Causal Effect Estimation and Optimal Dose Suggestions in Mobile Health International Conference on Machine Learning (ICML 2020)

Liangyu Zhu [1] **Wenbin Lu** [1] **Rui Song** [1]

Abstract

In this article, we propose novel structural nested models to estimate causal effects of continuous treatments based on mobile health data. To find the treatment regime that optimizes the expected short-term outcomes for patients, we define a weighted lag-K advantage as the value function. The optimal treatment regime is then defined to be the one that maximizes the value function. Our method imposes minimal assumptions on the data generating process. Statistical inference is provided for the estimated parameters. Simulation studies and an application to the Ohio type 1 diabetes dataset show that our method could provide meaningful insights for dose suggestions with mobile health data.

1. Introduction

There is a rapid-increasing interest in healthcare interventions using mobile apps. Mobile technologies allow physical conditions of patients to be collected in real time, measured by sensors or self-reported by patients. Studies have shown that mobile health interventions could be beneficial for the healthcare delivery process by improving disease management, enhancing communication with the healthcare provider and providing more precise and individualized medication (Free et al., 2013). However, analyzing mobile health data can be challenging because they typically have a large number of time points, time-varying treatments, and non-definite time horizon (Luckett et al., 2019).

One focus in analyzing mobile health data is to evaluate causal effects 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 & Zeger, 1986; Zhao & 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 & Moodie, 2015; Chakraborty & Moodie, 2013). A dynamic treatment regime is a set of decision rules that decide treatments to be assigned to patients according to their time-varying measurements during the ongoing treatment process. An optimal DTR is the one that yields the most favorable 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 & 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 & 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 hori-

*Equal contribution [1]Department of Statistics, North Carolina State University, Raleigh, NC, USA. Correspondence to: Liangyu Zhu <lzhu12@ncsu.edu>.

Proceedings of the 37^{th} International Conference on Machine Learning, Online, PMLR 119, 2020.

zons. Sequential decision making process in infinite horizon can be modeled as a Markov decision process (Puterman, 2014). Ertefaie & Strawderman (2018) defined the optimal DTR in the infinite horizon as the one which maximizes the expected cumulative discounted reward (the beneficial outcome). The optimal DTR is estimated first by positing a parametric model 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 treatments.

There is increasing attention in how mobile interventions can help managing diseases by monitoring physical conditions against high-risk events and providing frequent treatment adjustments (Maahs et al., 2012; Levine et al., 2001). Our research is motivated by the use of mobile health applications for diseases like diabetes and hypertension, where the main focus is to monitor adverse events in the near future (Haller et al., 2004; Heron & Smyth, 2010). For example, for diabetes patients, the main interest of using rapid-reacting insulin is to maintain a safe blood glucose level within 2 hours after a meal. Existing methodologies in reinforcement learning mainly aims at maximizing a discounted cumulative reward, which might not be the optimal criteria for treatment suggestions in this scenario. Furthermore, the treatments in this case are continuous and thus have infinite number of possible values. Studies on estimating causal effects and providing treatment suggestions under this setting are still absent.

In this article, we aim to find the treatment regime which optimizes the outcomes (or minimizes the risk of adverse events) within a time period of near future. We first extend Boruvka et al. (2018)'s definition of lagged treatment effect to continuous treatments and propose novel structural nested models for estimating causal effects of continuous treatments based on mobile health data. We then define a weighted advantage function. The optimal treatment regime at a specific time point is defined to be the one which optimizes the weighted advantage function. The rest of the article is structured as follows. In section 2, we formalize the problem in a statistical framework and present the proposed methodology for finding the optimal DTR. In section 3, we discuss the theoretical results of the proposed estimators. Simulations are conducted and the corresponding results are presented in section 4. In section 5, we apply the proposed method to the Ohio type 1 diabetes dataset. Discussions and conclusions are given in section 6.

2. Method

2.1. Notation

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. $X_t \in \mathbb{R}^p$ 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 $\{(X_1, A_1),(Y_2, X_2, A_2), \ldots, (Y_T, X_T, A_T), (Y_{T+1}, X_{T+1})\}$. In this article, we use capitalized letters to denote random variables and lowercase letters to denote realized values. Let the overbar denotes the history of a random variable. For example, $\bar{X}_t = (X_1, ..., X_t)$. All information accrued up to time t can be represented by $H_t = (\bar{X}_t, \bar{Y}_t, \bar{A}_{t-1})$. In the considered type-1 diabetes study, A_t is the rapid-reacting insulin dose taken at time t. Y_t measures the stability of the blood glucose between time $t-1$ and t, and X_t includes the food intake, exercise and blood glucose levels.

To define the treatment effects, we adopt the potential outcome framework by Rubin (1974). $X_t(\bar{a}_{t-1})$ 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 $\pi_t = \{p_{\pi,t}(a|h_t), a \in \mathcal{A}\}$; $p_{\pi,t}$ 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 $p_{\pi,t}(a|h_t) = \delta(a = g_t(h_t))$, 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 $\pi_t = g_t(h_t)$.

2.2. Lag k Treatment Effect

Define the conditional lag k ($k \geq 1$) treatment effect of treatment a with respect to a reference treatment a_0 at time t as:

$$\tau_{t,k}\big(a, a_0, H_t(\bar{A}_{t-1})\big) = E\big\{Y_{t+k}(\bar{A}_{t-1}, a, A_{t+1}^{a_t=a}, \ldots, A_{t+k-1}^{a_t=a}) - Y_{t+k}(\bar{A}_{t-1}, a_0, A_{t+1}^{a_t=a_0}, \ldots, A_{t+k-1}^{a_t=a_0}) | H_t(\bar{A}_{t-1})\big\}. \quad (1)$$

$A_{t+1}^{a_t=a}$ 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 $l = 2, \ldots, k-1$. The expectation in Equation (1) is taken over all the possible future treatments from time t to $t+k-1$. Notice that the treatment effect defined in (1) measures the effect of a one-time change in the decision strategy. The causal effect measuring a single-time decision change has been extensively used in various models for intensively collected longitudinal data (Schafer, 2006; Schwartz & Stone, 2007; Bolger & Laurenceau, 2013). Boruvka et al. (2018) also used a similar definition for estimating the effect of mobile application notifications.

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

1. Consistency: The potential outcomes had the treatments given to the patient equal to the observed treatment history are equal to the observed data. More specifically, for $\bar{a}_{t-1} = \bar{A}_{t-1}$, $Y_t(\bar{a}_{t-1}) = Y_t$, $X_t(\bar{a}_{t-1}) = X_t$ and $A_t(\bar{a}_{t-1}) = A_t$ for $2 \leq t \leq T$; 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 $\bar{a}_{t-1} = \bar{A}_{t-1}$.

2. Positivity: All treatments $a \in \mathcal{A}$ can possibly be observed given h_t for any $h_t \in \mathcal{H}_t$. More specifically, for $a \in \mathcal{A}$ and $h_t \in \mathcal{H}_t$, $p_{\pi,t}(a|h_t) > 0$, where $p_{\pi,t}(a|h_t)$ denotes the conditional density for the treatment A_t given the history $H_t = h_t$.

3. Sequential ignorability: The potentials outcomes $\{Y_{t+1}(\bar{a}_t), X_{t+1}(\bar{a}_t), A_{t+1}(\bar{a}_t), ..., Y_{T+1}(\bar{a}_T)\}$ are independent of A_t conditional on H_t, for $t \leq T$. This assumption 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.

Under these three assumptions, we can estimate the conditional lag k treatment effect with the observed data for any $a \in \mathcal{A}$ (see appendix for the proof):

$$E\big\{Y_{t+k}(\bar{A}_{t-1}, a, A_{t+1}^{a_t=a}, \ldots, A_{t+k-1}^{a_t=a}) - Y_{t+k}(\bar{A}_{t-1}, a_0, A_{t+1}^{a_t=a_0}, \ldots, A_{t+k-1}^{a_t=a_0}) | H_t(\bar{A}_{t-1})\big\} = E(Y_{t+k}|A_t = a, H_t) - E(Y_{t+k}|A_t = a_0, H_t). \quad (2)$$

2.3. Lag K Weighted Advantage

Furthermore, we define the lag K weighted advantage to be: $\tilde{\tau}_{t,K}(a, a_0, S_t(\bar{A}_{t-1})) = \sum_{k=1}^{K} w_k \tau_{t,k}(a, a_0, H_t(\bar{A}_{t-1}))$, where K is the largest lag of interest and $w_1, \ldots w_K$ are predefined non-negative weights and $w_1 + \cdots + w_K = 1$. 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 within four hours after the dose injection, we could define Y_t as the percentage of time the blood sugar being outside 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: $\pi_t^{opt} = \arg\max_{\pi_t} \tilde{\tau}_{t,K}\big\{a = \pi_t(H_t(\bar{A}_{t-1})), a_0, H_t(\bar{A}_{t-1})\big\}$. Notice that the choice of a_0 does not affect the optimal treatment regime. In this article, we take $a_0 = 0$. For simplicity of notation, we write $\tau_{t,k}(a, 0, H_t(\bar{A}_{t-1}))$ as $\tau_{t,k}(a, H_t(\bar{A}_{t-1}))$ in the rest of this article.

2.4. Estimation Method

We first use a nonparametric version of the structural nested models to estimate the lag k treatment effect. 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.

$$\tau_{t,k}(a, H_t) = \tau_k(a, S_t; \alpha_k, \beta_k) = \alpha_k a^2 + \{\beta_k^T f_k(S_t)\}a. \quad (3)$$

where f_k is a q_k dimensional function of S_t. Notice that we assume S_t for $t = 1, \ldots, T$ to be from the same vector space $\mathcal{S}$ and the parameters in this model do not vary with t. Boruvka et al. (2018) showed that the models for the lagged effects for different k do not constrain one another. 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 $f_1(S_t) = \cdots = f_K(S_t)$. (Otherwise, just let $f(S_t)$ be a vector of functions which includes all the functions from $\{f_1(S_t), \ldots, f_K(S_t)\}$ and substitute $f_1(S_t), \ldots, f_K(S_t)$ with $f(S_t)$.) Then the weighted lag K advantage is:

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

where $\tilde{\alpha}_K = w^T\alpha$, $\tilde{\beta}_K = w^T\beta$. Thus the lag K weighted advantage 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 $\pi_t^{opt} = \arg\max_{\pi_t} \tilde{\tau}_K(a = \pi_t, S_t; \alpha, \beta)$ 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: $\pi_t^{opt} = -\{\tilde{\beta}_K^T f(S_t)\}/2\tilde{\alpha}_K$. 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 $f(S_t)$ while having all the other covariates the same, $j = 1, \ldots, q$ where q is the dimension of $f(S_t)$. When $\tilde{\alpha}_K \geq 0$, the optimal treatment falls on the edge of $\mathcal{A}$.

We first present the standard structural nested models for estimating the lag k causal effect. Let $U_{t+k} = Y_{t+k} - \tau_k(A_t, H_t)$. Under the proposed model,

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

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

$$\begin{aligned} E\big[\big\{d(A_t, H_t) - E(d(A_t, H_t)|H_t)\big\}\times \\ \big\{U_{t+k} - E(U_{t+k}|H_t)\big\}\big] = 0, \end{aligned} \tag{4}$$

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 $\{H_{T+1}^1, \ldots, H_{T+1}^n\}$. Then we can estimate α_k, β_k with:

$$\begin{aligned} 0 =& \mathbb{P}_n \sum_{t=1}^{T-k+1} \big\{d_{t+k}(A_t, H_t) - E\big(d_{t+k}(A_t, H_t)|H_t\big)\big\} \\ &\times \big\{U_{t+k}(\alpha_k, \beta_k) - E\big(U_{t+k}(\alpha_k, \beta_k)|H_t\big)\big\}, \end{aligned} \tag{5}$$

where

$$d_{t+k}(A_t, H_t) = -\frac{\partial U_{t+k}(\alpha_k, \beta_k)}{\partial(\alpha_k, \beta_k)} = \begin{pmatrix} A_t^2 \\ A_t f_k(S_t) \end{pmatrix}. \tag{6}$$

and $\mathbb{P}_n$ denotes empirical mean of a function, $\mathbb{P}_n g(A_t, S_t, Y_t) = \sum_{i=1}^n g(A_t^i, S_t^i, Y_t^i)/n$ for any function $g(\cdot)$. Since $d_{t+k}(\cdot)$ depends on H_t only through S_t, we write it as $d_{t+k}(A_t, S_t)$. 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, the complexity of the model increases as t increases, leading to a high risk of model misspecification. If nonparametric estimators are used, the high dimension of H_t can also induce large variance. Therefore, we revise the estimation equation by first showing the following result (See the appendix for the proof).

Theorem 1. *Under model assumption (3), if the following assumption is satisfied:*

$$A_t \perp Y_{t+k}(\bar{a}_{t+k-1})|S_t \quad \textit{for} \quad \bar{a}_{t+k-1} \in \mathcal{A}^{t+k-1}, \tag{7}$$

then for an arbitrary function $d(\cdot) : \mathcal{A} \times \mathcal{S} \to \mathbb{R}^{q_k+1}$*:*

$$\begin{aligned} E\big[\big\{d(A_t, S_t) - E(d(A_t, S_t)|S_t)\big\}\times \\ \big\{U_{t+k} - E(U_{t+k}|S_t)\big\}\big] = 0. \end{aligned} \tag{8}$$

Therefore we can estimate α_k, β_k with:

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

where $d_{t+k}(A_t, S_t)$ is also taken to be (6). The advantage of this estimation equation is that the dimension of S_t does not increase with t. Therefore we can use nonparametric estimators for $E(U_{t+k}|S_t)$ and $E(d(A_t, S_t)|S_t)$ without imposing model assumptions on $A_t|S_t$ and $Y_{t+k}|S_t$. The above equation can thus be written as:

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

Let $B_t(s) = E(A_t^2|S_t = s)$, $C_t(s) = E(A_t|S_t = s)$, $D_{t,k}(s) = E(Y_{t+k}|S_t = s)$. We estimate $B_t(s)$, $C_t(s)$, $D_{t,k}(s)$ with kernel estimators: $\hat{B}_t(s) = \sum_{i=1}^n A_t^{i\,2} K_\Lambda(h - S_t^i)/\{\sum_{i=1}^n K_\Lambda(s - S_t^i)\}$, $\hat{C}_t(s) = \sum_{i=1}^n A_t^i K_\Lambda(s - S_t^i)/\{\sum_{i=1}^n K_\Lambda(s - S_t^i)\}$, $\hat{D}_{t,k}(s) = \sum_{i=1}^n Y_{t+k}^i K_\Lambda(s - S_t^i)/\{\sum_{i=1}^n K_\Lambda(s - S_t^i)\}$, 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. We can then derive the estimated parameters:

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

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 $\hat{\tilde{\alpha}}_K < 0$, π_t^{opt} can be estimated by $\hat{\pi}_t^{opt} = -\{\hat{\tilde{\beta}}_K^T f(S_t)\}/2\hat{\tilde{\alpha}}_K$. When $\hat{\tilde{\alpha}} \geq 0$, $\hat{\pi}_t^{opt}$ would be either 0 or the maximum possible dosage.

Since in model (3), 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. Then we can let:$\hat{B}_t(s) = \{\sum_{i=1}^n \sum_{t=1}^T A_t^{i\,2} K_\Lambda(s - S_t^i)\}/\{\sum_{i=1}^n \sum_{t=1}^T K_\Lambda(s - S_t^i)\}$, where the sum is taken over t as well (similar for $\hat{C}_t(s)$ and $\hat{D}_{t,k}(s)$). 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) and (7). 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 and the dependence between A_t and $Y_{t+k}(\bar{A}_{t-1}, a, A_{t+1}^{a_t=a}, \ldots, A_{t+k-1}^{a_t=a})$ for $k = 1, \ldots, K$. In our simulation study, we will also examine the performance of the model when assumption (7) is not valid.

3. 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:

$$(\hat{\alpha}_k, \hat{\beta}_k^T)^T = \big[\mathbb{P}_n L_1(H; \hat{B}, \hat{C})\big]^{-1}\big[\mathbb{P}_n L_2(H; \hat{B}, \hat{C}, \hat{D})\big],$$

where,

$$L_1(H; B, C) = \sum_{t=1}^{T-k+1} \begin{pmatrix} {A_t}^2 - B_t(S_t) \\ \{A_t - C_t(S_t)\} f_k(S_t) \end{pmatrix}^{\otimes 2},$$

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

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

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

From Equation (8), it is trivial to obtain that $(\alpha_k^*, \beta_k^{*T})^T$ are the true parameters of the model if the model assumption (3) is correct. To derive the asymptotic normality of the estimators, we need the following regularity assumptions:

Assumption 1. *The marginal density of S_t, p_{S_t}, is uniformly bounded away from 0 for all t:* $\inf_{s\in\mathcal{S}} p_{S_t}(s) > 0$.

Assumption 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.*

Assumption 3. $E(A_t|S_t = s)$, $E(A_t^2|S_t = s)$, $E(A_t^4|S_t = s)$, $E(Y_{t+k}|S_t = s)$, $E(Y_{t+k}^2|S_t = s)$, $p_{S_t}(s)$ *as functions of s are uniformly bounded for $s \in \mathcal{S}$. The first derivatives of these functions are also uniformly bounded.*

Assumption 1 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 $\hat{p}_{S_t}(s) = \frac{1}{n}\sum_{i=1}^n K_\Lambda(s - S_t^i)$, which converges in probability to $p_{S_t}(s)$, on the denominator. The first equation in Assumption 2 ensures the unbiasedness of the kernel estimator. When $\mathcal{S} = \mathbb{R}^d$, this assumption is satisfied by most 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 ensure that the limit distributions of the kernel estimators exist. Assumption 3 ensures that the higher order terms of the Taylor expansion of the kernel estimators converge to zero. Then we have the following theorem.

Theorem 2. *If assumptions 1–3 are satisfied, and 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 proof of the theorem is in the appendix.

4. 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 $A_t \in \mathcal{A} = \mathbb{R}$.

Notice that when $S_t = X_t$ and $\theta_2 = 0$, assumption (7) is satisfied (Proof is provided in the appendix). 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 also prove that for $k \geq 2$, the lag k effect under this generative model also follows a quadratic form (See appendix for

details). 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 with $T = 50$ and sample size $n = 100, 200, 400$. We take $S_t = X_t$ and use the proposed method to estimate the treatment effects for lag $1, 2$ and 3. 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, and $f(S_t) = S_t$. In practice, different kernels can be used, which usually will lead to similar results. Here we chose the Gaussian kernel over the others mainly for its computational simplicity. 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 [1]. The results are presented in Table 1.

As presented in Table 1, the proposed method was able to estimate the parameters with small bias. The standard deviation of the estimated parameters decreased 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 scenarios. However, the estimated standard errors slightly underestimated the standard deviation, leading to an undercoverage for the confidence intervals. From the proof of Theorem 2 in the appendix, we see that the variance of the estimated parameters consist of two parts, the variance from the estimation equation and the variance from the kernel estimation. The latter part of the variance converges to 0 as n goes to infinity and is thus excluded from the asymptotic variance formula. However, when the sample size is not large enough, excluding this part of the variance might lead to underestimation of the variance, as supported by the simulation result.

Table 2 presents 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 $\hat{\pi}_t^{opt} = -\hat{\tilde{\beta}}_K^T S_t/(2\hat{\tilde{\alpha}}_K)$ 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 $t = 1, \ldots, T+3$. Table 2 presents the average lag 3 weighted advantage across time $\bar{\tilde{\tau}}_K = \sum_{t=1}^T \tilde{\tau}_{t,K}(a = \hat{\pi}_t^{opt}, S_t)/T$. The average lag 3 weighted advantage of the true optimal treatment regime

[1] The R code for the simulation can be found in https://github.com/lz2379/Mhealth.

Table 1. Simulation results from 200 replicates for observational studies.

k	n	Parameter	Bias[1]	SD[1]	SE[1]	CP
1	100	α_k	0.9	16.5	14.8	93.0
		$\beta_{k,0}$	0.9	9.3	9.3	95.0
		$\beta_{k,1}$	1.1	39.6	35.0	95.0
	200	α_k	-0.4	11.1	10.4	93.5
		$\beta_{k,0}$	0.3	6.8	6.3	91.0
		$\beta_{k,1}$	-0.3	25.7	24.0	92.0
	400	α_k	0.4	7.9	7.4	91.5
		$\beta_{k,0}$	-0.1	4.5	4.3	92.0
		$\beta_{k,1}$	-0.1	18.3	16.7	93.5
2	100	α_k	1.7	31.6	29.0	92.5
		$\beta_{k,0}$	-1.0	23.0	22.3	93.0
		$\beta_{k,1}$	-3.1	79.7	67.9	92.0
	200	α_k	0.2	23.5	20.8	91.5
		$\beta_{k,0}$	-0.3	16.5	15.7	93.5
		$\beta_{k,1}$	-0.6	54.7	47.6	92.0
	400	α_k	-1.8	14.5	14.7	95.5
		$\beta_{k,0}$	0.7	11.8	11.1	92.5
		$\beta_{k,1}$	-1.2	33.6	33.0	94.5
3	100	α_k	0.3	32.4	26.8	88.5
		$\beta_{k,0}$	4.1	22.1	20.9	94.0
		$\beta_{k,1}$	6.2	75.2	67.0	90.5
	200	α_k	-3.8	19.4	18.8	94.0
		$\beta_{k,0}$	0.7	15.4	14.6	91.5
		$\beta_{k,1}$	4.4	47.9	45.3	91.0
	400	α_k	-1.7	15.6	13.3	88.5
		$\beta_{k,0}$	0.7	10.4	10.2	93.0
		$\beta_{k,1}$	2.2	36.1	31.4	91.0

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

[2] Note: Bias refers to the average bias from 200 replicates; 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%.

Table 2. Estimated Parameters for Lag 3 Weighted Advantage from 200 Replicates

n	$\tilde{\alpha}_3$ $(\times 10^{-2})$	$\tilde{\beta}_{3,0}$ $(\times 10^{-2})$	$\tilde{\beta}_{3,1}$ $(\times 10^{-2})$	$\bar{\tilde{\tau}}_K(\hat{\pi}_t^{opt}, S_t)$ $(\times 10^{-3})$
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: The numbers in the parenthesis are the standard deviations.

Table 3. Simulation results from 200 replicates when $\theta_2 = -0.1$.

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

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

[2] Note: Bias refers to the average bias from 200 replicates; 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%.

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 (7) is not satisfied, we generate the datasets with the same parameters except that $\theta_2 = -0.1$. Under this setting, assumption (7) is not satisfied for $k = 1$ when $S_t = X_t$ (see appendix for details). The result of the simulation is presented in Table 3. The estimated parameters for $\beta_{1,0}$ were biased, thus leading to wrong statistical inference of the parameters. Since for $k = 2, 3$, assumption (7) is still satisfied, the result remained unbiased (see appendix for the complete results). We also calculate the average lag 3 weighted advantage of the treatment regime suggested by the biased estimation equation with $w = (1/3, 1/3, 1/3)$. The average lag 3 weighted advantage of the true optimal treatment regime is 64.5×10^{-3}, while the average lag 3 weighted advantages of the estimated treatment regime are 63.9×10^{-3}, 64.0×10^{-3} and 64.1×10^{-3} for sample size 100, 200 and 400. In this particular setting, the recommended treatment regime was still close to optimal. However, it cannot be guaranteed that the suggested treatment regime would be close to optimal in a different setting. 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 with $S_t = (X_t, A_{t-1})$ are given in the appendix.

5. Type 1 Diabetes Data Analysis

Rapid-reacting insulin therapies 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 provide real-time tracking on blood glucose, physical activity and insulin injections of the patients and thus facilitate the dose adjustments to prevent adverse events (Maahs et al., 2012). We apply our method to the Ohio type 1 diabetes dataset collected by Marling & Bunescu (2018) to estimate the lagged treatment effects of the doses and then provide dose suggestions which maximize the weighted advantage[2].

This dataset contains six patients, each with eight weeks of data, including: 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 insulin usage and blood glucose levels. Therefore, we regard them as 6 separate datasets. We illustrate with the data of one patient and assume that the data from each day of this patient are independent from each other. Results of the other patients are presented in the appendix. 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. 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 and 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 planned ahead of time and $\text{Carb-Planned}_t = \text{Carb}_{t+1}$. A_t is the total bolus injection from $t-1$ to t. Let $A_{\max}$ be the maximum observed dose across all days and time. We estimate $\mathcal{A}$ with the interval $[0, A_{\max}]$. Y_t is taken to be the average of the index of glycemic control (IGC) between time $t-1$ and t calculated by:

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

where G is the measured blood glucose level (See Rodbard (2009) for various criterias for glycemic control evaluation).

[2] The R code for real data application can be found in https://github.com/lz2379/Mhealth.

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 $\text{Basal-4-8-hour}_t = \sum_{l=8}^{15} \text{Basal}_{t-l}/8$. These covariates are chosen because they are significantly correlated with A_t from exploratory analysis. To satisfy assumption (7), all covariates correlated with A_t need to be included in S_t. We take $f(S_t) = (\text{Carb}_t, \text{Carb-Planned}_t, \sum_{k=8}^{15} \text{Basal}_{t-k}/8, A_{t-1})$ and predict the treatment effect of the dosage within two hours, $k = 1, \ldots, 4$. Thus the model for the lag k causal effect can be written as:

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

We still use Gaussian kernel and 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 4. Estimated variables with the Ohio type 1 diabetes dataset

k	1	2	3	4	Weighted
α_k ($\times 10^{-1}$)	-1.3	-2.1	-1.4	-0.5	-1.3
	(0.9)	(1.5)	(1.3)	(1.2)	(1.1)
$\beta_{k,0}$ ($\times 10^{-1}$)	15.8	45.6	50.0	33.8	36.9
	(8.2)	(14.7)	(11.1)	(17.8)	(10.9)
$\beta_{k,1}$ ($\times 10^{-2}$)	2.2	1.7	1.5	2.5	1.5
	(1.1)	(1.6)	(1.9)	(2.0)	(1.4)
$\beta_{k,2}$ ($\times 10^{-2}$)	2.5	1.9	0.9	0.7	1.4
	(1.0)	(1.4)	(1.5)	(1.7)	(1.2)
$\beta_{k,3}$ ($\times 10^{-1}$)	-15.6	-40.5	-47.4	-35.2	34.9
	(9.7)	(14.8)	(12.5)	(18.7)	(11.8)
$\beta_{k,4}$ ($\times 10^{-1}$)	-0.6	-1.5	-0.8	-1.4	-1.0
	(1.2)	(2.0)	(2.5)	(2.6)	(1.9)

[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 4. The optimal treatment would be the one which maximizes the weighted advantage for two hours. Since the estimated $\tilde{\alpha}_K$ was negative, 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 be higher when the carbonhydrates intake was higher over the past half an hour or the planned carbonhydrates intake is higher for the next half an hour ($\hat{\tilde{\beta}}_{K,1}, \hat{\tilde{\beta}}_{K,2} > 0$); the optimal dose should be lower 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 result also implies that the past basal insulin infusion rate is an important factor in deciding the optimal insulin dosage for the current moment.

We then estimate the lag K weighted advantage on the test dataset using the estimated parameters $\hat{\tilde{\tau}}_{t,K}(a, S_t) = \hat{\tilde{\alpha}}_K a^2 + \hat{\tilde{\beta}}_K f(S_t)$. 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 two hours.

6. Discussion and Conclusion

In this article, 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 were 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 was defined to be the one which maximizes this advantage. The R code for the simulations and the real data application is provided in the supplementary material.

The proposed method fills the gap in the literature of sequential decision making where the goal is to provide dose suggestions which maximize short-term outcomes. This semiparametric model provides more robustness against model misspecification. By conditioning on partial information of the past history, the proposed method excludes irrelevant information for the estimation of the optimal treatment regime. Thus, the complexity of the suggested optimal dosage would not increase as T increases and is more practical when applied to infinite horizon data. 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 can 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 with our covariance function. The estimated treatment regime was close to maximizing 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 also subject to a few limitations. First, the proposed method is 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 of more interest in certain scenarios. However, the single-time change measurement can still be of great use in practice. In real life, when patients are taking medications, they are likely to take medications only a few times per day. It would be useful to measure what would be the best dosage if the patient takes the medication at the moment and conduct no further medical actions for the next few hours. Second, when the key assumption (7) is not satisfied, the proposed method might lead to biased results. In Appendix C.2, we showed that this bias can be avoided by including all variables that are influential to the decision making process. In practice, for diseases like diabetes and hypertension (which are the main applications we are interested in), patients typically receive education from clinicians on dosage calculation before starting the treatment. Take diabetes as an example. Key decision-making factors, including meals, exercise levels, physical indicators, are collected by most blood glucose monitoring applications. The proposed method can be applied to enhance the performance of the dosages when the key factors for dosing are well established for the patients and can be easily collected. However, the assumption (7) might be hard to examine when the decision making process of the patient is unknown. Therefore, it is definitely essential to let patients be aware of the possible fallacy of the method when an outside factor is guiding his/her decision making process. Third, due to the quadratic form of the model for the lagged treatment effects, a slight underestimate of the quadratic effects of the doses may lead to a large overestimate of the optimal dose. Possible future work would include improving the method to avoid overestimation in doses.

In the future, we are also interested in extending the method to incorporate both long-term optimization and short-term monitoring. Applying this method to online streaming data is also of great interest. For a fixed group of users, the proposed method can be extended to allow streaming data without much additional computation. If the number of users is large, kernel estimation would be computationally expensive. One potential solution is to divide users into subgroups according to certain demographic or medical similarities and then conduct kernel estimation within each subgroup.

References

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. Nutrition recommendations and interventions for diabetes: a position statement of the american diabetes association. *Diabetes care*, 31:S61–S78, 2008.

Bolger, N. and Laurenceau, J.-P. *Intensive longitudinal methods: An introduction to diary and experience sampling research*. Guilford Press, 2013.

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

Carrà, G., Crocamo, C., Bartoli, F., Carretta, D., Schivalocchi, A., Bebbington, P. E., and Clerici, M. 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, 2016.

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

Chen, H., Lu, W., and Song, R. Statistical inference for online decision-making: In a contextual bandit setting. *Journal of the American Statistical Association*, (just-accepted):1–22, 2020.

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

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

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

Fan, A., Lu, W., and Song, R. Sequential advantage selection for optimal treatment regime. *The annals of applied statistics*, 10(1):32, 2016.

Free, C., Phillips, G., Watson, L., Galli, L., Felix, L., Edwards, P., Patel, V., and Haines, A. The effectiveness of mobile-health technologies to improve health care service delivery processes: a systematic review and meta-analysis. *PLoS medicine*, 10(1):e1001363, 2013.

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

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

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

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

Levine, B.-S., Anderson, B. J., Butler, D. A., Antisdel, J. E., Brackett, J., and Laffel, L. M. Predictors of glycemic control and short-term adverse outcomes in youth with type 1 diabetes. *The Journal of pediatrics*, 139(2):197–203, 2001.

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

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

Liang, S., Lu, W., and Song, R. Deep advantage learning for optimal dynamic treatment regime. *Statistical theory and related fields*, 2(1):80–88, 2018.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Schafer, J. L. 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*, pp. 38–62, 2006.

Schulte, P. J., Tsiatis, A. A., Laber, E. B., and Davidian, M. 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, 2014.

Schwartz, J. E. and Stone, A. A. The analysis of real-time momentary data: A practical guide. *The science of real-time data capture: Self-reports in health research*, pp. 76–113, 2007.

Shi, C., Song, R., Lu, W., and Fu, B. Maximin projection learning for optimal treatment decision with heterogeneous individualized treatment effects. *Journal of the Royal Statistical Society. Series B, Statistical methodology*, 80(4):681, 2018.

Shi, C., Song, R., and Lu, W. Concordance and value information criteria for optimal treatment decision. *Annals of Statistics*, 2019.

Song, R., Luo, S., Zeng, D., Zhang, H. H., Lu, W., and Li, Z. Semiparametric single-index model for estimating optimal individualized treatment strategy. *Electronic journal of statistics*, 11(1):364, 2017.

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

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

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

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

Xiao, W., Zhang, H. H., and Lu, W. Robust regression for optimal individualized treatment rules. *Statistics in medicine*, 38(11):2059–2073, 2019.

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

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

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