^{T-k+1}\begin{pmatrix}A_t^{i2}-\frac{\sum_{j=1}^n A_t^{j\,2}K_\Lambda(S_t^j-S_t^i)}{\sum_{j=1}^n K_\Lambda(S_t^j-S_t^i)}\\ (A_t^i-\frac{\sum_{j=1}^n A_t^j K_\Lambda(S_t^j-S_t^i)}{\sum_{j=1}^n K_\Lambda(S_t^j-S_t^i)})f_k(S_t^i)\end{pmatrix}\Big(Y_{t+k}^i-\frac{\sum_{j=1}^n Y_{t+k}^j K_\Lambda(S_t^j-S_t^i)}{\sum_{j=1}^n K_\Lambda(S_t^j-S_t^i)}\Big)\right].
\end{aligned}
$$

The estimated $\tilde{\alpha}_K$ and $\tilde{\beta}_K$ can thus be calculated by $\hat{\tilde{\alpha}}_K = \sum_{k=1}^K w_k\hat{\alpha}_k$, $\hat{\tilde{\beta}}_K = \sum_{k=1}^K w_k\hat{\beta}_k$. When $\hat{\tilde{\alpha}}_K < 0$, the optimal treatment regime which maximizes the lag K weighted can thus be estimated by $\hat{\pi}_t^{opt} = -\{\hat{\tilde{\beta}}_K^T g(S_t)\}/2\hat{\tilde{\alpha}}_K$. When $\hat{\tilde{\alpha}} > 0$, the estimated optimal treatment regime would be either 0 or the maximum possible dosage. We can also estimate the parameters for the lag K weighted advantage directly by letting $\tilde{U}_{t+K}(\tilde{\alpha}_K,\tilde{\beta}_K) = \sum_{i=1}^K w_k Y_{t+k} - \tilde{\alpha}_K A_t^2 - \{\tilde{\beta}_K^T g(S_t)\}A_t$ and estimate $\tilde{\alpha}_K, \tilde{\beta}_K$ with:

$$
0 = \sum_{i=1}^n\sum_{t=1}^{T-K+1}\Big\{d_{t+K}^i(A_t,S_t) - E\big(d_{t+K}^i(A_t,S_t)|S_t^i\big)\Big\}\times\Big\{\tilde{U}_{t+K}^i(\tilde{\beta}_K) - E\big(\tilde{U}_{t+K}^i(\tilde{\beta}_K)|S_t\big)\Big\},
$$

where $E\big(\tilde{U}_{t+K}^i(\tilde{\beta}_K)|S_t\big)$ can be estimated similarly with kernel estimation. It is trivial to prove that estimated $\tilde{\alpha}_K$ and $\tilde{\beta}_K$ are the same with these two approaches.

Since in model (3.8), the parameters α_k and β_k are invariant across time, the estimation equation can thus be summed over the time index t. Also notice that the kernel estimation in our method averages over the n observations but not over the time index t. If we include enough information in S_t, then it might be possible to assume that the conditional

distributions $Y_{t+k}|S_t$ and $A_t|S_t$ are invariant across time. Thus we can let:

$$\hat{B}_t(s) = \{\sum_{i=1}^{n}\sum_{t=1}^{T} {A_t^i}^2 K_\Lambda(s - S_t^i)\}/\{\sum_{i=1}^{n}\sum_{t=1}^{T} K_\Lambda(s - S_t^i)\}$$
$$\hat{C}_t(s) = \{\sum_{i=1}^{n}\sum_{t=1}^{T} A_t^i K_\Lambda(s - S_t^i)\}/\{\sum_{i=1}^{n}\sum_{t=1}^{T} K_\Lambda(s - S_t^i)\}$$
$$\hat{D}_{t,k}(s) = \{\sum_{i=1}^{n}\sum_{t=1}^{T} Y_{t+k}^i K_\Lambda(s - S_t^i)\}/\{\sum_{i=1}^{n}\sum_{t=1}^{T} K_\Lambda(s - S_t^i)\}$$

This would be more preferable when we only observe the data of a small number of patients and each patient has a large number of observations over time.

The validity of our estimation equation is mainly based on assumptions (3.8) and (3.12). In other words, we assume that the summary statistics of the past history S_t contains all the information which influences the lag k treatment effect for $k = 1, \ldots, K$ and the dependence between A_t and $Y_{t+k}(\bar{A}_{t-1}, a, A_{t+1}^{a_t=a}, \ldots, A_{t+k-1}^{a_t=a})$. In our simulation study, we will also examine the performance of the model when assumption (3.12) is not valid.

3.5 Theoretical Results

In this section, we derive the consistency and asymptotic normality of the estimated parameters. For simplicity of notation, let $B = \{B_1(S_1), \ldots, B_T(S_T)\}$, $C = \{C_1(S_1), \ldots, C_T(S_T)\}$, $D = \{D_1(S_1), \ldots, D_{T-k+1}(S_T)\}$, and $\hat{B} = \{\hat{B}_1(S_1), \ldots, \hat{B}_T(S_T)\}$, $\hat{C} = \{\hat{C}_1(S_1), \ldots, \hat{C}_T(S_T)\}$, $\hat{D} = \{\hat{D}_1(S_1), \ldots, \hat{D}_{T-k+1}(S_T)\}$ and $H = H_{T+1}$. Then the solution to the estimating equation can be written as:

$$\begin{pmatrix} \hat{\alpha}_k \\ \hat{\beta}_k \end{pmatrix} = \left[\mathbb{P}_n L_1(H; \hat{B}, \hat{C})\right]^{-1} \left[\mathbb{P}_n L_2(H; \hat{B}, \hat{C}, \hat{D})\right],$$

where,

$$L_1(H; B, C) = \sum_{t=1}^{T-k+1} \begin{pmatrix} {A_t}^2 - B_t(S_t) \\ \{A_t - C_t(S_t)\} g_k(S_t) \end{pmatrix}^{\otimes 2},$$
$$L_2(H; B, C, D) = \sum_{t=1}^{T-k+1} \begin{pmatrix} {A_t}^2 - B_t(S_t) \\ \{A_t - C_t(S_t)\} g_k(S_t) \end{pmatrix} \Big\{Y_{t+k} - D_t(S_t)\Big\}.$$

Let $\hat{\phi}_k = (\hat{\alpha}_k, \hat{\beta}_k)^T$, and $\phi_k^* = (\alpha_k^*, \beta_k^*)^T$, where:

$$\begin{pmatrix} \alpha_k^* \\ \beta_k^* \end{pmatrix} = \Big\{ E\big[L_1(H; B, C)\big] \Big\}^{-1} E\Big[L_2(H; B, C, D) \Big].$$

To derive the asymptotic normality of the estimators, we need the following regularity assumptions:

Assumption 3.1. *The marginal density of S_t, f_{S_t}, is uniformly bounded away from 0 for all t:* $\inf_{s \in \mathcal{S}} f_{S_t}(s) > 0$ *. This assumption is to ensure that the kernel estimators $\hat{B}_t(s)$, $\hat{C}_t(s)$, $\hat{D}_t(s)$ do not diverge to infinity because of $\hat{f}_{S_t}(s) = \frac{1}{n} \sum_{i=1}^n K_\Lambda(s - S_t^i)$, which converges in probability to $f_{S_t}(s)$ (see Section 3.9 for the proof), on the denominator.*

Assumption 3.2. *As $\Lambda \to 0$, the kernel function satisfies the following equations:* $\inf_s\{\int_{\mathcal{V}_s} K(v)dv\} = 1 - O(\Lambda^{\frac{1}{2}})$*;* $\sup_s\{\int_{\mathcal{V}_s} vK(v)dv\} = O(1)$*;* $\sup_s\{\int_{\mathcal{V}_s} K^2(v)dv\} = O(1)$*;* $\sup_s\{\int_{\mathcal{V}_s} vK^2(v)dv\} = O(1)$*, where $\mathcal{V}_s = \{v : s - \Lambda^{\frac{1}{2}} v \in \mathcal{S}\}$ for $s \in \mathcal{S}$ and v is a vector with the same number of dimensions as s. The first equation in this assumption is to ensure the unbiasedness of the kernel estimator. When $\mathcal{S} = \mathbb{R}^d$, this assumption is satisfied by most of the commonly used kernel functions. However, when $\mathcal{S}$ is bounded, a kernel function defined on $\mathbb{R}^d$ might fail to satisfy this assumption. The rest three equations are to ensure that the limit distributions of the kernel estimators exist. See Section 3.9 for more details.*

Assumption 3.3. *$E(A_t|S_t = s)$, $E(A_t^2|S_t = s)$, $E(A_t^4|S_t = s)$, $E(Y_{t+k}|S_t = s)$, $E(Y_{t+k}^2|S_t = s)$, $f_{S_t}(s)$ as functions of s are uniformly bounded for $s \in \mathcal{S}$. The first derivatives of these functions are also uniformly bounded. These assumptions are to ensure that the higher order terms of the Taylor expansion of the kernel estimators converge to zero.*

Then we prove the following theorem.

Theorem 3.1. *Under the assumptions (3.8), (3.12) and the regularity assumptions 3.1-3.3, if Λ satisfies $n|\Lambda| \to \infty$ and $\Lambda \to 0$ as $n \to \infty$, then $\sqrt{n}(\hat{\phi}_k - \phi_k^*)$ converges to a normal distribution with mean 0 and variance*

$$E^{-1}\Big\{ H; L_1(B, C) \Big\} \Sigma(H; \phi_k^*, B, C, D) E^{-1}\Big\{ L_1(H; B, C) \Big\},$$

where,

$$\Sigma(H;\phi_k^*,B,C,D) = \\ Var\Big\{\mathbb{P}_n L_1(H;B,C)\phi_k^* - \mathbb{P}_n L_2(H;B,C,D)\Big\}.$$

The variance covariance function above can be estimated consistently with:

$$\mathbb{P}_n^{-1}\Big\{L_1(H;\hat{B},\hat{C})\Big\}\Sigma(H;\phi_k^*,\hat{B},\hat{C},\hat{D})\mathbb{P}_n^{-1}\Big\{L_1(H;\hat{B},\hat{C})]\Big\}.$$

The theorem above provides us with the asymptotic distribution of the estimated parameters $\hat{\alpha}_k$ and $\hat{\beta}_k$. For the proof of the theorem, see Section 3.9.

3.6 Simulation Studies

We evaluate the proposed method using a simulation study. The following generative model simulates an observational study where the treatment at each time point is correlated with past treatments and covariates. For each individual, data $(X_1, A_1, ..., X_{T+1}, Y_{T+1})$ are generated as follows:

$$X_1 \sim \text{Normal}(0, \sigma^2),$$

$$A_1 \sim \text{Uniform}(0, 1);$$

For $t \geq 1$,

$$X_{t+1} \sim \text{Normal}(\eta_1 X_t + \eta_2 A_t, \sigma^2),$$

$$A_{t+1} \sim \text{Normal}(\tau_1 X_{t+1} + \tau_2 A_t, \sigma^2);$$

$$Y_{t+1} = \theta_1 X_t + \theta_2 A_{t-1} - A_t(A_t - \beta_0 - \beta_1 X_t) + \epsilon_{t+1},$$

where $\epsilon_t \sim \text{Normal}(0, \sigma^2)$ and the correlation between ϵ_{t_1} and ϵ_{t_2} for any $t_1, t_2 \in \{2, \ldots, T+1\}$ is $\sigma^{|t_1 - t_2|/2}$. Here we assume that the data is observed starting from $t = 1$ and the dosages have been transformed so that $A_t \in \mathcal{A} = \mathbb{R}$. $(\sigma, \theta_1, \theta_2, \eta_1, \eta_2, \tau_1, \tau_2, \beta_0, \beta_1)$ are predefined constants.

Notice that when $S_t = X_t$ and $\theta_2 = 0$, assumption (3.12) is satisfied (Proofs are provided in Section 3.9.6).Under the simulation setting above, the true value for the lag 1 treatment effect is: $\tau_{t,1}(a, S_t) = -a^2 + (\beta_0 + \beta_1 S_t)a$. We can prove that for $k \geq 2$, the lag k effect under this generative model also follows a quadratic form: $\tau_{t,k}(a, S_t) = \alpha_k a^2 + (\beta_{k,0} a + \beta_{k,1} S_t)a$ (See Section 3.9 for details). In our simulation study,

Table 3.1: Simulation results from 200 replicates for observational studies.

		α_k				$\beta_{k,0}$				$\beta_{k,1}$			
k	n	Bias[1]	SD[1]	SE[1]	CP	Bias[1]	SD[1]	SE[1]	CP	Bias[1]	SD[1]	SE[1]	CP
1	100	0.9	16.5	14.8	93.0	0.9	9.3	9.3	95.0	1.1	39.6	35.0	95.0
	200	-0.4	11.1	10.4	93.5	0.3	6.8	6.3	91.0	-0.3	25.7	24.0	92.0
	400	0.4	7.9	7.4	91.5	-0.1	4.5	4.3	92.0	-0.1	18.3	16.7	93.5
2	100	1.7	31.6	29.0	92.5	-1.0	23.0	22.3	93.0	-3.1	79.7	67.9	92.0
	200	0.2	23.5	20.8	91.5	-0.3	16.5	15.7	93.5	-0.6	54.7	47.6	92.0
	400	-1.8	14.5	14.7	95.5	0.7	11.8	11.1	92.5	-1.2	33.6	33.0	94.5
3	100	2.0	32.2	26.9	88.5	4.1	22.1	21.0	93.5	3.1	74.8	67.3	91.0
	200	-3.3	19.6	18.9	94.5	0.6	15.3	14.6	92.0	5.8	50.5	45.6	91.5
	400	1.0	15.9	13.4	89.5	0.7	10.6	10.2	93.0	-2.2	36.8	31.6	91.0

[1] Note: These columns are in 10^{-3} scale

[2] Note: SD refers to the standard deviation of the estimated parameters from 200 replicates, SE refers to the mean of the estimated standard errors calculated by our covariance function, CP refers to the coverage probability of the 95% confidence intervals calculated using the estimated standard errors.

[3] Note: The worst case Monte Carlo standard error for proportions is 2.3%.

we take $\sigma = 0.5$, $\theta_1 = 0.8$, $\theta_2 = 0$, $\eta_1 = -0.2$, $\eta_2 = 0.2$, $\tau_1 = 1$, $\tau_2 = -0.5$, $\beta_0 = 0$, $\beta_1 = 2$ and $S_t = X_t$. The true parameters for the lag 1, lag 2, lag 3 treatment effects can thus be calculated: $(\alpha_1, \beta_{1,0}, \beta_{1,1}) = (-1, 0, 2)$; $(\alpha_2, \beta_{2,0}, \beta_{2,1}) = (-0.21, 0.16, -0.08)$; $(\alpha_3, \beta_{3,0}, \beta_{3,1}) = (-0.0125, -0.08, -0.03)$. The true parameters for the lag 3 weighted advantage with $w_1 = w_2 = w_3 = 1/3$ are $(\tilde{\alpha}_3, \tilde{\beta}_{3,0}, \tilde{\beta}_{3,1}) = (-0.4075, 0.0267, 0.63)$.

We generate the dataset according to the above setting with $T = 50$ and sample size $n = 100, 200, 400$. Then we use the proposed method to estimate the treatment effects for lag $1, 2$ and 3 when $S_t = X_t$. We use the Gaussian kernel $K_\Lambda(s) = (2\pi)^{-q/2}|\Lambda|^{-1/2}\exp(-s^T\Lambda s/2)$, where $q = 1$ is the dimension of S_t, Λ is a $q \times q$ diagonal matrix with $\Lambda_{j,j} = \lambda_j^2$. We take $\lambda_j = 0.305 \times n^{-1/3}\text{sd}(S_{t,j})$, $j = 1, \ldots, q$. The simulation is replicated for 200 times with each sample size. The estimation results are presented in Table 3.1.

As presented in Table 3.1, the proposed method was able to estimate the parameters with small bias. Both the bias and the standard deviation of the estimated parameters showed a tendency to decrease with the sample size increasing. The standard errors estimated with our covariance function provided a close estimate of the standard deviation. The 95% confidence intervals provided a coverage of the true parameters close to 95% in most cases. However, the estimated standard errors tended to slightly underestimate the standard deviation and thus leading to an under-coverage for the confidence intervals. From the proof of Theorem 3.1 in the Section 3.9, we see that the variance of the

Table 3.2: Estimated Parameters for Lag 3 Weighted Advantage from 200 Replicates

n	$\tilde{\alpha}_3$	$\tilde{\beta}_{3,0}$	$\tilde{\beta}_{3,1}$	$\tilde{\tau}_K(\hat{\pi}_t^{opt}, S_t)$
100	-40.6 (2.1)	2.8 (1.1)	63.0(4.7)	64.7(0.27)
200	-40.9 (1.4)	2.7 (0.8)	63.3(3.2)	64.8(0.13)
400	-40.7 (1.1)	2.7 (0.5)	62.9(2.3)	64.9(0.06)

[1] Note: Columns 2-4 are in 10^{-2} scale; column 5 is in 10^{-3} scale.
[2] Note: The numbers in the parenthesis are the standard deviations.
[3] Note: The last column $\tilde{\tau}_K(\hat{\pi}_t^{opt}, S_t) = \sum_{t=1}^{T} \tilde{\tau}_{t,K}(a = \hat{\pi}_t^{opt}, S_t)/T$.

Table 3.3: Simulation results from 200 replicates when $\theta_2 = -0.1$.

		α_k				$\beta_{k,0}$				$\beta_{k,1}$			
k	n	Bias[1]	SD[1]	SE[1]	CP	Bias[1]	SD[1]	SE[1]	CP	Bias[1]	SD[1]	SE[1]	CP
1	100	1.2	16.6	14.8	92.0	73.5	9.4	9.5	0.0	0.0	36.3	35.3	94.0
	200	-0.4	11.1	10.4	92.0	71.7	7.2	6.3	0.0	0.5	26.0	24.1	92.5
	400	1.5	7.9	7.4	93.0	71.5	4.5	4.3	0.0	-2.1	19.0	16.8	93.5

[1] Note: These columns are in 10^{-3} scale
[2] Note: SD refers to the standard deviation of the estimated parameters from 200 replicates, SE refers to the mean of the estimated standard errors calculated by our covariance function, CP refers to the coverage probability of the 95% confidence intervals calculated using the estimated standard errors.
[3] Note: The worst case Monte Carlo standard error for the non-zero proportions is 1.9%.

Table 3.4: Estimated Parameters for Lag 3 Weighted Advantage from 200 Replicates When $\theta_2 = -0.1$

n	$\tilde{\alpha}_3$	$\tilde{\beta}_{3,0}$	$\tilde{\beta}_{3,1}$	$\tilde{\tau}_K(\hat{\pi}_t^{opt}, S_t)$
100	-40.6 (2.1)	2.9 (1.1)	62.9(4.7)	63.9(0.46)
200	-40.9 (1.4)	2.7 (0.8)	63.3(3.2)	64.0(0.25)
400	-40.7 (1.0)	2.8 (0.5)	62.9(2.3)	64.1(0.17)

[1] Note: Columns 2-4 are in 10^{-2} scale; column 5 is in 10^{-3} scale.
[2] Note: The numbers in the parenthesis are the standard deviations.
[3] Note: The last column $\tilde{\tau}_K(\hat{\pi}_t^{opt}, S_t) = \sum_{t=1}^{T} \tilde{\tau}_{t,K}(a = \hat{\pi}_t^{opt}, S_t)/T$.

estimated parameters consist of two parts, the variance from the estimation equation and the variance from the kernel estimation of the conditional expectation. The latter part of the variance converges to 0 as n goes to infinity and is thus not included in the asymptotic variance formula. However, when the sample size is not large enough, excluding this part of the variance may lead to underestimation of the variance, which was supported by the simulation results.

Table 3.2 presents the average of the estimated parameters for the lag 3 weighted advantage with $w_1 = w_2 = w_3 = 1/3$ from 200 replicates for each sample size. For each replicate, we obtain the estimated optimal treatment regime at time t: $\hat{\pi}_t^{opt} = -\hat{\tilde{\beta}}_K^T S_t/(2\hat{\tilde{\alpha}}_K)$ and calculate the lag 3 weighted advantage of this suggested treatment regime. The lag 3 weighted advantage is calculated on a test dataset with 5000 subjects each with observations from time $t = 1, \ldots, T+3$ (We generated extra two time points for the purpose of evaluating the lag 3 weighted advantage for time $t = T-1$ and T). Table 3.2 presents the mean and standard deviation of the average lag 3 weighted advantage across time, $\sum_{t=1}^{T} \tilde{\tau}_{t,K}(a = \hat{\pi}_t^{opt}, S_t)/T$. The average lag 3 weighted advantage of the true optimal treatment regime $\sum_{t=1}^{T} \tilde{\tau}_{t,K}(a = \pi_t^{opt}, S_t)/T$ is 65.0×10^{-3}. As the result shows, the treatment regimes estimated by the proposed method was close to optimal.

In order to see how the model performs when assumption (3.12) is not satisfied, we take the same parameters except that $\theta_2 = -0.1$. When $S_t = X_t$, assumption (3.12) is not satisfied for $k = 1$ under this setting (see Section 3.9.6 for details). The true parameters for $k = 1$ under this setting are still $(\alpha_1, \beta_{1,0}, \beta_{1,1}) = (-1, 0, 2)$. The result of the simulation for $k = 1$ is presented in Table 3.3. The estimated parameters for $\beta_{1,0}$ were biased using the proposed estimation equation, thus leading to wrong statistical inference of the parameters. Since for $k = 2, 3$, assumption (3.12) is still satisfied, the result remained unbiased. The estimated parameters for $k = 2, 3$ under this setting are presented in Section 3.9.7. In Table 3.3, we also present the lag 3 weighted advantage of the optimal treatment regime estimated by the biased estimation equation, where $w = (1/3, 1/3, 1/3)$. The actual lag 3 weighted advantage of the optimal treatment regime is 64.7×10^{-3}. In this particular setting, the recommended treatment regime was still close to optimal. However, when a different w is taken, it might not be the case. One solution to the bias is to include more information in S_t. Under this specific setting, it is trivial to prove that $Y_{t+k}, A_t | X_t, A_{t-1}$. Therefore, by taking $S_t = (X_t, A_{t-1})$, we could obtain unbiased estimates of the parameters using the same estimation equation. The estimated results are given in Section 3.9.7.

3.7 Type 1 Diabetes Data Analysis

According to American Diabetes Association, there are approximately 1.25 million American adults and children with type 1 diabetes in the United States in year 2015. Large intervention trials have shown that hyperglycemia among diabetes patients could be better managed by a tight glycemic control (Torrent-Fontbona and López 2018). Rapid-reacting insulin are frequently used for diabetes patients before meals to prevent hyperglycemia events. However, the patients under the insulin therapies may be constantly under the risk of hypoglycemia or hyperglycemia due to overdosing or underdosing. Mobile technologies can be used to collect data on physical activity, glucose, and insulin of the patients and thus facilitates the dose adjustments to prevent adverse events (Maahs et al. 2012). However, there is a lack of literature on methodologies for providing dose suggestions which aims at maximizing short-term outcomes based on data collected by mobile applications. We apply our method to the Ohio type 1 diabetes dataset collected by Marling and Bunescu (2018) to estimate the lagged treatment effects of the doses and provide dose suggestions which maximize the weighted advantage.

This dataset contains six patients, each with eight weeks of data, including: continuously monitored blood glucose; insulin dosages, including rapid reacting insulin taken before meals (bolus insulin doses), and long-term insulin infused continuously through out the day (basal insulin doses); sensor-collected physiological measurements including heart rate, body temperature and steps; and self-reported life-events including carbonhydrates intake and exercises. Through exploratory analysis, we found that each patient has distinct patterns in recorded covariates and blood glucose fluctuations. Therefore, we illustrate with the data of one of the patients and assume that the data from each day of this patient are independent from each other. Results for the other patients will be presented in Table 3.8 in Section 3.9. Figure 3.1 provides a visualisation of 7 days of 3 patients recorded data.

There are 54 days of data available. We further take the first 44 days as the training data and the last 10 days as the testing data. Thus $n = 44$ for the training dataset. We summarize the measurements every 30 minutes, resulting in $T = 48$ time intervals each day. For each 30-minute time interval, the covariates we consider include total carbonhydrates intake, planned total carbonhydrates intake in the next time interval, average glucose level, average heart rate, basal insulin level. We denote these covariates as: $X_t = (\text{Carb}_t, \text{Carb-Planned}_t, \text{Glucose}_t, \text{Heartrate}_t, \text{Basal}_t)^T$. Since education of meal planning is typically incorporated as a part of the insulin therapy for diabetes patients (Bantle et al. 2008), we assume that all the carbonhydrates intake within 30 minutes are

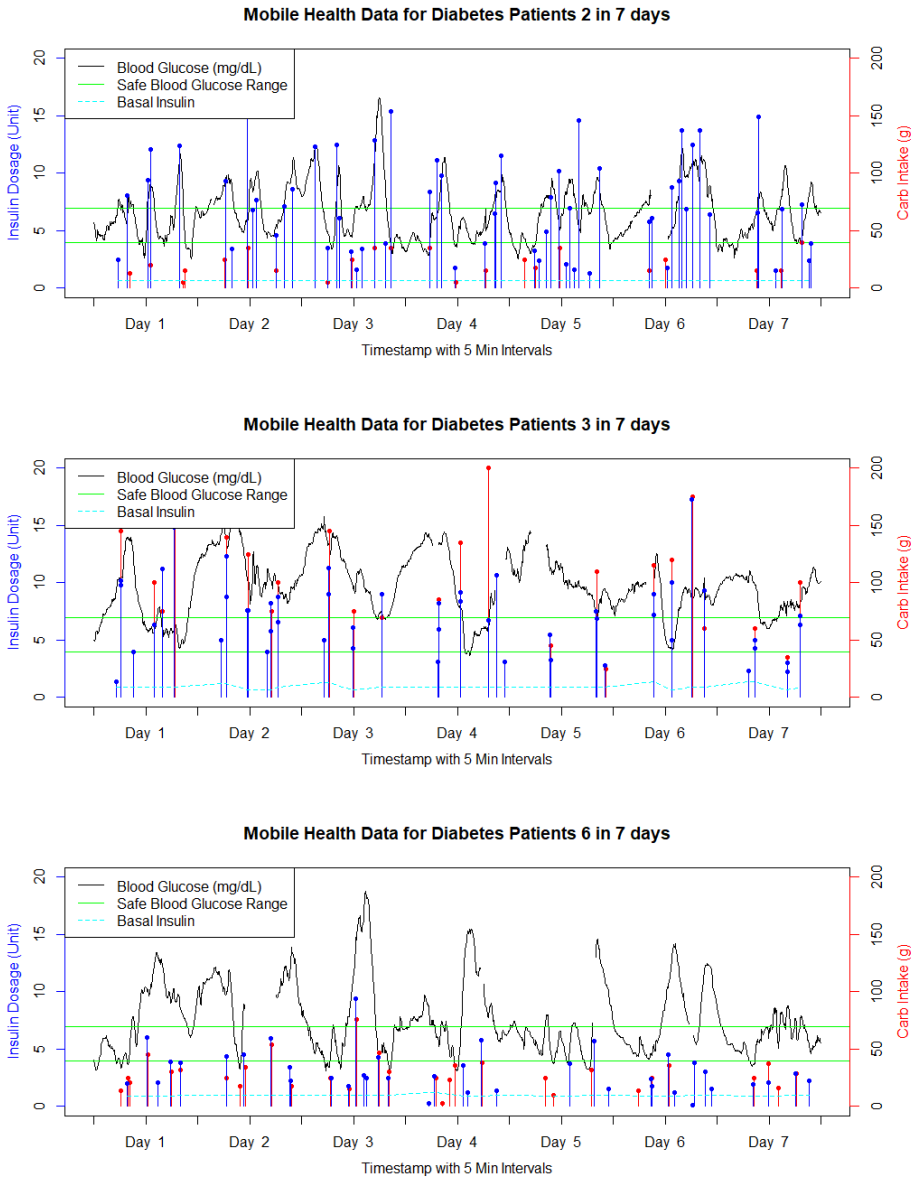

Figure 3.1: Data of 3 Patients for 7 Days

planned ahead of time and the actual carbonhydrates intake is the same as the planned intake: Carb-Planned$_t$ = Carb$_{t+1}$. The treatment A_t is the total bolus injection within the time interval from $t-1$ to t. Let $A_{\max}$ be the maximal observed dose across all days and time. We estimate $\mathcal{A}$ with the interval $[0, A_{\max}]$. The outcome Y_t is taken to be the average of the index of glycemic control (IGC) between time $t-1$ and t calculated by:

$$\text{IGC} = -\frac{I(\text{G} < 80)|80 - \text{G}|^2}{30} - \frac{I(\text{G} > 140)|\text{G} - 140|^{1.35}}{30},$$

where G is the blood glucose level measured in mg/dL (See Rodbard (2009) for various criterias for glycemic control evaluation). Higher Y_t indicates a better glycemic control within the time interval. We take $S_t = (X_t^T,\ \text{Basal-4-8-hour}_t,\ A_{t-1})^T$, where Basal-4-8-hour$_t$ = $\sum_{l=8}^{15}$ Basal$_{t-l}$/8. These covariates are chosen because they are significantly correlated with A_t with exploratory analysis. To support assumption (3.12), it is preferable to include all covariates that are correlated with A_t into S_t. We predict the treatment effect of the dosage within two hours, $k = 1, 2, 3, 4$. For predicting the causal effect, we take $g(S_t) = (\text{Carb}_t, \text{Carb-Planned}_t, \sum_{k=8}^{15} \text{Basal}_{t-k}/8, A_{t-1})$. Thus the model for the lag k causal effect can be written as:

$$\tau_k(a, S_t) =$$
$$\alpha_k a^2 + (\beta_{k,0} + \beta_{k,1}\text{Carb}_t + \beta_{k,2}\text{Carb-Planned}_t + \beta_{k,3}\text{Basal-4-8-hour}_t + \beta_{k,4}A_{t-1})a.$$

When the proposed method is applied, we still use Gaussian kernel. The bandwidth Λ is chosen to be a $q \times q$ diagonal matrix with $\Lambda_{j,j} = \lambda_j^2$ and $\lambda_j = 0.305 \times n^{-1/8}\text{sd}(S_{t,j})$, where $j = 1, ..., q$ and $q = 7$.

Table 3.5: Estimated variables with the Ohio type 1 diabetes dataset

k	1	2	3	4	Weighted
α_k ($\times 10^{-2}$)	-12.7(9.0)	-20.6(14.8)	-13.7(12.8)	-5.2(12.1)	-13.0(11.0)
$\beta_{k,0}$ ($\times 10^{-1}$)	15.8(8.2)	45.6(14.7)	50.0(11.1)	33.8(17.8)	36.9 (10.9)
$\beta_{k,1}$ ($\times 10^{-3}$)	21.8(10.8)	17.4(15.5)	15.3(18.7)	25.3(20.3)	15.2 (13.7)
$\beta_{k,2}$ ($\times 10^{-3}$)	25.1(10.1)	18.8(13.8)	8.8(14.5)	6.5(16.5)	13.8(11.9)
$\beta_{k,3}$ ($\times 10^{-1}$)	-15.6(9.7)	-40.5(14.8)	-47.4(12.5)	-35.2(18.7)	34.9(11.8)
$\beta_{k,4}$ ($\times 10^{-2}$)	-6.1(11.6)	-15.4(19.6)	-8.2(24.9)	-14.3(25.8)	-9.1(18.3)

[1] Note: The numbers in the parenthesis are the estimated standard errors calculated by the covariance formula.

[2] Note: The last column presents the estimated parameters for the lag 4 weighted advantage with $w_1 = w_2 = w_3 = w_4 = 1/4$.

The estimated parameters are presented in Table 3.5. The numbers in the parenthesis are the estimated standard errors for the parameters based on our covariance function. Thus the optimal treatment would be the one which maximizes the expected weighted advantage for the next two hours. The estimated parameters α_k are negative for $k = 1, 2, 3, 4$ resulting in a negative value for the estimated $\tilde{\alpha}_K$. Thus, the optimal treatment regime at time t can be estimated by $\hat{\pi}_t^{opt} = -\{\hat{\tilde{\beta}}_K^T S_t\}/(2\hat{\tilde{\alpha}}_K)$ when $-\{\hat{\tilde{\beta}}_K^T S_t\}/(2\hat{\tilde{\alpha}}_K) \in [0, A_{\max}]$; 0 when $-\{\hat{\tilde{\beta}}_K^T S_t\}/(2\hat{\tilde{\alpha}}_K) < 0$; $A_{\max}$ when $-\{\hat{\tilde{\beta}}_K^T S_t\}/(2\hat{\tilde{\alpha}}_K) > A_{\max}$. Since $\hat{\tilde{\alpha}}_K < 0$, the parameters $\hat{\tilde{\beta}}_{K,j}$ can be interpreted as the units of increase in optimal insulin with $-2\tilde{\alpha}_K$ extra units in the $S_{t,j}$ had the other covariates held constant. The results implied that the optimal dose should increase when the carbonhydrates intake is higher over the last half an hour or in the next half an hour ($\hat{\tilde{\beta}}_{K,1}, \hat{\tilde{\beta}}_{K,2} > 0$); the optimal dose should decrease when the average basal insulin rate 4 to 8 hours ago was higher or the dose in the last half an hour was higher ($\hat{\tilde{\beta}}_{K,3}, \hat{\tilde{\beta}}_{K,4} < 0$). These results are consistent with the fact that carbonhydrates intake increases the blood glucose and past insulin injections lower the blood glucose. The most significant effects estimated in the result were Carb-Planned$_t$ and A_{t-1}. We then estimate the lag K weighted advantage on the test dataset using the estimated parameters $\hat{\tilde{\tau}}_{t,K}(a, S_t) = \hat{\tilde{\alpha}}_K a^2 + \hat{\tilde{\beta}}_K g(S_t)$. The mean of the estimated average lag 4 weighted advantage $\sum_{t=1}^{T} \tilde{\tau}_{t,K}(a, S_t)/T$ is 0.63 for the suggested treatment regime and 0.13 for the original doses. If the model was correct, this method could be used to provide dose suggestions which enhance the stability of the blood glucose for diabetes patients within a short time period.

3.8 Discussion and Conclusion

In this chapter, we defined the lag k treatment effects for continuous treatments following the framework by Boruvka et al. (2018). Nonparametric structural nested models with a quadratic form are used for estimating the causal effects of continuous treatments based on mobile health data. We also defined the weighted lag K advantage to measure the effect of the treatments within a short time period in the future. The optimal treatment regime is defined to be the one which maximizes this advantage. The proposed method provides dose suggestions which maximizes short-term outcome. The semiparametric model is robust against model misspecification. Compared to other infinite horizon methods where stationarity or Markovian property is required, this method imposes minimal assumptions on the data generating process. Statistical inference was also be provided for the estimated parameters. The simulation studies showed that the method was capable of estimating the parameters accurately and the variance could be approximated comparatively well

with our covariance function. The estimated treatment regime was close to maximize the weighted lag K advantage. Application to the Ohio type 1 diabetes dataset showed that this method could provide meaningful insights for dose suggestions based on observed history of the patients. In practice, to ensure the unbiasedness of proposed estimation equation, it is essential to include all confounders which influence both A_t and Y_{t+k} into S_t.

The proposed method is an extension of Boruvka et al. (2018)'s method to continuous dose settings. However, there are several major differences in our works. First, although both works are under the mobile health setting, our work actually addresses a different research question. Boruvka et al. (2018)'s method deals with binary interventions and assumes that the data come from micro-randomized trials where the conditional density of the treatment given the covariate history $p(A_t|H_t)$ is known. The type of treatments considered for these trials are mobile interventions such as sending alerts or reminders. However, micro-randomized trials are not available when the treatment options are actual medications, considering the ethical issue and the infeasibility of enforcing drug usage through mobile apps. Therefore, Boruvka et al. (2018)'s estimation method is not directly applicable in this case. Under our setting, the mobile application functions as strategy recommendation system based on health tracking data rather than an intervention party. Second, although both our estimation equations are conditional on a summary statistics S_t of the full past history H_t, the motivations behind this approach are different. Boruvka et al. (2018) aim to address specific scientific questions by conditioning on subgroups of the subjects, while our interest lies in estimating the exact lag k treatment effect given the full history, but basing the estimation equations only on factors which are influential to the treatment effects and the actual treatment decision process. Conditioning on S_t also allows a time-invariant dynamic treatment regime $\pi_1(\cdot) = \pi_2(\cdot) = \cdots = \pi_T(\cdot)$ which can be applied to data with infinite horizons. The complexity of the suggested optimal dosage would not increase tremendously as T increases since the dimension of S_t is fixed. Notice that assumption (3.12) is a weaker assumption than the stationarity assumption and the Markovian assumption typically used in decision making literature under infinite horizon. Third, we defined the proximal advantage, which is a concept that has not been proposed in previous research works. This definition allows decision making strategies under the mobile health setting where the goal is to optimize an outcome (or minimize the possibility of adverse events) in a short term. The proposed work fills the research gap of methodologies on sequential decision making for the purpose of optimizing short term outcomes.

One drawback of this method is that it is not applicable when the dimension of S_t

gets large due to the nonparametric nature of the kernel estimation. From the real data application, we also noticed that this method tend to suggest none-zero doses for all time points which would not be feasible for patients to do manually in reality. In addition, a slight underestimate of the quadratic effects of the doses may lead to a large overestimate of the optimal dose due to the form of the suggested treatment regime. Possible future work would include extending the method to avoid overestimation in doses. The proposed method is also limited to estimating the causal effect of a one-time change in the treatment history. Estimating the cumulative treatment effects if all future treatments follow the suggested treatment regime would be the direction for our future research.

3.9 Proof and Additional Results

3.9.1 Proof of Equation (3.7)

First, for the first term in equation (3.7), we can derive:

$$
\begin{aligned}
&E\Big[Y_{t+k}(\bar{A}_{t-1}, a, A_{t+1}^{a_t=a}, \ldots, A_{t+k-1}^{a_t=a})|H_t(\bar{A}_{t-1})\Big] \\
&= E\Big[Y_{t+k}(\bar{A}_{t-1}, a, A_{t+1}^{a_t=a}, \ldots, A_{t+k-1}^{a_t=a})|H_t(\bar{A}_{t-1}), A_t = a\Big] \\
&= E\Big[Y_{t+k}(\bar{A}_{t-1}, A_t, A_{t+1}^{a_t=a}, \ldots, A_{t+k-1}^{a_t=a})|H_t(\bar{A}_{t-1}), A_t = a\Big] \\
&= E\Big[E\Big\{Y_{t+k}(\bar{A}_{t-1}, A_t, A_{t+1}^{a_t=a}, \ldots, A_{t+k-1}^{a_t=a})|H_t(\bar{A}_{t-1}), A_t = a, A_{t+1} = A_{t+1}^{a_t=a}, \ldots, \\
&A_{t+k-1} = A_{t+k-1}^{a_t=a}\Big\}|H_t(\bar{A}_{t-1}), A_t = a\Big] \\
&= E\Big[E\Big\{Y_{t+k}|H_t(\bar{A}_{t-1}), A_t = a, A_{t+1} = A_{t+1}^{a_t=a}, \ldots, A_{t+k-1} = A_{t+k-1}^{a_t=a}\Big\}|H_t(\bar{A}_{t-1}), A_t = a\Big] \\
&= E\Big[Y_{t+k}|H_t(\bar{A}_{t-1}), A_t = a\Big] \\
&= E\Big[Y_{t+k}|H_t, A_t = a\Big],
\end{aligned}
$$

where the first equation is based on the sequential ignorability assumption; The second, the third and the fifth equations are based on the property of the conditional expectation; The fourth and the last equations are based on the consistency assumption. Equation (3.7) can thus be proved.

3.9.2 Proof of Equation (3.13)

We need to show that under assumptions (3.8) and (3.12) :

$$
E\left[\Big\{d(A_t, S_t) - E(d(A_t, S_t)|S_t)\Big\} \times \Big\{U_{t+k} - E(U_{t+k}|S_t)\Big\}\right] = 0.
$$

By the property of conditional expectation, it is trivial to obtain that:

$$
\begin{aligned}
&E\Big[\Big\{d(A_t, S_t) - E(d(A_t, S_t)|S_t)\Big\} \times E(U_{t+k}|S_t)\Big] \\
&= E\left(E\left[\Big\{d(A_t, S_t) - E(d(A_t, S_t)|S_t)\Big\} \times E(U_{t+k}|S_t)\middle| S_t\right]\right) \\
&= E\Big[E\Big\{d(A_t, S_t)|S_t\Big\} - E\Big\{d(A_t, S_t)|S_t)\Big\} \times E(U_{t+k}|S_t)\Big] \\
&= 0.
\end{aligned}
$$

Thus Equation (3.13) is equivalent to: $E\Big[\Big\{d(A_t,S_t)-E(d(A_t,S_t)|S_t)\Big\}\times U_{t+k}\Big]=0$. By the property of conditional expectation, it is sufficient to show that:

$$E\Big[\Big\{d(A_t,S_t)-E(d(A_t,S_t)|S_t)\Big\}\times U_{t+k}\Big|S_t\Big]=0,$$

which is equivalent to:

$$E\Big[d(A_t,S_t)U_{t+k}|S_t\Big]=E\Big[d(A_t,S_t)|S_t\Big]E\Big[U_{t+k}|S_t\Big]. \tag{3.14}$$

From the definition of U_{t+k}, we can obtain that:

$$U_{t+k}=Y_{t+k}-\tau_k(A_t,a_0,S_t).$$

With consistency assumption, $Y_{t+k}=Y_{t+k}(\bar{A}_{t-1},a=A_t,A_{t+1}^{a=A_t},\ldots,A_{t+k-1}^{a=A_t})$. Thus,

$$U_{t+k}=Y_{t+k}(\bar{A}_{t-1},a_t=A_t,A_{t+1}^{a_t=A_t},\ldots,A_{t+k-1}^{a_t=A_t})-\tau_k(A_t,a_0,S_t).$$

By the consistency assumption , it is trivial to prove that $S_t(\bar{A}_{t-1})=S_t$

$$E(U_{t+k}|S_t,A_t)=E\Bigg[Y_{t+k}(\bar{A}_{t-1},a_t=A_t,A_{t+1}^{a_t=A_t},\ldots,A_{t+k-1}^{a_t=A_t})-\tau_k(A_t,a_0,S_t)|S_t(\bar{A}_{t-1}),A_t\Bigg]$$

$$=E\Bigg[Y_{t+k}(\bar{A}_{t-1},a_t=A_t,A_{t+1}^{a_t=A_t},\ldots,A_{t+k-1}^{a_t=A_t})-E\Big\{Y_{t+k}(\bar{A}_{t-1},a_t=A_t,A_{t+1}^{a_t=A_t},\ldots,A_{t+k-1}^{a_t=A_t})-$$

$$Y_{t+k}(\bar{A}_{t-1},a_t=a_0,A_{t+1}^{a_t=a_0},\ldots,A_{t+k-1}^{a_t=a_0})\Big|H_t(\bar{A}_{t-1}),A_t\Big\}\Big|S_t(\bar{A}_{t-1}),A_t\Bigg].$$

We first take the conditional expectation with respect to $H_t(\bar{A}_{t-1}),A_t$. Then the first term and the second term are both

$$E\{Y_{t+k}(\bar{A}_{t-1},a_t=A_t,A_{t+1}^{a_t=A_t},\ldots,A_{t+k-1}^{a_t=A_t})|H_t(\bar{A}_{t-1}),A_t\}$$

and can be canceled. Thus the right side of the above equation is equal to:

$$E\Big[E\{Y_{t+k}(\bar{A}_{t-1},a_t=a_0,A_{t+1}^{a_t=a_0},\ldots,A_{t+k-1}^{a_t=a_0})|H_t(\bar{A}_{t-1}),A_t\}|S_t(\bar{A}_{t-1}),A_t\Big]$$

$$=E\Big[Y_{t+k}(\bar{A}_{t-1},a_t=a_0,A_{t+1}^{a_t=a_0},\ldots,A_{t+k-1}^{a_t=a_0})|S_t(\bar{A}_{t-1}),A_t\Big]$$

$$=E\Big[Y_{t+k}(\bar{A}_{t-1},a_t=a_0,A_{t+1}^{a_t=a_0},\ldots,A_{t+k-1}^{a_t=a_0})|S_t(\bar{A}_{t-1})\Big],$$

where the last equation is based on assumption (3.12). Therefore,

$$E(U_{t+k}|S_t, A_t) = E\Big[Y_{t+k}(\bar{A}_{t-1}, a_t = a_0, A_{t+1}^{a_t=a_0}, \ldots, A_{t+k-1}^{a_t=a_0})|S_t(\bar{A}_{t-1})\Big]$$

Take expectation with respect to S_t for both sides, we obtain that:

$$E(U_{t+k}|S_t) = E\Big[Y_{t+k}(\bar{A}_{t-1}, a_t = a_0, A_{t+1}^{a_t=a_0}, \ldots, A_{t+k-1}^{a_t=a_0})|S_t(\bar{A}_{t-1})\Big] = E(U_{t+k}|S_t, A_t).$$

Therefore,

$$\begin{aligned}
&E\Big[d(A_t, S_t)U_{t+k}|S_t\Big] \\
&= E\Big[E\{d(A_t, S_t)U_{t+k}|S_t, A_t\}\Big|S_t\Big] \\
&= E\Big[d(A_t, S_t)E\{U_{t+k}|S_t, A_t\}\Big|S_t\Big] \\
&= E\Big[d(A_t, S_t)E\{U_{t+k}|S_t\}\Big|S_t\Big] \\
&= E\Big\{d(A_t, S_t)|S_t\Big\}E\Big\{U_{t+k}|S_t\Big\}.
\end{aligned}$$

Thus Equation (3.14) can be proved. Therefore, Equation (3.13) is proved.

3.9.3 Proof of Theorem 3.1

First of all,

$$\begin{aligned}
&\sqrt{n}(\hat{\phi}_k - \phi_k^*) \\
=&\left[\mathbb{P}_n L_1(H; \hat{B}, \hat{C})\right]^{-1}\left[\sqrt{n}\mathbb{P}_n\Big\{L_2(H; \hat{B}, \hat{C}, \hat{D}) - L_1(H; \hat{B}, \hat{C})\phi_k^*\Big\}\right].
\end{aligned}$$

The second part on the right side can be written as:

$$\begin{aligned}
&\sqrt{n}\mathbb{P}_n\Big\{L_2(H; \hat{B}, \hat{C}, \hat{D}) - L_1(H; \hat{B}, \hat{C})\phi_k^*\Big\} \\
=&\sqrt{n}\Big[\mathbb{P}_n\{L_2(H; \hat{B}, \hat{C}, \hat{D}) - L_1(H; \hat{B}, \hat{C})\phi_k^*\} - \mathbb{P}_n\{L_2(H; B, C, D) - L_1(H; B, C)\phi_k^*\}\Big] \\
&+ \sqrt{n}\Big[\mathbb{P}_n\{L_2(H; B, C, D) - L_1(H; B, C)\phi_k^*\} - E\{L_2(H; B, C, D) - L_1(H; B, C)\phi_k^*\}\Big].
\end{aligned}$$

Therefore, to prove Theorem 3.1, it is enough to show the following three equations:

$$\mathbb{P}_n L_1(H;\hat{B},\hat{C}) \xrightarrow{p} E[L_1(H;B,C)], \tag{3.15}$$

$$\sqrt{n}\Big[\mathbb{P}_n\{L_2(H;\hat{B},\hat{C},\hat{D}) - L_1(H;\hat{B},\hat{C})\phi_k^*\} - \mathbb{P}_n\{L_2(H;B,C,D) - L_1(H;B,C)\phi_k^*\}\Big] = o_p(1) \tag{3.16}$$

$$\sqrt{n}\Big[\mathbb{P}_n\{L_1(H;B,C)\phi_k^* - L_2(H;B,C,D)\} - E\{L_1(H;B,C)\phi_k^* - L_2(H;B,C,D)\}\Big] \xrightarrow{d} N\Big\{0, \Sigma(\phi_k^*;B,C,D)\Big\}. \tag{3.17}$$

Then with Slutsky's theorem, we can obtain that $\sqrt{n}(\hat{\phi}_k - \phi_k^*)$ converges in distribution to a mean zero normal random vector with variance-covariance matrix given by:

$$E^{-1}\Big\{L_1(H;B,C)\Big\}\Sigma(H;\phi_k^*,B,C,D)E^{-1}\Big\{L_1(H;B,C)\Big\}.$$

Proof of Equation (3.15)

First, we can obtain:

$$\begin{aligned}&\mathbb{P}_n L_1(H;\hat{B},\hat{C}) - E[L_1(H;B,C)]\\ &= \{\mathbb{P}_n L_1(H;\hat{B},\hat{C}) - \mathbb{P}_n L_1(H;B,C)\} + \{\mathbb{P}_n L_1(H;B,C) - E[L_1(H;B,C)]\}\end{aligned}$$

The second part on the right is $o_p(1)$ by the law of large numbers. Therefore, we just need to prove that the first part is $o_p(1)$. With Taylor expansion and the mean value theorem, we can obtain:

$$\begin{aligned}&\Big|\mathbb{P}_n L_1(H;\hat{B},\hat{C}) - \mathbb{P}_n L_1(H;B,C)\Big|\\ &= \Big|\mathbb{P}_n\Big\{\frac{\partial L_1(H;B,C)}{\partial B}\Big|_{B'}(\hat{B}-B) + \frac{\partial L_1(H;B,C)}{\partial C}\Big|_{C'}(\hat{C}-C)\Big\}\Big|\\ &\le \Big|\mathbb{P}_n\Big\{\frac{\partial L_1(H;B,C)}{\partial B}\Big|_{B'}(\hat{B}-B)\Big\}\Big| + \Big|\mathbb{P}_n\Big\{\frac{\partial L_1(H;B,C)}{\partial C}\Big|_{C'}(\hat{C}-C)\Big\}\Big|\end{aligned}$$

for some B' between $\hat{B}$ and B, and C' between $\hat{C}$ and C.

$$\Big|\mathbb{P}_n\Big\{\frac{\partial L_1(H;B,C)}{\partial B}\Big|_{B'}(\hat{B}-B)\Big\}\Big| \le \mathbb{P}_n\Big|\frac{\partial L_1(H;B,C)}{\partial B}\Big|_{B'}\Big|\sup_{s\in\mathcal{S}}\Big|\hat{B}-B\Big| \tag{3.18}$$

Notice that:

$$E\left|\frac{\partial L_1(H;B,C)}{\partial B_t}\Big|_{B'}\right| = E\begin{pmatrix} 2|B'_t - A_t^2| & |C_t - A_t|f_k(S_t) \\ |C_t - A_t|f_k(S_t) & 0 \end{pmatrix}$$

By assumption 3.3, we can obtain that $E\{|C_t - A_t|f_k(S_t)\} < \infty$. Furthermore, $E|B'_t - A_t^2| \leq E|B'_t - B_t| + E|B_t - A_t^2| \leq E|\hat{B}_t - B_t| + E|B_t - A_t^2|$. By assumption 3.3, we can obtain that $E\{B_t - A_t^2\} < \infty$. Thus, If we can prove that:

$$\sup_s |\hat{B}_t(s) - B_t(s)| = o_p(1), \tag{3.19}$$

then $E\left|\frac{\partial L_1(H;B,C)}{\partial B}|_{B'}\right| < \infty$. Since

$$\mathbb{P}_n\left|\frac{\partial L_1(H;B,C)}{\partial B}\Big|_{B'}\right| \xrightarrow{p} E\left|\frac{\partial L_1(H;B,C)}{\partial B_t}\Big|_{B'}\right|,$$

we obtain that:

$$\mathbb{P}_n\left|\frac{\partial L_1(H;B,C)}{\partial B}\Big|_{B'}\right| = O_p(1).$$

Together with Equation (3.19), we obtain that the right side of Equation (3.18) is $o_p(1)$. Similarly, if we can prove that:

$$\sup_s |\hat{C}_t(s) - C_t(s)| = o_p(1), \tag{3.20}$$

and we can obtain:

$$\mathbb{P}_n\left|\frac{\partial L_1(H;B,C)}{\partial C}\Big|_{C'}(\hat{C} - C)\right| = o_p(1).$$

Then we can obtain that:

$$\left|\mathbb{P}_n L_1(H;\hat{B},\hat{C}) - \mathbb{P}_n L_1(H;B,C)\right| = o_p(1).$$

Together with $\mathbb{P}_n L_1(H;B,C) - E[L_1(H;B,C)] \xrightarrow{p} 0$ by the law of large numbers, we can finish the proof for equation (3.15).

Below, we prove equation (3.20). Proof of equation (3.19) can be derived similarly. First, let the density of S_t be f_{S_t} and $\hat{f}_{S_t}(s) = \{\sum_{i=1}^n K_\lambda(s - S_t^i)\}/n$. Write $\hat{C}_t(s)$ as: $\hat{C}_t(s) =$

$\hat{C}_{t,1}(s)/\hat{f}_{S_t}(s)$, where $\hat{C}_{t,1}(s) = \{\sum_{i=1}^{n} A_t^i K_\lambda(s - S_t^i)\}/n$. Also let $C_{t,1}(s) = C_t(s) f_{S_t}(s)$, then:

$$\sup_s |\hat{C}_t(s) - C(s)| = \sup_s |\frac{\hat{C}_{t,1}(s)}{\hat{f}_{S_t}(s)} - \frac{C_{t,1}(s)}{f_{S_t}(s)}| =$$
$$\sup_s \Big| \frac{\big\{\hat{C}_{t,1}(s) - C_{t,1}(s)\big\} f_{S_t}(s) - C_{t,1}(s)\big\{\hat{f}_{S_t}(s) - f_{S_t}(s)\big\}}{\hat{f}_{S_t}(s) f_{S_t}(s)} \Big|$$
$$\leq \sup_s \Big| \frac{\hat{C}_{t,1}(s) - C_{t,1}(s)}{\hat{f}_{S_t}(s)} \Big| + \sup_s \Big| \frac{C_{t,1}(s)\big\{\hat{f}_{S_t}(s) - f_{S_t}(s)\big\}}{\hat{f}_{S_t}(s) f_{S_t}(s)} \Big|.$$

Under the boundedness of $C_{t,1}(s)$ and the assumption that $p_{S_t}(s)$ is uniformly bounded away from 0, it suffices to show that:

$$\sup_s |\hat{C}_{t,1}(s) - C_{t,1}(s)| \to 0 \tag{3.21}$$
$$\sup_s |\hat{f}_{S_t}(s) - f_{S_t}(s)| \to 0 \tag{3.22}$$

We demonstrate the proof for equation (3.21). Equation (3.22) can be proved similarly. First notice:

$$\sup_s |\hat{C}_{t,1}(s) - C_{t,1}(s)| \leq \sup_s |\hat{C}_{t,1}(s) - E\{\hat{C}_{t,1}(s)\}| + \sup_s |E\{C_{t,1}(s)\} - C_{t,1}(s)| \tag{3.23}$$

We prove the uniform convergence of the two parts on the right separately. First, we obtain:

$$E\{\hat{C}_{t,1}(s)\} = E\{A_t K_\Lambda(s - S_t)\}$$
$$= \int_{a_t} \int_{s_t} a_t K_\Lambda(s - s_t) f_{A_t|S_t}(a_t|s_t) f_{S_t}(s_t) ds_t da_t$$
$$= \int_{s_t} C_t(s) |\Lambda^{-1/2}| K\big(\Lambda^{-1/2}(s - s_t)\big) f_{S_t}(s_t) ds_t.$$

Let $v = \Lambda^{-1/2}(s - s_t)$, then $s_t = s - \Lambda^{1/2} v$. Let $\mathcal{V}_s = \{v : s - \Lambda^{-1/2} v \in \mathcal{S}\}$, then the

above equation is equal to:

$$\begin{aligned}
&\int_{\mathcal{V}_s} C_t(s-\Lambda^{1/2}v)K(v)f_{S_t}(s-\Lambda^{1/2}v)dv \\
&= \int_{\mathcal{V}_s} \Big\{C_t(s) - v^T\Lambda^{1/2}\dot{C}_t(s')\Big\}K(v)\Big\{f_{S_t}(s) - v^T\Lambda^{1/2}\dot{f}_{S_t}(s'')\Big\}dv \\
&= C_t(s)f_{S_t}(s)\Big\{\int_{\mathcal{V}_s} K(v)dv\Big\} - \Big\{C_t(s)\dot{f}_{S_t}(s'')^T + f_{S_t}\dot{C}_t(s')^T\Big\}\Lambda^{\frac{1}{2}}\Big\{\int_{\mathcal{V}_s} vK(v)dv\Big\},
\end{aligned}$$

where the first equation above is obtained by Taylor expansion; s' and s'' are vectors on the segment connecting s and S_t; for any function $g(s)$, $\dot{g}(s) = \partial g(s)/\partial s$. From the assumptions, $\Lambda \to 0$ as $n \to \infty$, thus $\inf_s \mathcal{V}_s \to \mathcal{S}$. $\inf_s\{\int_{\mathcal{V}_s} K(v)dv\} = 1 - O(\Lambda^{1/2})$. $\int_{\mathcal{V}_s} vK(v)dv \le \int_{\mathcal{S}} vK(v) = O(1)$. Thus $E\{\hat{C}_{t,1}(s)\} = C_{t,1}(s) + O(\Lambda^{1/2})$. Next, we prove the uniform convergence of the first part of equation (3.23).

$$\begin{aligned}
&\sup_s |\hat{C}_{t,1}(s) - E\{\hat{C}_{t,1}(s)\}| \\
=&\sup_s |\frac{1}{n}\{\sum_{i=1}^{n} A_t^i K_\Lambda(s - S_t^i)\} - E\{A_t^i K_\Lambda(s - S_t^i)\}| \\
=&\sup_s \Big|\int_{s_t} C_t(s_t)|\Lambda^{-\frac{1}{2}}|K\Big(\Lambda^{-\frac{1}{2}}(s-s_t)\Big)d\big\{F_n(s_t) - F(s_t)\big\}\Big|,
\end{aligned}$$

where $F_n(s_t)$ and $F(s_t)$ denote the empirical cumulative distribution and the cumulative distribution of S_t. Then with integration by part, the above equation is less or equal to:

$$\begin{aligned}
&|\Lambda^{-\frac{1}{2}}| \sup_{s,s_t} \Big|C_t(s_t)K\big(\Lambda^{-\frac{1}{2}}(s-s_t)\big)\Big\{F_n(s_t) - F(s_t)\Big\}\Big| \\
&+ \sup_s \Big|\int_{\mathcal{S}} \Big[\big\{F_n(s_t) - F(s_t)\big\}dC_t(s_t)K\big(\Lambda^{-\frac{1}{2}}(s-s_t)\big)\Big]\Big| \\
&\le \xi_1|\Lambda^{-\frac{1}{2}}| \sup_{s_t} |F_n(s_t) - F(s_t)|,
\end{aligned}$$

where ξ is a constant and the last inequality can be derived by the assumption for the boundedness of $C_t(s_t)$ and $K(\cdot)$. By lemma 2.1 of Schuster et al. (1969), we obtain

that:$P_{S_t}\{\sup_{s_t}|F_n(s_t)-F(s_t)|>\epsilon\}\leq \xi_2\exp(-2n\epsilon^2)$. Then:

$$\begin{aligned}
&P(\sup_s|\hat{C}_{t,1}(s)-E\{\hat{C}_{t,1}(s)\}|>\epsilon)\\
&\leq P(\xi_1|\Lambda^{-\frac{1}{2}}|\sup_{s_t}|F_n(s_t)-F(s_t)|>\epsilon)\\
&= P(\sup_{s_t}|F_n(s_t)-F(s_t)|>\frac{\epsilon|\Lambda^{\frac{1}{2}}|}{\xi_1})\\
&\leq \xi_2\exp(-\frac{2n\epsilon^2|\Lambda|}{\xi_1^2})
\end{aligned}$$

Thus, if $2n|\Lambda|\to\infty$ as $n\to\infty$, then the first part of equation (3.23) converges to 0. Equation (3.21) is then proved. With similar proof for equation (3.22), we can obtain formula (3.20). This ends the proof for formula (3.15).

Proof of Equation (3.16)

First we write the left side of the equation as:

$$\begin{aligned}
&\sqrt{n}\mathbb{P}_n\left[\left\{L_2(H;\hat{B},\hat{C},\hat{D})-L_1(H;\hat{B},\hat{C},\hat{D})\phi_k^*\right\}-\left\{L_2(H;B,C,D)-L_1(H;B,C,D)\phi_k^*\right\}\right]\\
&=\sqrt{n}\mathbb{P}_n\Big[\sum_{t=1}^{T-k+1}\{\hat{M}_{t,1}\hat{M}_{t,2}-M_{t,1}M_{t,2}\}\Big]\\
&=\sum_{t=1}^{T-k+1}\sqrt{n}\mathbb{P}_n\Big[\{M_{t,1}(\hat{M}_{t,2}-M_{t,2})+M_{t,2}(\hat{M}_{t,1}-M_{t,1})+(\hat{M}_{t,1}-M_{t,1})(\hat{M}_{t,2}-M_{t,2})\}\Big],
\end{aligned}$$

where,

$$\begin{aligned}
\hat{M}_{t,1}&=\begin{pmatrix}A_t^2-\hat{B}_t(S_t)\\ \{A_t-\hat{C}_t(S_t)\}g_k(S_t)\end{pmatrix}\\
\hat{M}_{t,2}&=Y_{t+k}-\hat{D}_t(S_t)-\begin{pmatrix}A_t^2-\hat{B}_t(S_t)\\ \{A_t-\hat{C}_t(S_t)\}g_k(S_t)\end{pmatrix}^T\phi_k^*\\
M_{t,1}&=\begin{pmatrix}A_t^2-B_t(S_t)\\ \{A_t-C_t(S_t)\}g_k(S_t)\end{pmatrix}\\
M_{t,2}&=Y_{t+k}-D_t(S_t)-\begin{pmatrix}A_t^2-B_t(S_t)\\ \{A_t-C_t(S_t)\}g_k(S_t)\end{pmatrix}^T\phi_k^*.
\end{aligned}$$

Thus, it is sufficient to show that:

$$\sqrt{n}\mathbb{P}_n M_{t,1}(\hat{M}_{t,2} - M_{t,2}) = o_p(1) \tag{3.24}$$
$$\sqrt{n}\mathbb{P}_n M_{t,2}(\hat{M}_{t,1} - M_{t,1}) = o_p(1) \tag{3.25}$$
$$\sqrt{n}\mathbb{P}_n (\hat{M}_{t,1} - M_{t,1})(\hat{M}_{t,2} - M_{t,2}) = o_p(1) \tag{3.26}$$

We first prove equation (3.24). Let $G_{t,1} = A_t^2 - B_t(S_t)$, $G_{t,2} = \{A_t - C_t(S_t)\}f_k(S_t)$, $G_{t,3} = Y_{t+k} - D_t(S_t)$ and $\hat{G}_{t,1} = A_t^2 - \hat{B}_t(S_t)$, $\hat{G}_{t,2} = \{A_t - \hat{C}_t(S_t)\}f_k(S_t)$, $\hat{G}_{t,3} = Y_{t+k} - \hat{D}_t(S_t)$. Then equation (3.24) can be written as:

$$\sqrt{n}\mathbb{P}_n \begin{pmatrix} G_{t,1} \\ G_{t,2} \end{pmatrix} \left\{ \hat{G}_{t,3} - G_{t,3} + \begin{pmatrix} \hat{G}_{t,1} - G_{t,1} \\ \hat{G}_{t,2} - G_{t,2} \end{pmatrix}^T \phi_k^* \right\} = o_p(1)$$

Therefore, it is equivalent to show all the following equations :

$$\begin{aligned}
\sqrt{n}\mathbb{P}_n G_{t,1}\{\hat{G}_{t,3} - G_{t,3}\} &= o_p(1) \\
\sqrt{n}\mathbb{P}_n G_{t,2}\{\hat{G}_{t,3} - G_{t,3}\} &= o_p(1) \\
\sqrt{n}\mathbb{P}_n G_{t,1}\{\hat{G}_{t,1} - G_{t,1}\} &= o_p(1) \\
\sqrt{n}\mathbb{P}_n G_{t,2}\{\hat{G}_{t,2} - G_{t,2}\} &= o_p(1) \\
\sqrt{n}\mathbb{P}_n G_{t,1}\{\hat{G}_{t,2} - G_{t,2}\} &= o_p(1) \\
\sqrt{n}\mathbb{P}_n G_{t,2}\{\hat{G}_{t,1} - G_{t,1}\} &= o_p(1)
\end{aligned} \tag{3.27}$$

We show the proof of the last equation above. The rest of the equations can be proved similarly. First write it as:

$$\begin{aligned}
&\sqrt{n}\mathbb{P}_n G_{t,2}\{\hat{G}_{t,1} - G_{t,1}\} \\
&= -\sqrt{n}\mathbb{P}_n \{A_t - C_t(S_t)\}\{\hat{B}_t(S_t) - B_t(S_t)\}
\end{aligned}$$

Let $\hat{B}_{t,1}(s) = \sum_{j=1}^n A_t^{j\,2} K_\Lambda(S_t^j - s)/n$, $B_{t,1}(s) = B_t(s) f_{S_t}(s)$. Then $\hat{B}_t(s) = \hat{B}_{t,1}(s)/\hat{f}_{S_t}(s)$. If we can obtain that

$$\lim_{n\to\infty} \mathrm{Var}\Big\{ \sqrt{n|\Lambda^{1/2}|}\Big(\hat{B}_{t,1}(s) - B_t(s)\hat{f}_{S_t}(s) \Big) \Big\} < \infty, \tag{3.28}$$

then from appendix B.1 of Zhu et al. (2019), we obtain that: under the assumptions: $\sqrt{n|\Lambda^{1/2}|}\Big(\hat{B}_{t,1}(s) - B_t(s)\hat{f}_{S_t}(s) \Big)$ converge in distribution to a mean 0 normal distribution. Together with equation 3.22 and the assumption that $f_{S_t}(s)$ is bounded away from 0, we

can obtain that,

$$\begin{aligned}
&\sqrt{n|\Lambda|^{\frac{1}{2}}}\{\hat{B}_t(S_t) - B_t(S_t)\} \\
&= \sqrt{n|\Lambda|^{\frac{1}{2}}}\Big\{\frac{\hat{B}_{t,1}(S_t) - B_t(S_t)\hat{f}_{S_t}(S_t)}{f_{S_t}(S_t)}\Big\} + o_p(1) \\
&= \sqrt{n|\Lambda|^{\frac{1}{2}}}\Big\{\frac{1}{n}\sum_{j=1}^{n}\tilde{B}_t^j(S_t)K_\Lambda(S_t^j - S_t)\Big\} + o_p(1)
\end{aligned}$$

where $\tilde{B}_t^j(s) = \{A_t^2 - E(A_t^2|S_t = s)\}/f_{S_t}(s)$.

Then:

$$\begin{aligned}
&\sqrt{n}\mathbb{P}_n G_{t,2}\{\hat{G}_{t,1} - G_{t,1}\} \\
&= -\sqrt{n}\mathbb{P}_n\{A_t - C_t(S_t)\}\{\hat{B}_t(S_t) - B_t(S_t)\} \\
&= -\frac{1}{n\sqrt{|\Lambda^{\frac{1}{2}}|}}\sum_{i=1}^{n}\Big\{A_t^i - C_t(S_t^i)\Big\}\Big[\sqrt{n|\Lambda^{\frac{1}{2}}|}\Big\{\hat{B}_t(S_t^i) - B_t(S_t^i)\}\Big] \\
&= -\frac{1}{n\sqrt{|\Lambda^{\frac{1}{2}}|}}\sum_{i=1}^{n}\Big\{A_t^i - C_t(S_t^i)\Big\}\Big[\sqrt{n|\Lambda^{\frac{1}{2}}|}\Big\{\frac{1}{n}\sum_{j=1}^{n}\tilde{B}_t^j(S_t^i)K_\Lambda(S_t^j - S_t^i)\Big\} + o_p(1)\Big] \\
&= -\frac{\sqrt{n}}{n^2}\sum_{i=1}^{n}\Big\{A_t^i - C_t(S_t^i)\Big\}\Big\{\sum_{j=1}^{n}\tilde{B}_t^j(S_t^i)K_\Lambda(S_t^j - S_t^i)\Big\} + o_p(1) \\
&= -\frac{\sqrt{n}}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\Big\{A_t^i - C_t(S_t^i)\Big\}\tilde{B}_t^j(S_t^i)K_\Lambda(S_t^j - S_t^i) + o_p(1)
\end{aligned}$$

The third equation above is based on $\sqrt{n|\Lambda^{\frac{1}{2}}|} \to \infty$ and $\frac{\sqrt{n}}{n}\sum_{i=1}^{n}(A_t^i - C_t(S_t^i)) \xrightarrow{d} N(0, E\{\mathrm{Var}(A_t|S_t\})$. Thus we just need to prove that the first term is $o_p(1)$. The first term above is a $\sqrt{n}$ times a U-statistic plus an $o_p(1)$ term when written as:

$$\begin{aligned}
&\frac{\sqrt{n}}{n^2}\sum_{j=1}^{n}\sum_{i<j}\Bigg[\tilde{B}_t^j(S_t^i)K_\Lambda(S_t^j - S_t^i)\{A_t^i - C_t(S_t^i)\} + \tilde{B}_t^i(S_t^j)K_\lambda(S_t^j - S_t^i)\Big\{A_t^j - C_t(S_t^j)\Big\}\Bigg] \\
&+ \frac{1}{n}\Bigg[\frac{\sqrt{n}}{n}\sum_{i=1}^{n}\tilde{B}_t^i(S_t^i)\Big\{A_t^i - C_t(S_t^i)\Big\}\Bigg].
\end{aligned}$$

The second term above is $o_p(1)$ because of the law of large numbers. The expectation

of the U-statistics is equal to :

$$\begin{aligned}
&\frac{n-1}{n}E\Big[\tilde{B}_t^j(S_t^i)K_\Lambda(S_t^j - S_t^i)\Big\{A_t^i - C_t(S_t^i)\Big\}\Big] \\
&= \frac{n-1}{n}E\Big[\frac{{A_t^j}^2 - E(A_t^2|S_t = S_t^i)}{f_{S_t}(S_t^i)}K_\Lambda(S_t^j - S_t^i)\Big\{A_t^i - E(A_t|S_t = S_t^i)\Big\}\Big] \\
&= \frac{n-1}{n}E\left(E\Big[\Big\{A_t^i - E(A_t|S_t = S_t^i)\Big\}\Big|S_t^i, A_t^j, S_t^j\Big]\frac{{A_t^j}^2 - E(A_t^2|S_t = S_t^i)}{f_{S_t}(S_t^i)}K_\lambda(S_t^j - S_t^i)\right) \\
&= 0.
\end{aligned}$$

By the properties of U-statistics, the variance of $\sqrt{n}$ times the U-statistics converge to:

$$\mathrm{Var}\Big\{E\Big[\tilde{B}_t^j(S_t^i)K_\Lambda(S_t^i - S_t^j)\{A_t^i - C_t(S_t^i)\} + \tilde{B}_t^i(S_t^j)K_\Lambda(S_t^j - S_t^i)\{A_t^j - C_t(S_t^i)\}\Big|S_t^i, A_t^i\Big]\Big\}.$$

We can obtain:

$$\begin{aligned}
&E\Big[\tilde{B}_t^j(S_t^i)K_\Lambda(S_t^j - S_t^i)\{A_t^i - C_t(S_t^i)\} + \tilde{B}_t^i(S_t^j)K_\lambda(S_t^i - S_t^j)\{A_t^j - C_t(S_t^j)\}\Big|S_t^i, A_t^i\Big] \\
&= E\Big[\tilde{B}_t^j(S_t^i)K_\Lambda(S_t^j - S_t^i)\{A_t^i - C_t(S_t^i)\}|S_t^i, A_t^i\Big] \\
&= E\Big[\{{A_t^j}^2 - B_t(S_t^i)\}K_\Lambda(S_t^j - S_t^i)|S_t^i, A_t^i\Big]\frac{A_t^i - C_t(S_t^i)}{f_{S_t}(S_t^i)}
\end{aligned}$$

From calculation in section 3.9.3, we can obtain that:

$$\sup_s |E\{{A_t^j}^2 K_\Lambda(S_t^j - s)\} - B_t(s)f_{S_t}(s)| = O(|\Lambda^{\frac{1}{2}}|)$$
$$\sup_s |E\{K_\Lambda(S_t^j - s)\} - f_{S_t}(s)| = O(|\Lambda^{\frac{1}{2}}|).$$

Thus,

$$E\Big[\{{A_t^j}^2 - B_t(S_t^i)\}K_\Lambda(S_t^j - S_t^i)|S_t^i, A_t^i\Big] \le \xi_3|\Lambda|^{\frac{1}{2}}$$

for some constant ξ_3. Thus,

$$\begin{aligned}
&\mathrm{Var}\Big\{E\Big[\tilde{B}_t^j(S_t^i)K_\Lambda(S_t^i - S_t^j)\{A_t^i - C_t(S_t^i)\} + \tilde{B}_t^i(S_t^j)K_\Lambda(S_t^j - S_t^i)\{A_t^j - C_t(S_t^i)\}\Big|S_t^i, A_t^i\Big]\Big\} \\
&\le \xi^2|\Lambda|\mathrm{Var}\{\frac{A_t^i - C_t(S_t^i)}{f_{S_t}(S_t^i)}\}
\end{aligned}$$

Then as long as $\mathrm{Var}\{(A_t^i - C_t(S_t^i))/f_{S_t}(S_t^i)\} < \infty$, the variance of the U-statistics converges

to 0. Since we assumed that $f_{S_t}(s)$ is bounded away from 0 and $E(A_t^2|S_t = S_t^i) < \infty$, this conditional can be satisfied. Thus, both the expectation and the variance of the $\sqrt{n}$ times the U-statistics converge to 0, so $\sqrt{n}$ times the U-statistic converges in probability to 0. Thus equation (3.27) can be proved. With similar proof for the other equations above equation (3.27), we can obtain equation (3.24). Equation (3.25) can be proved similarly. Equation (3.26) can be proved with similar calculation. We omit the details here due to the length of the proof. This completes the proof for equation (3.16).

3.9.4 Proof for Equation (3.17)

Equation 3.17 can be simply obtained with the central limit theorem. Thus the proof for theorem is completed.

3.9.5 Form for lag k effect under the simulation setting

The true value for the lag 2 effect is:

$$\begin{aligned}
&E(Y_{t+2}|A_t = a, S_t) - E(Y_{t+2}|A_t = 0, S_t) = \\
&- (\tau_1\eta_2 + \tau_2 - \beta_1\eta_2)(\tau_2 + \tau_1\eta_2)a^2 + \Big\{\theta_1\eta_2 + \theta_2 + \beta_0(\tau_1\eta_2 + \tau_2)\Big\}a \\
&+ \Big\{(\tau_1\eta_2 + \tau_2)(-2\tau_1\eta_1 + \beta_1\eta_1) + \beta_1\tau_1\eta_1\eta_2\Big\}aX_t.
\end{aligned}$$

For $k \geq 3$: If we have:

$$E(Y_{t+k-1}|A_t = a, S_t) = \alpha_{k-1,1}X_t + \alpha_{k-1,2}X_t^2 + \alpha_{k-1,3}A_t^2 + \alpha_{k-1,4}A_t + \alpha_{k-1,5}A_tX_t.$$

Then

$$\begin{aligned}
&E(Y_{t+k}|A_t = a, S_t) \\
&= \alpha_{k-1,1}X_{t+1} + \alpha_{k-1,2}X_{t+1}^2 + \alpha_{k-1,3}A_{t+1}^2 + \alpha_{k-1,4}A_{t+1} + \alpha_{k-1,5}A_{t+1}X_{t+1} \\
&= \alpha_{k-1,1}(\eta_1X_t + \eta_2A_t) + \alpha_{k-1,2}(\eta_1X_t + \eta_2A_t)^2 + \alpha_{k-1,3}\Big\{\tau_1\eta_1X_t + (\tau_1\eta_2 + \tau_2)A_t\Big\}^2 \\
&+ \alpha_{k-1,4}\Big\{\tau_1\eta_1X_t + (\tau_1\eta_2 + \tau_2)A_t\Big\} + \alpha_{k-1,5}\Big\{\tau_1\eta_1X_t + (\tau_1\eta_2 + \tau_2)A_t\Big\}(\eta_1X_t + \eta_2A_t) \\
&= \alpha_{k,1}X_t + \alpha_{k,2}X_t^2 + \alpha_{k,3}A_t^2 + \alpha_{k,4}A_t + \alpha_{k,5}A_tX_t,
\end{aligned}$$

where

$$\alpha_{k,1} = \Big\{\alpha_{k-1,1} + \alpha_{k-1,4}\tau_1\Big\}\eta_1,$$
$$\alpha_{k,2} = \Big\{\alpha_{k-1,2} + \alpha_{k-1,3}\tau_1^2 + \alpha_{k-1,5}\tau_1\Big\}\eta_1^2,$$
$$\alpha_{k,3} = \Big\{\alpha_{k-1,2}\eta_2^2 + \alpha_{k-1,3}(\tau_1\eta_2 + \tau_2)^2 + \alpha_{k-1,5}(\tau_1\eta_2 + \tau_2)\eta_2\Big\},$$
$$\alpha_{k,4} = \Big\{\alpha_{k-1,1}\eta_2 + \alpha_{k-1,4}(\tau_1\eta_2 + \tau_2)\Big\},$$
$$\alpha_{k,5} = \Big\{2\eta_1\eta_2\alpha_{k-1,2} + \alpha_{k-1,3}2\tau_1\eta_1(\tau_1\eta_2 + \tau_2) +$$
$$\alpha_{k-1,5}\big[\tau_1\eta_1\eta_2 + \eta_1(\tau_1\eta_2 + \tau_2)\big]\Big\}.$$

Then lag k effect is:

$$\alpha_{k,3}A_t^2 + \alpha_{k,4}A_t + \alpha_{k,5}A_tX_t.$$

3.9.6 Proof of Assumption (3.12) under the Simulation Setting

According to the data generation model for our simulation setting,

$$Y_{t+1}(\bar{A}_{t-1}, a_t = a) = \theta_1X_t + \theta_2A_{t-1} - a(a - \beta_0 - \beta_1X_t) + \epsilon_{t+1}.$$

$$A_t \sim \text{Normal}(\tau_1X_t + \tau_2A_{t-1}).$$

When $\theta_2 = 0$,

$$Y_{t+1}(\bar{A}_{t-1}, a_t = a) = \theta_1X_t - a(a - \beta_0 - \beta_1X_t) + \epsilon_{t+1}.$$

is independent of A_t given $S_t = X_t$. Thus assumption (3.12) is satisfied for $k = 1$. However, when $\theta_2 \neq 0$, this assumption is not satisfied for $k = 1$.

For $k = 2$, first since $X_{t+1}(\bar{A}_{t-1}, a_t = 2) \sim \text{Normal}(\eta_1X_t + \eta_2A_t, \sigma^2)$, it is trivial to see that:

$$X_{t+1}(\bar{A}_{t-1}, a_t = a) \perp A_t|X_t. \tag{3.29}$$

Since

$$A_{t+1}(\bar{A}_{t-1}, a_t = a) \sim \text{Normal}\Big(\tau_1X_{t+1}(\bar{A}_{t-1}, a_t = a) + \tau_2a_t, \sigma^2\Big),$$

we can obtain that

$$A_{t+1}(\bar{A}_{t-1}, a_t = a) \perp A_t|X_t. \tag{3.30}$$

Table 3.6: Simulation results from 200 replicates when $\theta_2 = -0.1$.

		α_k				$\beta_{k,0}$				$\beta_{k,1}$			
k	n	Bias[1]	SD[1]	SE[1]	CP	Bias[1]	SD[1]	SE[1]	CP	Bias[1]	SD[1]	SE[1]	CP
2	100	1.9	31.4	29.2	91.5	-1.2	23.4	22.4	93.5	-4.7	79.1	68.3	93.0
	200	-1.0	23.8	20.9	91.5	-0.1	16.6	15.8	93.0	3.6	56.3	47.9	90.5
	400	-1.1	14.8	14.8	95.5	1.0	11.9	11.1	93.5	1.0	33.6	33.2	95.0
3	100	2.0	32.2	26.8	88.5	4.2	22.2	21.1	94.0	2.9	75.1	67.1	90.5
	200	-3.1	19.5	18.8	93.0	0.6	15.6	14.7	91.5	5.3	50.8	45.7	92.0
	400	1.1	15.7	13.3	89.5	0.7	10.8	10.3	94.0	-2.5	36.7	31.5	92.0

[1] Note: These columns are in 10^{-3} scale

[2] Note: SD refers to the standard deviation of the estimated parameters from 200 replicates, SE refers to the mean of the estimated standard errors calculated by our covariance function, CP refers to the coverage probability of the 95% confidence intervals calculated using the estimated standard errors.

[3] Note: The worst case Monte Carlo standard error for proportions is 2.3%.

Therefore,

$$
\begin{aligned}
&Y_{t+2}(\bar{A}_{t-1}, a_t = a, A_{t+1}^{a_t=a}) = \\
&\theta_1 X_{t+1}(\bar{A}_{t-1}, a_t = a) + \theta a - A_{t+1}(\bar{A}_{t-1}, a_t = a)\Big\{ A_{t+1}(\bar{A}_{t-1}, a_t = a) - \beta_0 - \\
&\beta_1 X_{t+1}(\bar{A}_{t-1}, a_t = a)\Big\} + \epsilon_{t+2}
\end{aligned}
$$

is independent of A_t given X_t based on Equation (3.29) and (3.30). This is true even when $\theta_2 \neq 0$.

Using induction, we can also prove that for any $k \geq 3$,

$$Y_{t+k}(\bar{A}_{t-1}, a_t = a, A_{t+1}, \ldots, A_{t+k-1}) \perp A_t | X_t.$$

3.9.7 Additional Simulation Results

The true parameters for $k = 2, 3$ when $\theta_2 = -0.1$ are : $(\alpha_2, \beta_{2,0}, \beta_{2,1}) = (-0.21, 0.06, -0.08)$; $(\alpha_3, \beta_{3,0}, \beta_{3,1}) = (-0.0125, -0.05, -0.03)$. Table 3.6 presents the result for the estimated parameters when $S_t = X_t$. As shown in the table, the estimated parameters appeared to be unbiased.

For $k = 1$, we can correct the bias by estimating the parameters with $S_t = (X_t, A_{t-1})$. The model for the lag 1 treatment effect is thus: $\tau_{t,1} = \alpha_1 a^2 + (\beta_{1,0} a + \beta_{1,1} X_t + \beta_{1,2} A_{t-1})a$. The true parameters are: $(\alpha_1, \beta_{1,0}, \beta_{1,1}, \beta_{1,2}) = (-1, 0, 2, 0)$. Since the dimension of S_t has increased, we use the bandwidth $\lambda_j = n^{-1/4}\text{sd}(S_{t,j})$. The estimated parameters are

Table 3.7: Simulation results from 200 replicates when $\theta_2 = -0.1$.

k	n	Parameter	Bias[1]	SD[1]	SE[1]	CP
1	100	α_k	18.3	23.4	21.8	83.5
		$\beta_{k,0}$	11.7	15.9	14.8	82.0
		$\beta_{k,1}$	-20.5	59.3	54.5	91.5
		$\beta_{k,2}$	12.8	32.9	31.2	90.5
	200	α_k	9.2	13.8	15.0	92.5
		$\beta_{k,0}$	6.6	12.2	10.3	85.5
		$\beta_{k,1}$	-5.2	35.5	37.5	97.0
		$\beta_{k,2}$	4.6	21.1	21.3	95.5
	400	α_k	5.3	11.0	10.5	92.0
		$\beta_{k,0}$	3.4	8.1	7.3	88.0
		$\beta_{k,1}$	-2.6	28.4	26.1	93.0
		$\beta_{k,2}$	1.7	14.5	14.9	96.0

[1] Note: These columns are in 10^{-3} scale

[2] Note: SD refers to the standard deviation of the estimated parameters from 200 replicates, SE refers to the mean of the estimated standard errors calculated by our covariance function, CP refers to the coverage probability of the 95% confidence intervals calculated using the estimated standard errors.

[3] Note: The worst case Monte Carlo standard error for proportions is 2.3%.

presented in Table 3.7. From the results we see that the estimated parameters appeared to be unbiased. However, the estimated standard deviation was smaller than the actual standard deviation, leading to lower coverage probability when sample size was small. This implies that when the dimension of covariates increases, the estimated standard error converges slower to the actual standard deviation.

3.9.8 Additional Results for Ohio Type 1 Diabetes Dataset

We applied the proposed method for the Ohio Type 1 Diabetes Dataset. The dataset consists of data from 6 patients. The result for patient 6 has been presented in Section 3.7. In Table 3.8, we present the additional results from the other patients. It is likely that the decision process for insulin dosage is different for each patient. Thus using the same set of S_t for all patients might not be the optimal choice.

Table 3.8: Estimated variables with the Ohio type 1 diabetes dataset

Patient 1					
k	1	2	3	4	Weighted
α_k	17(7.6)	21.8(10.6)	16.6(10.6)	10.9(12.7)	16(9.1)
$\beta_{k,0}$	-172.3(73.1)	-92.3(103.9)	-11(112.4)	-50(116.4)	-63.2(86.3)
$\beta_{k,1}$	1.4(0.7)	3.1(1.4)	3.6(1.6)	3.3(1.7)	3.1(1.3)
$\beta_{k,2}$	-0.8(3.1)	2.1(4.7)	5.9(4.8)	3.9(5.9)	2.9(3.6)
$\beta_{k,3}$	-9.1(51.5)	-104.4(80.4)	-93.7(88.3)	34.9(66.7)	-59.6(56.9)
$\beta_{k,4}$	12.9(8.9)	2.6(19.2)	-11.6(23)	-170(35.2)	-110.4(33.6)
Patient 2					
k	1	2	3	4	Weighted
α_k	0.2(0.7)	0.5(0.8)	0.6(1)	-0.8(1.1)	0.1(0.8)
$\beta_{k,0}$	-8.8(11.1)	14.1(18)	18(22.6)	14.7(24.2)	10.7(17.8)
$\beta_{k,1}$	0.6(0.3)	-0.2(0.2)	0(0.2)	-0.1(0.5)	0.1(0.2)
$\beta_{k,2}$	0.4(0.1)	0.2(0.3)	0.1(0.3)	0.3(0.3)	0.3(0.2)
$\beta_{k,3}$	-2.6(6.8)	-19.1(16.1)	-22.2(22.5)	-5.3(23.8)	-13.1(16.9)
$\beta_{k,4}$	-2.1(2.2)	-3(4)	-4.6(4.5)	-6.3(5.2)	-4.2(4)
Patient 3					
k	1	2	3	4	Weighted
α_k	0.6(1)	-2.2(1.8)	-5(2.6)	-4.6(2.9)	-2.9(1.9)
$\beta_{k,0}$	-14(39.7)	69.2(57.3)	159.8(81.9)	173.5(88.3)	98.8(63.9)
$\beta_{k,1}$	0.1(0.1)	0.4(0.3)	0.6(0.4)	0.7(0.4)	0.4(0.3)
$\beta_{k,2}$	0.1(0.1)	0.2(0.2)	0.4(0.3)	0.3(0.3)	0.3(0.2)
$\beta_{k,3}$	-11.2(46.6)	-78.1(67.3)	-148.4(93.1)	-166(98.1)	-103.3(74.8)
$\beta_{k,4}$	2.3(2.5)	3.4(2.9)	5.9(3.3)	7.5(3.2)	4.8(2.7)
Patient 4					
k	1	2	3	4	Weighted
α_k	-1.2(5.1)	-7.4(5.9)	-6.5(4.8)	-5.8(4.6)	-6.2(4.4)
$\beta_{k,0}$	-129(138.1)	162.5(95.5)	154.2(124.4)	167.4(104.2)	97.8(72.4)
$\beta_{k,1}$	0.2(0.4)	0.1(0.5)	-0.8(0.7)	-0.4(1)	-0.1(0.5)
$\beta_{k,2}$	0.4(0.5)	1.3(0.9)	1.2(0.8)	1.2(0.8)	0.7(0.6)
$\beta_{k,3}$	130.6(169.5)	-173.4(130.2)	-146.4(162.6)	-182.3(143.9)	-96(97)
$\beta_{k,4}$	1.2(3.5)	-8.2(10.1)	-13.9(12.3)	-21.2(9.7)	-10.3(7.7)
Patient 5					
k	1	2	3	4	Weighted
α_k	2.7(2.9)	6.7(4.4)	6.6(4.9)	2.7(5.2)	3.9(3.7)
$\beta_{k,0}$	-271(110.8)	-514.8(139.1)	-321(160.8)	-110.6(151.9)	-303.4(127.8)
$\beta_{k,1}$	0.4(0.3)	1.2(0.7)	0.1(0.7)	-0.6(0.8)	0.3(0.6)
$\beta_{k,2}$	0.6(0.3)	0.5(0.6)	0.5(0.7)	0.8(0.7)	0.6(0.5)
$\beta_{k,3}$	173.1(87.1)	324.2(119.8)	201.9(134.1)	78.7(125.6)	198.1(103.7)
$\beta_{k,4}$	28.8(13)	12.4(10.8)	-5.1(6)	-1.6(13.2)	8.4(7.8)

[1] Note: These columns are in 10^{-2} scale .

[2] Note: The numbers in the parenthesis are the estimated standard errors calculated by the covariance formula.

[3] Note: The last column presents the estimated parameters for the lag 4 weighted advantage with $w_1 = w_2 = w_3 = w_4 = 1/4$.

REFERENCES

Arnhold, M., Quade, M., and Kirch, W. (2014). Mobile applications for diabetics: a systematic review and expert-based usability evaluation considering the special requirements of diabetes patients age 50 years or older. *Journal of medical Internet research*, 16(4):e104.

Bantle, J. P., Wylie-Rosett, J., Albright, A. L., Apovian, C. M., Clark, N. G., Franz, M. J., Hoogwerf, B. J., Lichtenstein, A. H., Mayer-Davis, E., Mooradian, A. D., et al. (2008). Nutrition recommendations and interventions for diabetes: a position statement of the american diabetes association. *Diabetes care*, 31:S61–S78.

Bisio, I., Lavagetto, F., Marchese, M., and Sciarrone, A. (2015). A smartphone-centric platform for remote health monitoring of heart failure. *International Journal of Communication Systems*, 28(11):1753–1771.

Blatt, D., Murphy, S. A., and Zhu, J. (2004). A-learning for approximate planning. *Ann Arbor*, 1001:48109–2122.

Boruvka, A., Almirall, D., Witkiewitz, K., and Murphy, S. A. (2018). Assessing time-varying causal effect moderation in mobile health. *Journal of the American Statistical Association*, 113(523):1112–1121.

Breiman, F. (1984). Olshen, and stone. *Classification and Regression trees*.

Cao, W., Tsiatis, A. A., and Davidian, M. (2009). Improving efficiency and robustness of the doubly robust estimator for a population mean with incomplete data. *Biometrika*, 96(3):723–734.

Carrà, G., Crocamo, C., Bartoli, F., Carretta, D., Schivalocchi, A., Bebbington, P. E., and Clerici, M. (2016). Impact of a mobile e-health intervention on binge drinking in young people: The digital–alcohol risk alertness notifying network for adolescents and young adults project. *Journal of Adolescent Health*, 58(5):520–526.

Chakraborty, B. and Moodie, E. (2013). *Statistical methods for dynamic treatment regimes*. Springer.

Chen, G., Zeng, D., and Kosorok, M. R. (2016). Personalized dose finding using outcome weighted learning. *Journal of the American Statistical Association*, 111(516):1509–1521.

Chen, J., Fu, H., He, X., Kosorok, M. R., and Liu, Y. (2018). Estimating individualized treatment rules for ordinal treatments. *Biometrics*, 74(3):924–933.

Chevret, S. (2006). *Statistical methods for dose-finding experiments*, volume 24. Wiley Online Library.

Consortium, I. W. P. (2009). Estimation of the warfarin dose with clinical and pharmacogenetic data. *New England Journal of Medicine*, 360(8):753–764.

Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20(3):273–297.

Dempsey, W., Liao, P., Klasnja, P., Nahum-Shani, I., and Murphy, S. A. (2015). Randomised trials for the fitbit generation. *Significance*, 12(6):20–23.

Duong, T. and Hazelton, M. L. (2005). Cross-validation bandwidth matrices for multivariate kernel density estimation. *Scandinavian Journal of Statistics*, 32(3):485–506.

Ertefaie, A. and Strawderman, R. L. (2018). Constructing dynamic treatment regimes over indefinite time horizons. *Biometrika*, 105(4):963–977.

Evans, W. D., Wallace, J. L., and Snider, J. (2012). Pilot evaluation of the text4baby mobile health program. *BMC public health*, 12(1):1031.

Gustafson, D. H., Boyle, M. G., Shaw, B. R., Isham, A., McTavish, F., Richards, S., Schubert, C., Levy, M., and Johnson, K. (2011). An e-health solution for people with alcohol problems. *Alcohol Research & Health*.

Haller, M. J., Stalvey, M. S., and Silverstein, J. H. (2004). Predictors of control of diabetes: monitoring may be the key. *The Journal of pediatrics*, 144(5):660–661.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). Unsupervised learning. In *The elements of statistical learning*, pages 485–585. Springer.

Henderson, R., Ansell, P., and Alshibani, D. (2010). Regret-regression for optimal dynamic treatment regimes. *Biometrics*, 66(4):1192–1201.

Heron, K. E. and Smyth, J. M. (2010). Ecological momentary interventions: incorporating mobile technology into psychosocial and health behaviour treatments. *British journal of health psychology*, 15(1):1–39.

Hood, M., Wilson, R., Corsica, J., Bradley, L., Chirinos, D., and Vivo, A. (2016). What do we know about mobile applications for diabetes self-management? a review of reviews. *Journal of behavioral medicine*, 39(6):981–994.

Horowitz, J. L. (2001). The bootstrap. In *Handbook of econometrics*, volume 5, pages 3159–3228. Elsevier.

Huckvale, K., Adomaviciute, S., Prieto, J. T., Leow, M. K.-S., and Car, J. (2015). Smartphone apps for calculating insulin dose: a systematic assessment. *BMC medicine*, 13(1):106.

Johnson, J. A., Gong, L., Whirl-Carrillo, M., Gage, B. F., Scott, S. A., Stein, C., Anderson, J., Kimmel, S. E., Lee, M. T. M., Pirmohamed, M., et al. (2011). Clinical pharmacogenetics implementation consortium guidelines for cyp2c9 and vkorc1 genotypes and warfarin dosing. *Clinical Pharmacology & Therapeutics*, 90(4):625–629.

Kirwan, M., Vandelanotte, C., Fenning, A., and Duncan, M. J. (2013). Diabetes self-management smartphone application for adults with type 1 diabetes: randomized controlled trial. *Journal of medical Internet research*, 15(11):e235.

Klasnja, P., Hekler, E. B., Shiffman, S., Boruvka, A., Almirall, D., Tewari, A., and Murphy, S. A. (2015). Microrandomized trials: An experimental design for developing just-in-time adaptive interventions. *Health Psychology*, 34(S):1220.

Kosorok, M. R. (2008). *Introduction to empirical processes and semiparametric inference.* Springer.

Kosorok, M. R. and Moodie, E. E. (2015). *Adaptive TreatmentStrategies in Practice: Planning Trials and Analyzing Data for Personalized Medicine*, volume 21. SIAM.

Laber, E. and Zhao, Y. (2015). Tree-based methods for individualized treatment regimes. *Biometrika*, 102(3):501–514.

Larson, R. R. (2010). Introduction to information retrieval.

Le Thi Hoai, A. and Tao, P. D. (1997). Solving a class of linearly constrained indefinite quadratic problems by dc algorithms. *Journal of Global Optimization*, 11(3):253–285.

Liang, K.-Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73(1):13–22.

Liang, K.-Y., Zeger, S. L., and Qaqish, B. (1992). Multivariate regression analyses for categorical data. *Journal of the Royal Statistical Society: Series B (Methodological)*, 54(1):3–24.

Liao, P., Klasnja, P., Tewari, A., and Murphy, S. A. (2015). Micro-randomized trials in mhealth. *arXiv preprint arXiv:1504.00238*.

Liao, P., Klasnja, P., Tewari, A., and Murphy, S. A. (2016). Sample size calculations for micro-randomized trials in mhealth. *Statistics in medicine*, 35(12):1944–1971.

Lopes, I. M., Silva, B. M., Rodrigues, J. J., Lloret, J., and Proença, M. L. (2011). A mobile health monitoring solution for weight control. In *2011 International Conference on Wireless Communications and Signal Processing (WCSP)*, pages 1–5. IEEE.

Luckett, D. J., Laber, E. B., Kahkoska, A. R., Maahs, D. M., Mayer-Davis, E., and Kosorok, M. R. (2019). Estimating dynamic treatment regimes in mobile health using v-learning. *Journal of the American Statistical Association*, pages 1–34.

Maahs, D. M., Mayer-Davis, E., Bishop, F. K., Wang, L., Mangan, M., and McMurray, R. G. (2012). Outpatient assessment of determinants of glucose excursions in adolescents with type 1 diabetes: proof of concept. *Diabetes technology & therapeutics*, 14(8):658–664.

Maei, H. R., Szepesvári, C., Bhatnagar, S., and Sutton, R. S. (2010). Toward off-policy learning control with function approximation. In *ICML*, pages 719–726.

Marling, C. and Bunescu, R. C. (2018). The ohiot1dm dataset for blood glucose level prediction. In *KHD@ IJCAI*, pages 60–63.

Martínez-Pérez, B., De La Torre-Díez, I., and López-Coronado, M. (2013a). Mobile health applications for the most prevalent conditions by the world health organization: review and analysis. *Journal of medical Internet research*, 15(6):e120.

Martínez-Pérez, B., De La Torre-Díez, I., López-Coronado, M., and Herreros-González, J. (2013b). Mobile apps in cardiology. *JMIR mHealth and uHealth*, 1(2):e15.

Moodie, E. E., Platt, R. W., and Kramer, M. S. (2009). Estimating response-maximized decision rules with applications to breastfeeding. *Journal of the American Statistical Association*, 104(485):155–165.

Moodie, E. E., Richardson, T. S., and Stephens, D. A. (2007). Demystifying optimal dynamic treatment regimes. *Biometrics*, 63(2):447–455.

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

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

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

Puterman, M. L. (2014). *Markov Decision Processes.: Discrete Stochastic Dynamic Programming*. John Wiley & Sons.

Qian, M. and Murphy, S. A. (2011). Performance guarantees for individualized treatment rules. *Annals of statistics*, 39(2):1180.

Rich, B., Moodie, E. E., and Stephens, D. A. (2014). Simulating sequential multiple assignment randomized trials to generate optimal personalized warfarin dosing strategies. *Clinical Trials*, 11(4):435–444.

Robins, J. M. (1986). A new approach to causal inference in mortality studies with a sustained exposure periodapplication to control of the healthy worker survivor effect. *Mathematical modelling*, 7(9-12):1393–1512.

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

Robins, J. M. (1999). Association, causation, and marginal structural models. *Synthese*, 121(1):151–179.

Robins, J. M. (2004). Optimal structural nested models for optimal sequential decisions. In *Proceedings of the second seattle Symposium in Biostatistics*, pages 189–326. Springer.

Robins, J. M., Hernan, M. A., and Brumback, B. (2000). Marginal structural models and causal inference in epidemiology.

Rodbard, D. (2009). Interpretation of continuous glucose monitoring data: glycemic variability and quality of glycemic control. *Diabetes technology & therapeutics*, 11(S1):S–55.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology*, 66(5):688.

Rubin, D. B. (1978). Bayesian inference for causal effects: The role of randomization. *The Annals of statistics*, pages 34–58.

Schafer, J. L. (2006). Marginal modeling of intensive longitudinal data by generalized estimating equations. *Models for Intensive Longitudinal Data. Walls TA, Schafer JL (Eds). Oxford University Press, New York*, pages 38–62.

Schulte, P. J., Tsiatis, A. A., Laber, E. B., and Davidian, M. (2014). Q-and a-learning methods for estimating optimal dynamic treatment regimes. *Statistical science: a review journal of the Institute of Mathematical Statistics*, 29(4):640.

Schuster, E. F. et al. (1969). Estimation of a probability density function and its derivatives. *The Annals of Mathematical Statistics*, 40(4):1187–1195.

Sieverdes, J. C., Treiber, F., Jenkins, C., and Hermayer, K. (2013). Improving diabetes management with mobile health technology. *The American journal of the medical sciences*, 345(4):289–295.

Silva, B. M., Rodrigues, J. J., de la Torre Díez, I., López-Coronado, M., and Saleem, K. (2015). Mobile-health: A review of current state in 2015. *Journal of biomedical informatics*, 56:265–272.

Smola, A. J. and Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and computing*, 14(3):199–222.

Sutton, R. S., Barto, A. G., et al. (1998). *Introduction to reinforcement learning*, volume 2. MIT press Cambridge.

Tang, Y. and Kosorok, M. R. (2012). Developing adaptive personalized therapy for cystic fibrosis using reinforcement learning.

Torrent-Fontbona, F. and López, B. (2018). Personalized adaptive cbr bolus recommender system for type 1 diabetes. *IEEE journal of biomedical and health informatics*, 23(1):387–394.

Vapnik, V. (2013). *The nature of statistical learning theory*. Springer science & business media.

Wallace, M. P. and Moodie, E. E. (2015). Doubly-robust dynamic treatment regimen estimation via weighted least squares. *Biometrics*, 71(3):636–644.

Watkins, C. J. and Dayan, P. (1992). Q-learning. *Machine learning*, 8(3-4):279–292.

Zhang, B., Tsiatis, A. A., Davidian, M., Zhang, M., and Laber, E. (2012a). Estimating optimal treatment regimes from a classification perspective. *Stat*, 1(1):103–114.

Zhang, B., Tsiatis, A. A., Laber, E. B., and Davidian, M. (2012b). A robust method for estimating optimal treatment regimes. *Biometrics*, 68(4):1010–1018.

Zhao, L. P. and Prentice, R. L. (1990). Correlated binary regression using a quadratic exponential model. *Biometrika*, 77(3):642–648.

Zhao, Y., Kosorok, M. R., and Zeng, D. (2009). Reinforcement learning design for cancer clinical trials. *Statistics in medicine*, 28(26):3294–3315.

Zhao, Y., Zeng, D., Rush, A. J., and Kosorok, M. R. (2012). Estimating individualized treatment rules using outcome weighted learning. *Journal of the American Statistical Association*, 107(499):1106–1118.

Zhao, Y.-Q., Zeng, D., Laber, E. B., and Kosorok, M. R. (2015). New statistical learning methods for estimating optimal dynamic treatment regimes. *Journal of the American Statistical Association*, 110(510):583–598.

Zhu, F., Bosch, M., Woo, I., Kim, S., Boushey, C. J., Ebert, D. S., and Delp, E. J. (2010). The use of mobile devices in aiding dietary assessment and evaluation. *IEEE journal of selected topics in signal processing*, 4(4):756–766.

Zhu, L., Lu, W., Kosorok, M. R., and Song, R. (2019). Kernel assisted learning for personalized dose finding. *Manuscript sumbitted for publication*.

Zisser, H., Robinson, L., Bevier, W., Dassau, E., Ellingsen, C., Doyle III, F. J., and Jovanovic, L. (2008). Bolus calculator: a review of four smart insulin pumps. *Diabetes technology & therapeutics*, 10(6):441–444.