CLC_____ Number_____

UDC_____ Available for reference ☐Yes ☐No

Undergraduate Thesis

Thesis Title: Dynamic Treatment Regime in Parkinson's Disease

Student Name: ZHANG ZIYI

Student ID: 11712621

Department: Department of Statistics and Data Science

Program: Statistics

Thesis Advisor: ZHANG ZHEN

Date: June 4, 2021

分类号_____ 编 号_____

U D C_____ 密 级_____

南方科技大学
SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

本科生毕业设计（论文）

题　　目：Dynamic Treatment Regime in Parkinson's Disease

姓　　名：张子怡

学　　号：11712621

系　　别：统计与数据科学系

专　　业：统计学

指导教师：张振

2021 年 6 月 4 日

COMMITMENT OF HONESTY

1. I solemnly promise that the paper presented comes from my independent research work under my supervisor's supervision. All statistics and images are real and reliable.

2. Except for the annotated reference, the paper contents no other published work or achievement by person or group. All people making important contributions to the study of the paper have been indicated clearly in the paper.

3. I promise that I did not plagiarize other people's research achievement or forge related data in the process of designing topic and research content.

4. If there is violation of any intellectual property right, I will take legal responsibility myself.

Signature：

Date：

Dynamic Treatment Regime in Parkinson's Disease

张子怡

（统计与数据科学系　指导教师：张振）

[**ABSTRACT**]: Parkinson's disease (PD) is a common long-term disorder that mainly affects the human's motor system. Researchers have developed several dopamine replacement drugs, like L-Dopa and DA-agonists. However, patients and doctors always face a dilemma: whether to delay the treatment initiation time because most of these drugs come with unexpected side effects. Decision-making guidance is both necessary and expected.

Nowadays personalized medicine attracts more and more attention, it emphasizes the exploitation of patient's historical information in making treatment decisions, and the dynamic treatment regime (DTR) is a concept extending core ideas of personalized medicine to chronic conditions. For the DTR, the randomized trial is generally the golden rule to gather data as there are no confounding effects. While on the other hand, observational data, which is much richer and cheaper, also offers a complimentary valuable source that should be paid with great attention.

Combining problems in these two fields, we decided to build a treatment management rule set according to analyzing a PD observational dataset from Parkinson's Progression Markers Initiative (PPMI), a public Parkinson’s patient data collection study. We first paid great effect in variable selection, and then used dynamic weighted ordinary least square, an easy-to-apply optimal DTR estimation technique with robustness, to

figure out an optimal rule set for the when-to-treat problem in Parkinson's disease, the treatment regime performs well in the DTR model diagnostics. This project shows the value of applying DTR's ideas in observational studies.

The paper's structure is as follows: it first outlines DTR problem settings and planning methods, then illustrates details of data analysis. Meanwhile, several difficulties that are always accompanied by analyzing observational data and their solutions are talked about.

[**Keywords**]: dynamic treatment regimes; Parkinson's Progression Markers Initiative; observational data; time-varying confounding effects

[摘要]：帕金森症（Parkinson's disease）是一种常见的慢性疾病，对人体的运动系统有很大影响。如今帕金森症的研究人员已经开发了数种有效的药物以缓解病症，如左旋多巴和多巴胺促进剂。但由于这些药物副作用较大，患者和医生始终面临着一个难题：是否推迟药物治疗的开始时间。因此，对帕金森症的药物决策指导是十分必要的。

另一方面，近年来个性化医疗（personalized medicine）方兴未艾，它强调在制定决策时充分利用患者的历史医疗信息，而动态治疗方案（dynamic treatment regime）把这种思想推广到慢性病上。在动态治疗规划的研究中，往往会通过随机试验收集数据，以期减少混杂效应（confounding effects）。但是，更丰富更便宜的观测数据也提供了宝贵的数据资源，如何利用动态治疗方案研究观测数据很值得思考。

结合这两个领域的问题，我们决定通过分析研究帕金森症进展标记倡议（Parkinson's Progression Markers Initiative）——一个致力于收集帕金森氏病患者数据的公共项目——提供的观测数据来构建动态治疗方案。我们首先进行了变量选择，然后应用了动态加权的最小二乘法（一种易用且鲁棒性好的规划方法），为帕金森氏病“何时用药”的问题找到了最佳指导方案，该方案在之后的模型诊断中表现良好。本项目充分体现了用动态治疗方案探索观测数据的价值。

全文主体结构如下：首先概述动态治疗方案的问题设置和规划方法，然后展示数据分析的详细过程。同时讨论了一些经常随分析观察数据而来的问题及其解决方法。

[关键词]：动态治疗规划；帕金森症进展标记倡议；观测数据分析；动态混杂效应

CONTENTS

1. Introduction

1.1 Dynamic Treatment Regimes

Personalized Medicine (PM, also known as precision medicine) is an increasingly popular theme that adapts treatments (type, dosage, and timing) individually by patient-specific data, like medical respond, genetic information. PM has a close relationship with the chronic care model (CCM)[29], which contains three main features against the traditional acute care model: individualized (vs. one-size-fits-all), dynamic treatments (vs. static treatments), and evidence-based (vs. expert options).

A dynamic treatment regime (DTR)[12], also called adaptive treatment strategy or treatment policy, is indeed a set of treatment guiding rules and its position is like the "agent" in reinforcement learning (RL). The major goal in DTR problems, similar to that in reinforcement learning, is finding an optimal agent with the largest outcome. This concept was proposed to generalize ideas of PM and CCM from single-stage to sequential decision-making processes where treatments are tailored to a patient's time-varying (dynamic) states. Thus DTRs offer effective treatment management vehicles especially suitable for those long-term intervention conditions (weight loss, alcohol/smoke abuse, etc.).

We have mentioned relationships between DTRs and well-known classical RL settings (Markov decision process), but we also need to remind every reader of the difference between them[5], which makes methods in these two fields different.

- DTRs do not obey the Markov property: decision rules in each stage need to rely on historical information, not just on the previous stage information. This setting is intuitive, doctors would make treatment decisions according to patients' whole medical records.

- DTRs focus on unknown system cases and there may be significant confounding effects (for observational data): in terms of model information

like transition probability, finding an optimal DTR is like searching for an agent in a model-free RL case.

- DTRs need to fully use the data compared to the RL problem:
  - Generating new clinical data is not as cheap as that in classical RL game problems.
  - Sample size n is usually not comparatively large than the state size, and that means we have better use semi-parametric methods instead of parametric ones.

Data for DTR researches are mainly from three resources: randomized trials, observational studies, and dynamic models, for details of this please check[4].

Before building a DTR from collected data, there are two basic assumptions[18] ensuring that we can make reliable inferences on treatment allocation and outcomes from the collected data:

- Consistency: patients' treatment effects are independent and cannot interact, the counterexample is herd immunity. This assumption subsumes Rubin's stable unit treatment value assumption (SUTVA)[23].
- No unmeasured confounders: indicates that for any subject in stage k, his/her treatment a_k completely depends on the history $\mathbf{h}_k$, not on any future information. Take attention that this is not the Markov property in RL, which describes the link of two consecutive stages.

What's more, if using observational data, researchers can only explore those "viable" DTRs[4]: rule regimes that have already been followed by some participants.

In constructing optimal decision rules, dynamic program[2] was the traditional way, but it suffers from two problems: the curse of modeling (one need complete knowledge of the model, like all transition probabilities, this is unrealistic in clinical data analysis) and curse of dimensionality. So nowadays researchers have developed

two other method classes for finding DTRs, one is value search, methods include inverse probability weighting and marginal structural models[21], the doubly robust augmented inverse probability of treatment weighting method[35], and outcome weighted learning[36]; another class is based on regression, contains classical Q-learning[26][34], G-estimation[19], and dynamic weighted ordinary least squares (dWOLS)[30]. This paper focuses on the last class.

Notice that most works we state above only check binary treatment situations, but there are already some articles notifying continuous treatment allocations. For example, in value search, there are extended inverse probability weighting[14] and marginal structural models[6], while for regression-based method, luckily, Q-learning and G-estimation naturally fit any type of treatment. Despite this, the paper mainly talks about binary treatment cases.

1.2 Parkinson's Disease

Parkinson's disease (PD) is a long-term neurodegenerative disorder, in which disease progression is hard to reverse. Most PD treatments focus on pharmacological dopamine replacement, they were shown effective in slowing PD's progression[15]. But as these drugs themselves are also new sources of disability due to their unexpected side effects[8][16], many patients and doctors would like to begin the drug treatment as late as possible, we thus want to construct an optimal DTR to help understand whether delaying is a good choice. To achieve this, we first need to quantify the disease severity, we found that previous PD researchers have developed some rating scales to assess PD's progress, in this paper we will focus on the most popular one of them: Movement Disorder Society-Unified Parkinson’s Disease Rating Scale (MDS-UPDRS)[7], this is a four-part measurement considering not only PD patient’s motor symptom but also non-motor performance.

As for analyzed data, although the randomized trial is the golden standard in both clinical studies and DTRs analysis, van den Heuvel et al.[28] pointed out its weaknesses: short follow-up time, artificial inclusion/exclusion criteria, and not

reflecting realistic treatment decision. In their viewpoint, once confounding effects can be adjusted properly, observational data will show more value. Thus this project's other goal is to gain experience to solve difficulties in observational DTR analysis.

The data we use is from Parkinson's Progression Markers Initiative (PPMI), this is a worldwide ongoing observational study launched in 2010, it aims to help researchers to speed therapeutic development and collects the longitude states of patients. PPMI (www.ppmi-info.org) is sponsored by The Michael J. Fox Foundation for Parkinson's Research and supported by various industry, non-profit and private partners. All data we used in the code was accessed online on April 2, 2021.

The rest of the article is organized as follows. In Section 2, we briefly go through the framework of DTR problems and two optimal DTR building methods: Q-learning and dynamic weighted ordinary least squares (dWOLS). Detailed PD's data analysis and findings are shown in Section 3. Finally, we discuss the shortcoming of this research and give some future work directions, inspired by several most recent papers.

2. Methodology

2.1 Basic Concepts and Notations

As we said, this paper applies the indirect regression method to identify optimal DTR. So first let us introduce fundamental notions and concepts in DTRs[4][30], consider a K-stage **binary** treatment (treatments can be regarded as on/off or drug/placebo) case. As a traditional standard, lowercase letters denote realizations of random variables.

- y: (Final) patient's outcome, the long-term utility, can be one specific score or a composite outcome combining several scores, typically the larger the better.
- a_k: The kth treatment decision. Values are taken from $\{0, 1\}$ and $a = 0$ denotes reference treatment.

- $\mathbf{x}_k$: State information (age, disease duration, etc.) observed prior to the kth treatment decision.

- $\mathbf{h}_k$: History information before the kth treatment decision, includes previous states information $(x_1, \ldots, x_k)$ and treatments $(a_1, \ldots, a_{k-1})$. Usually, we like to denote $\underline{\mathbf{a}}_k = (a_{k+1}, a_{k+2}, \ldots, a_K)$ (similarly, $\underline{\mathbf{x}}_k = (x_{k+1}, x_{k+2}, \ldots, x_K)$ and future treatment sequence notation is $\overline{\mathbf{a}}_k = (a_1, \ldots, a_k)$), then $\mathbf{h}_k = (\overline{\mathbf{x}}_k, \overline{\mathbf{a}}_{k-1})$.

Our goal is to find the best sequence of decision rules with the optimal expected outcome y^{opt}. Notice that in many DTR articles including this one, the immediate reward of action at each stage is not in consideration (set it as 0), our focuses locate on the terminal reward. From this point of view, the word "reward" can be seen as the same thing as "outcome"[4].

There are two special functions in DTRs:

- Blip function[18]:

$$\gamma_k(\mathbf{h}_k, a_k) = \mathbb{E}\left[Y^{\overline{\mathbf{a}}_k, \underline{\mathbf{a}}_{k+1}^{opt}} - Y^{\overline{\mathbf{a}}_{k-1}, a_k=0, \underline{\mathbf{a}}_{k+1}^{opt}} \mid \mathbf{H}_k = \mathbf{h}_k\right]$$

  Which is also called the contrast function. At stage k, under identical history and assuming follow optimal treatments thereafter, blip function is to quantify the additive effect in (expected) outcome for selecting an action a_k instead of the reference one.

- Regret function[12]:

$$\mu_k(\mathbf{h}_k, a_k) = \mathbb{E}\left[Y^{\overline{\mathbf{a}}_{k-1}, \underline{\mathbf{a}}_k^{opt}} - Y^{\overline{\mathbf{a}}_k, \underline{\mathbf{a}}_{k+1}^{opt}} \mid \mathbf{H}_k = \mathbf{h}_k\right]$$

  Reflecting the loss in expected outcome by selecting a_k instead of the optimal one. For binary treatment and continuous outcome, blip and regret have a relation: $\mu_k(\mathbf{h}_k, a_k) = \gamma_k\left(\mathbf{h}_k, a_j^{opt}\right) - \gamma_k(\mathbf{h}_k, a_k)$.

From concepts of two above functions, we can infer that at each stage k, the best treatment is the one maximizing kth blip or minimizing kth regret, thus left work is how to model blip/regret functions for every stage. We can see that exploitation of blip and regret converts an optimizing problem to a more usual model fit problem.

Blip functions, expressing addictive effects of treatments, can be naturally applied to decompose expected outcomes (like the V function in RL) into two parts: treatment-free and treatment-related. For example, Wallace and Moodie showed one form of separation[30]:

$$\mathbb{E}[Y^a \mid \mathbf{H} = \mathbf{h}; \boldsymbol{\beta}, \boldsymbol{\psi}] = \underbrace{f(\mathbf{h}_0; \boldsymbol{\beta})}_{\text{treatment-free model}} + \sum_{k=1}^{K} \underbrace{\gamma_k(\mathbf{h}_k, a_k; \boldsymbol{\psi}_k)}_{\text{blip function}} \quad (1)$$

where f is some function irresponsive to treatment a_k, and $\mathbf{h}_0$, a subset of $\mathbf{h}_k$, includes those covariates not related to treatment. Y^a here is to emphasize the treatment regime a may be counterfactual.

2.2 Q-learning in DTRs

Q-learning[26] is a usually seen method in RL, and Murphy (2005) discussed Q-learning form when it is applied in a DTR problem[13], the DTR Q-learning form is close to the fitted Q-iteration algorithm with function approximation in batch RL fields[34]. We first define Q-functions (Q here means "quality") for a K-stage process:

$$\begin{aligned} Q_K(\mathbf{h}_K, a_K) &= \mathbb{E}[Y \mid \mathbf{H}_K = \mathbf{h}_k] \quad \text{and} \\ Q_k(\mathbf{h}_k, a_k) &= \mathbb{E}\left[\max_{A_{k+1}} Q_{k+1}(\mathbf{H}_{k+1}, A_{k+1} \mid \mathbf{H}_k = \mathbf{h}_k, A_k = a_k)\right] \quad \text{for } k < K. \end{aligned}$$

Inspired by the conditional expectation structure, researchers suggest using regression models (linear, regression trees, kernels, etc.) to fit Q-functions. For example, borrowing the idea from (1), a fitted linear Q-function can be divided into two parts, stage-k treatment-free (may contains previous treatment states) and stage-k blip for each stage k:

$$Q_k(\mathbf{h}_k, a_k; \boldsymbol{\beta}_k, \boldsymbol{\psi}_k) = \boldsymbol{\beta}_k^T \mathbf{h}_k^{\beta} + \boldsymbol{\psi}_k^T a_k \mathbf{h}_k^{\psi} \quad (2)$$

where $\mathbf{h}_k^{\beta}$ and $\mathbf{h}_k^{\psi}$ are subsets of $\mathbf{h}_k$.

To estimate parameters $\boldsymbol{\beta}_k$ and $\boldsymbol{\psi}_k$, one can start from stage K (may use classical ordinary least square (OLS) estimation method), and work backward stage-by-stage. After getting the estimated Q-functions, we can prescribe the best treatment following the rule: give treatment $a_k = 1$ if $\widehat{\boldsymbol{\psi}}_k^T \mathbf{h}_k^{\psi} > 0$, and $a_k = 0$ otherwise (comparing blip values). The main difference between (1) and (2) is that (1) accepts all model forms while (2) only selects the linear one. Indeed, when considering continuous treatment cases like drug dose allocation, non-linear terms can be added to restrict results inside reasonable ranges[17].

Mahar et al.[10] and Moodie et al.[11] have talked about the adjusted Q-learning in observational studies, and Moodie et al. imported inverse probability of treatment weighting (IPTW) to deal with confounding effects. These works helped to develop the following important technique.

2.3 Doubly-robust Estimation Method: dWOLS

Though Q-learning's idea is common and easy to be implemented, simplicity of it also leads to a severe limitation: lacking robustness. To get consistent results from Q-learning we need to make sure correct specifications of the whole outcome model[30]. While those more robust methods, like G-estimation, are more complex in statistics and hard for clinical researchers' use. Thus, Wallace and Moodie[30] proposed an advanced estimation method based on Q-learning and G-estimation: dynamic weighted ordinary least square (dWOLS). At the first glance, the technique looks more like Q-learning than G-estimation, it updates Q-learning mainly through two ways, one is to attach weights $w(a, \mathbf{h})$ to the subject with treatment a and covariate $\mathbf{h}$, attached weights in dWOLS must satisfy the balance equation $\pi(\mathbf{h})w(1, \mathbf{h}) = \big(1 - \pi(\mathbf{h})\big)w(0, \mathbf{h})$, where $\pi(\mathbf{h}) = P(A = 1|\mathbf{h})$ is the propensity score[22][27], describing the probability of the subject receive treatment a with history $\mathbf{h}$. Among all weight families satisfying the above equation, Wallace and Moodie[30]

suggested that "absolute value weight" $w(a, \mathbf{h}) = |a - \mathbb{E}[A \mid \mathbf{H} = \mathbf{h}]|$ deserves consideration because of its good performance, and this weight can be naturally interpreted: the more "unusual", the heavier weight will be given. Another improvement is that dWOLS regresses on **pseudo outcomes** instead of Q-value in Q-learning, the pseudo outcome's concept is originally from G-estimation, they are defined as:

$$\begin{aligned} \tilde{y}_K &= y \quad \text{and} \\ \tilde{y}_k &= y + \sum_{k+1}^{K} \mu_k \left(\mathbf{h}_k, a_k; \widehat{\boldsymbol{\psi}}_k\right) \quad \text{for } k < K. \end{aligned}$$

Similar to (2), authors restricted treatment-free models and blip functions in linear form, then the regression equation becomes:

$$\mathbb{E}[\tilde{Y}_k \mid \mathbf{h}_k, a_k; \boldsymbol{\beta}_k, \boldsymbol{\psi}_k] = \boldsymbol{\beta}_k^T \mathbf{h}_k^{\beta} + \boldsymbol{\psi}_k^T a_k \mathbf{h}_k^{\psi} \tag{3}$$

The blip term $\boldsymbol{\psi}_k^T a_k \mathbf{h}_k^{\psi}$ reflects treatment effect, it is generally in the form of addition of main treatment term and interaction terms between treatment and tailoring variables, like $a_k(\psi_{k0} + \psi_{k1} x_k)$.

dWOLS inherits the simplicity of Q-learning and the double robustness of G-estimation: it can get consistent results as long as either the treatment-free model or treatment model is correct. Besides, dWOLS helps to remove the confounding effects from treatment allocation because it applies propensity scores, this makes dWOLS quite suitable for observational data analysis.

Let us do a summary, in dWOLS, there are three models: treatment models (usually in the form of logistics models) to predict propensity scores, treatment-free models, and blip functions. Actually, in the observational data cases, researchers try best to confirm the model specification for the last two, this is because correct treatment-free models can hold double robustness and are useful in predicting pseudo outcomes, while reliable blip functions are cores to make optimal decisions. dWOLS is a backward estimation method because it regresses on pseudo outcomes that depend on

the next stages' parameters, when applying dWOLS, one first needs to estimate the terminal stage's blip parameters, and then go back stage by stage, keep creating pseudo outcomes and estimating blip parameters. Wallace et al.[32] published the DTRreg R package to help researchers use dWOLS.

After several years, dWOLS was extended to continuous treatment cases[25], authors applied generalized propensity scores to build continuous treatment case weight balancing equations, and they proved that weights suitable continuous treatment can be also applied in the binary treatment situations, while the converse is wrong.

2.4 Model Assessment and Selection for dWOLS

dWOLS provides us an easy-to-apply method with double-robust property, but its robustness is still based on the fact that at least two models (blip + treatment/treatment-free) are correct, techniques which can help us judge our models' suitability now are thus necessary and expected. Usually in the realistic world, experts will give a set of candidate models and then select, so ideal techniques should address the model selection problem. Wallace et, al.[33] showed two scientific methods separately, one is for nuisance model (treatment and treatment-free models) assessment by making use of double robustness to assess treatment/treatment-free models, the other is for blip model selection according to a quasilikelihood information criterion (QIC). They are both required in a complete model diagnostics.

First, suppose we already have the correct form of blip function (or at worst over-specified), we can then conduct the model assessment. This method's intuition is from the double robust property, if one of the nuisance models is fixed and misspecified, the blip model parameter estimates will shift larger (larger variance) when the other model changes compared to the proper specified one. According to this, Wallace et, al. [33] said that we should select those nuisance models with the least variance in results when varying another model. This method, however, is asymptotically held, authors suggested doing a bootstrap sample test in practice: keep selecting the lowest-variance (fixed) model in each bootstrap, then identify the one

which is picked the most frequently across bootstraps. Grounded by the double robustness property, this model assessment technique can be extended to many other methods. Another notable thing is that we may regard the whole blip estimated value ($\psi_{k0} + \psi_{k1}x_k$, also called treatment effect) instead of blip parameter estimates (ψ_{k0}, ψ_{k1}) as the result, the change reduces the pressure of specifying blip functions but becomes less powerful, so we mixed-used these two result forms.

Then we turn our attention to model selection, inspired by Akaike[1], Wallace et, al. [33] introduced the quasilikelihood-based method in DTRs, they did not use likelihood-based methods as dWOLS is not built on likelihood theory. This method uses the QIC criterion to choose blip functions analog to what AIC does. Again, bootstrap was shown helpful in improving QIC selection performance.

Both of the two techniques have limitations. For model assessment, it firstly has a problem that prefers over-specified models. However, luckily in DTRs, over-fitting is of much less concern than underfitting, and the authors suggested adding covariates one by one if overfitting is after all beyond sufferance. Another problem for assessment is that the result is related to the blip function's correctness, we have to assume the blip model to be well defined or at least overfitting. Wallace et, al.[33] also gave two ways, one is to include blip-dependent covariates, the second way is to alternate between model assessment and model selection, iteratively select three models until all outputs converge. Although the QIC selection method also comes with overfitting problems, authors believe that there is a modified version of QIC (like the corrected AIC) waiting for future research.

3. Data Analysis

PPMI is a longitude observational data, we set stage 0 as the beginning time, stage 1 as 6 months, stage 2 as 12 months, and so on. And then like many other papers[9][28], we chose the MDS-UPDRS III (motor examination score in MDS-UPDRS) OFF (the

difference between ON/OFF state is whether patients took medicines in the last 6 hours when recording) scores in stage 8 (48 months) as our primary outcome y. In the DTRreg R package, a larger primary outcome means better condition while in our PD case, but a larger MDS-UPDRS III score indicates worse progression, therefore we used its negative value in model fit.

3.1 Preprocessing the Data

Almost all collected observational datasets suffer from missing data and confounding effects problems. As we mentioned above, confounders can be addressed by dWOLS, while for missing data, we mainly followed the imputation steps in[28], where those authors gave several methods to deal with missing, all methods follow one major rule: For any missing variable, use its previous and next states' value to predict by a regression model built upon those subjects who do not miss measurements at the same time point.

From the initial 423 PD observations, we removed 7 who stopped PD medicine after treatment started, and those who still have too many missing variables (imputation did not handle subjects), what's more, we only left subjects with MDS-UPDRS III score values in stage 8. These operations made 286 subjects left, some of their demographics and disease characteristics information is listed in TABLE 1.

TABLE 1. Descriptive Statistics of the 286 PPMI Study Participants.

Variables	Mean (SD; Min; Max)	Percentage (%)
Age	61.37 (9.82; 33.50; 84.83)	
Gender		Male: 66.08
		Female: 33.92
Family History of PD		0 family: 72.03
		1 family: 21.68
		2 families: 4.90
		3 families: 0.70
		4 families: 0.35
		5 families: 0.35
Duration (Months)	6.40 (6.31; 1.00; 35.00)	
MDS-UPDRS part III score (OFF)	20.31 (8.59; 6.00; 47.00)	
RBDSQ score	4.42 (2.84; 0.00; 13.00)	
MoCA score	27.14 (2.29; 17.00; 30.00)	

SD, Standard deviation; Min, Minimum; Max, Maximum; MDS-UPDRS, Movement Disorder Society-Unified Parkinson's Disease Rating Scale; RBDSQ, REM Sleep Behavior Disorder Screening Questionnaire; MoCA, Montreal Cognitive Assessment

TABLE 2 shows the frequency of 286 patients' treatment starting time, after considering sample size, we decided to analyze situations from stage 1 to stage 4 to build up a four-stage DTR with largest outcome. From doctors we knew that PD's medicines cannot be stopped abruptly once initiated, this indicates that once start treatment, patients have no chance to make any other decisions. Thus, in every stage k, we only analyzed those who had not begun treatments until stage k ($a_1, a_2 \ldots a_{k-1} = 0$). Thus, there are separately 286, 258, 105 and 63 observations in each stage analysis.

TABLE 2. Treatment Start Time Analysis

Treatment Initiation Stage (k)	Frequency
6 months (1)	28
12 months (2)	153
18 months (3)	42
24 months (4)	25
>24 months (>4)	38

3.2 Variable Selection

There are more than 100 variables in the PPMI dataset, so it was very important to do variable selection at first. Seeing that dWOLS is indeed doing (weighted) linear regression, we paid attention to previous work on linearly predict PD progress (especially regress on MDS-UPDRS III scores) in PPMI data, we found[9][24]. Besides, we also did stepAIC selection ourselves trying to select appropriate variables. Meanwhile, notice that usually in clinal cases, expert opinions are important references, despite results from data. Following all findings and adding PD domain knowledge from experts into consideration, we finally set our models as (for stage 4):

- Blip function ($\boldsymbol{\psi}_4^T a_4 \mathbf{h}_4^{\psi}$ in equation (3)): $a_4 * (\psi_{k0} + \psi_{k1}$MDS-UPDRS III.4 $+ \psi_{k2}$RBDSQ.4 $+ \psi_{k3}$MOCA.4 $+ \psi_{k4}$MDS-UPDRS II.4$)$. MDS-UPDRS II is the second part of MDS-UPDRS scores. The numbers behind indicate the stage position. Blip functions in stage 1, 2, and 3 are all 0 as $a_k = 0,\ k = 1,2,3$.

- Treatment model: Using logistics regression regress 4th treatment decision a_4 on variables: MDS-UPDRS III.4, RBDSQ.4, MoCA.4, MDS-UPDRS I (DDS).4 and Duration. MDS-UPDRS I (DDS) is the score of the MDS-UPDRS 1.6 problem: FEATURES OF DOPAMINE DYSREGULATION SYNDROME.

- Treatment-free model: linear model contains variables ($\mathbf{h}_4^{\beta}$ in equation (3)): MDS-UPDRS III.4, RBDSQ.4, MOCA.4, MDS-UPDRS I (ANXS).4, MDS-UPDRS III.0, RBDSQ.0, MOCA.0, MDS-UPDRS I (ANXS).0, Duration, Age, and FAM_HISTORY. MDS-UPDRS I (ANXS) is the score of the MDS-UPDRS 1.4 problem: ANXIOUS MOOD.

Using DTRreg, the key code is:

```
R> blip.mod.4 <- list(~MDS_UPDRS_III_other.4+RBDSQ.4+MoCA.4)
R> tr.mod.4 <- list(treatment.4~MDS_UPDRS_III_other.4+RBDSQ.4+MoCA.4+NP1DDS.4+Durati
on)
R> trf.mod.4 <- list(~MDS_UPDRS_III_other.4+RBDSQ.4+MoCA.4+NP1ANXS.4
R>                   +MDS_UPDRS_III_other.0+RBDSQ.0+MoCA.0+NP1ANXS.0
R>                   +Duration+Age+FAM_HISTORY)
R> # We remove vairalbe NP1DDS.4 in trf.mod.4 due to singularity problem.
R>
R> DTRs.mod.4 <- DTRreg(-MDS_UPDRS_III_other.8,bl.mod.4,tr.mod.4,trf.mod.4,df_DTRs_4,me
R> thod="dwols",weight="ipcw",var.estim="bootstrap")
```

where "other" means that the score has been imputed. Here we only show the fourth stage case, others are similar despite that they use pseudo outcomes. Notice that we repeatedly exploited three variables: MDS-UPDRS Part III score, MoCA score and RBDSQ score, they are so crucial as previous work[9] selected these three variables in all five models they tried.

3.3 Results

The parameter estimation results are listed in TABLE 3.

TABLE 3 DTR Analysis of PPMI Data (Using dWOLS)

Stage (k)	$\hat{\psi}_{k0}$	$\hat{\psi}_{k1}$	$\hat{\psi}_{k2}$	$\hat{\psi}_{k3}$	$\hat{\psi}_{k4}$
6 months (1)	-43.79	-0.12	-0.08	1.77	-0.35
12 months (2)	14.46	0.01	-0.27	-0.38	-0.26
18 months (3)	-44.64	-0.17	1.47	1.58	0.1
24 months (4)	21.09	-0.36	-2.05	-0.24	0.76

To use this DTR, one just need to identify the sign of the equation $\hat{\psi}_{k0} + \hat{\psi}_{k1}$MDS-UPDRS III.4 $+ \hat{\psi}_{k2}$RBDSQ.4 $+ \hat{\psi}_{k3}$MOCA.4 $+ \hat{\psi}_{k4}$MDS-UPDRS II.4 in each stage. For example, in stage 4, the regime recommends starting treating patients if his/her equation $21.09 - 0.36 *$ MDS-UPDRS III.4 $- 2.05 *$ RBDSQ.4 $- 0.24 *$ MOCA.4 $+ 0.76 *$ MDS-UPDRS II.4 > 0.

We should declare that our results are of severe non-regularity, non-regularity is a complex problem in the DTR field, we recommend those interested in it to Robins's paper[20]. In short, it means that the results are not significant and the equation's clinical interpretation is not strong. Besides, we noticed a strange phenomenon that may be introduced by non-regularity: consider the meaning of the above equation, it seems to exist Matthew effect (the rich get richer and the poor get poorer). For stages 1, 3 and 4, regimes say that if one has larger MDS-UPDRS III scores (remember that this score is the higher the worse), he/she will be less likely to be treated. The effect can even be observed in the original DTRreg paper[32], where the less weighted infants were suggested to be breastfeed with lower probability. We hope future work can pay attention to this odd effect.

3.4 Model Diagnostics

Now that we got a DTR, it is time to apply techniques for model assessment and selection. Because these methods were still under development, we just simply apply them in stage 4 for an illustration purpose.

We first prepared 3 treatment models and 3 treatment-free models for selection:

```
R> # treatment model, the intercept-only model was originally used to increase variance
R> if (t == 1) {TR <- list(treatment.4~1)}
R> if (t == 2) {TR <- list(treatment.4~MDS_UPDRS_III_other.4+RBDSQ.4+MoCA.4)}
R> # used one
R> if (t == 3) {TR <- list(treatment.4~MDS_UPDRS_III_other.4+RBDSQ.4+MoCA.4+NP1DDS.
R> 4+Duration)}
R>
R> # treatment-free model
R> if (f == 1) {TF <- list(~1)}
R> if (f == 2) {TF <- list(~MDS_UPDRS_III_other.4+RBDSQ.4+MoCA.4+NP1ANXS.4
R>                         +MDS_UPDRS_III_other.0+RBDSQ.0+MoCA.0+NP1ANXS.0
R>                         +Duration)}
```

```
R> # used one
R> if (f == 3) {TF <- list(~MDS_UPDRS_III_other.4+RBDSQ.4+MoCA.4+NP1ANXS.4
R>                         +MDS_UPDRS_III_other.0+RBDSQ.0+MoCA.0+NP1ANXS.0
R>                         +Duration+Age+FAM_HISTORY)}
```

We did 100 simulation runs, each run includes 3*3=9 nuisance model combinations and for every combination, there are 200 bootstraps from the original analyzed data, we built a DTR for each bootstrap. Thus in each run, we constructed 9*200=1800 DTRs. We calculated standard deviations of treatment effects $\hat{\psi}_{k0} + \hat{\psi}_{k1}$MDS-UPDRS III.4 $+ \hat{\psi}_{k2}$RBDSQ.4 $+ \hat{\psi}_{k3}$MOCA.4 $+ \hat{\psi}_{k4}$MDS-UPDRS II.4 when one of nuisance models fixed and the other changed, we chose the fixed model with the least standard deviation. After that in each run, there were 400 results (two kinds of models (treatment and treatment-free)), then following the guidance from[31], we picked the model that appears the most frequently, set it as this run's final choice. In the end, we got 100 final choices for both treatment and treatment-free models. TABLE 4 shows them.

TABLE 4. Lowest SD Model Picked Frequency in 100 Runs

	1	**2**	**3**
treatment model picked times	85	14	1
treatment-free model picked times	0	7	93

Additionally, we took the last run results' standard deviation boxplots out for a closer look, as FIGURE 1 shows. Here for simplicity, we only took two parameters plus treatment effect estimates as an example. In FIGURE 1, each plot contains 1800 DTRs' estimators in the last run, and each box corresponds to one combination of treatment and treatment-free model, with each plot containing the results where one of the two models is held fixed. The top row is to assess different treatment models and the second row is for treatment-free models. We can see that three treatment models perform very likely across the first three plots, in this situation, we chose the T1: the intercept-only model for it is simplest. While turned to treatment-free models F1, F2 and F3, they show close SD of treatment effects, but F2 and F3 outperform F1 when

comparing two parameters' SD. Because treatment effect is not as powerful as parameter estimates, we suggest erring on the side of caution and using F3.

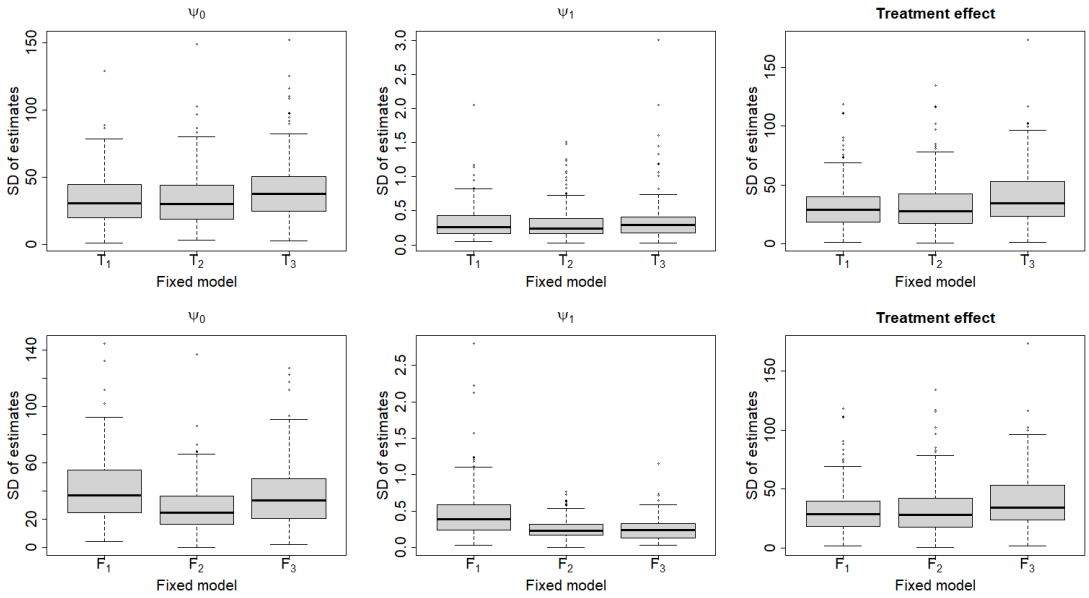

FIGURE 1. Last Run's Bootstrap DTR Estimates for the PPMI Analysis.

After choosing proper nuisance models, we turned our attention to blip models check, we compared two blip models:

R> B1 <- cbind(rep(1,n),MDS_UPDRS_III_other.4,RBDSQ.4,MoCA.4,MDS_UPDRS_II.4) # used one

R> B2 <- cbind(rep(1,n),MDS_UPDRS_III_other.4,RBDSQ.4,MoCA.4)

R> B3 <- cbind(rep(1,n),MDS_UPDRS_III_other.4)

The QIC of them are separately -29369.2, -29280.95 and -29290.19, which suggest that our first choice of blip function is preferable.

Model assessment and selection show that we need to change the treatment model for better results. However, because of the double robustness of dWOLS, we predicted that the estimated parameter will not vary a lot as we had used the proper treatment-free model. The coding analysis also supported our guess.

4. Conclusion and Discussion

In this paper, we mainly review the framework of DTRs and some optimal DTR estimation methods, our focus is on constructing DTRs from PD observational data. PPMI is a dataset that quite suitable for DTR analysis, we searched for the optimal DTR guiding patients when to start treatment to get the largest return. In building a best DTR from PD observational study, we spent a lot of effort solving missing data, confounding effects, and DTR model diagnostics.

Our research work has some limitations, we thought they are easy to be met in observational data analysis and deserve future attention. First, in the PD field, Poewe[16] already pointed out that PD's progression cannot be completely summarized by motor scores, next stage work can try to think about how to combine non-motor scores to build up a composite primary outcome. Then is the imputation methods, the missing data problem becomes much more terrible in DTR's multi-stage setting because we need to consider variables' relationship across stages, we had better compare performances of different kinds of imputation method. Next, in model diagnostics, the number of candidate models seems a little small, and the difference among them is not significant, that is to say, all candidates have about 10 variables, one or two variable's addiction or deletion may not affect the results with great effect. Finally, the most serve problem is the non-regularity of our results, to make clear of it in observational studies still requires a lot of hard work.

In the process of this project, we found at least two directions for future researches. One is variable selection, remember that dWOLS is a regression-based technique, thus the idea that adding penalty terms for variable selection naturally comes into mind, and we are happy to see this idea is under development[3]. Another is the Matthew effect we mentioned above.

Optimal DTR building methods grew fast in the last decade, we believed that its combination with observational data analysis will be promoted both in theorem researches and industrial applications.

5. Supplementary Material

The R code is stored in: https://github.com/Zi-Yi-ZHANG/NUS_Research_Program.

REFERENCES

[1] AKAIKE H. Information Theory and an Extension of the Maximum Likelihood Principle [Z]. Springer Series in Statistics. Springer New York. 1998: 199-213.10.1007/978-1-4612-1694-0_15

[2] BELLMAN R. Dynamic Programming [J]. Science, 1966, 153(3731): 34-7.

[3] BIAN Z, MOODIE E E, SHORTREED S M, et al. Variable Selection in Regression-based Estimation of Dynamic Treatment Regimes [J]. arXiv preprint arXiv:210107359, 2021.

[4] CHAKRABORTY B, MOODIE E E M. Statistical Reinforcement Learning [Z]. Statistical Methods for Dynamic Treatment Regimes. Springer New York. 2013: 31-52.10.1007/978-1-4614-7428-9_3

[5] CHAKRABORTY B, MURPHY S A. Dynamic Treatment Regimes [J]. Annual Review of Statistics and Its Application, 2014, 1(1): 447-64.

[6] CHEN G, ZENG D, KOSOROK M R. Personalized Dose Finding Using Outcome Weighted Learning [J]. Journal of the American Statistical Association, 2016, 111(516): 1509-21.

[7] GOETZ C G, TILLEY B C, SHAFTMAN S R, et al. Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): Scale presentation and clinimetric testing results [J]. Movement Disorders, 2008, 23(15): 2129-70.

[8] HELY M A, MORRIS J G L, REID W G J, et al. Sydney multicenter study of Parkinson's disease: Non‐ L ‐dopa‐responsive problems dominate at 15 years [J]. Movement Disorders, 2004, 20(2): 190-9.

[9] MA L-Y, TIAN Y, PAN C-R, et al. Motor Progression in Early-Stage Parkinson's Disease: A Clinical Prediction Model and the Role of Cerebrospinal Fluid Biomarkers [J]. Frontiers in Aging Neuroscience, 2021, 12.

[10] MAHAR R K, MCGUINNESS M B, CHAKRABORTY B, et al. A scoping review of studies using observational data to optimise dynamic treatment regimens [J]. BMC Medical Research Methodology, 2021, 21(1).

[11] MOODIE E E M, CHAKRABORTY B, KRAMER M S. Q-learning for estimating optimal dynamic treatment rules from observational data [J]. Canadian Journal of Statistics, 2012, 40(4): 629-45.

[12] MURPHY S A. Optimal dynamic treatment regimes [J]. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 2003, 65(2): 331-55.

[13] MURPHY S A. A generalization error for Q-learning [J]. 2005.

[14] NAIMI A I, MOODIE E E M, AUGER N, et al. Constructing Inverse Probability Weights for Continuous Exposures [J]. Epidemiology, 2014, 25(2): 292-9.

[15] Parkinson Study Group. Levodopa and the progression of Parkinson's disease [J]. New England Journal of Medicine, 2004, 351(24): 2498-508.

[16] POEWE W. The natural history of Parkinson’s disease [J]. Journal of Neurology, 2006, 253(S7): vii2-vii6.

[17] RICH B, MOODIE E E M, STEPHENS D A. Optimal individualized dosing strategies: A pharmacologic approach to developing dynamic treatment regimens for continuous-valued treatments [J]. Biometrical Journal, 2015, 58(3): 502-17.

[18] ROBINS J M. The analysis of randomized and non-randomized AIDS treatment trials using a new approach to causal inference in longitudinal studies [J]. Health service research methodology: a focus on AIDS, 1989: 113-59.

[19] ROBINS J M. Causal Inference from Complex Longitudinal Data [Z]. Lecture Notes in Statistics. Springer New York. 1997: 69-117.10.1007/978-1-4612-1842-5_4

[20] ROBINS J M. Optimal Structural Nested Models for Optimal Sequential Decisions [Z]. Proceedings of the Second Seattle Symposium in Biostatistics. Springer New York. 2004: 189-326.10.1007/978-1-4419-9076-1_11

[21] ROBINS J M, HERNáN M Á, BRUMBACK B. Marginal Structural Models and Causal Inference in Epidemiology [J]. Epidemiology, 2000, 11(5): 550-60.

[22] ROSENBAUM P R, RUBIN D B. The central role of the propensity score in observational studies for causal effects [J]. Biometrika, 1983, 70(1): 41-55.

[23] RUBIN D B. Randomization Analysis of Experimental Data: The Fisher Randomization Test Comment [J]. Journal of the American Statistical Association, 1980, 75(371): 591.

[24] SCHRAG A, SIDDIQUI U F, ANASTASIOU Z, et al. Clinical variables and biomarkers in prediction of cognitive impairment in patients with newly diagnosed Parkinson's disease: a cohort study [J]. The Lancet Neurology, 2017, 16(1): 66-75.

[25] SCHULZ J, MOODIE E E M. Doubly Robust Estimation of Optimal Dosing Strategies [J]. Journal of the American Statistical Association, 2020, 116(533): 256-68.

[26] SUTTON R S, BARTO A G. Reinforcement learning: An introduction [M]. MIT press, 2018.

[27] THAVANESWARAN A, LIX L. Propensity score matching in observational studies [J]. Manitoba Center for Health Policy Retrieved from: https://www.umanitoba.ca/faculties/health_sciences/medicine/units/chs/departmental_units/mchp/protocol/media/propensity_score_matching.pdf, 2008.

[28] van den HEUVEL L, EVERS L J W, MEINDERS M J, et al. Estimating the Effect of Early Treatment Initiation in Parkinson's Disease Using Observational Data [J]. Movement Disorders, 2020, 36(2): 407-14.

[29] WAGNER E H, AUSTIN B T, DAVIS C, et al. Improving Chronic Illness Care: Translating Evidence Into Action [J]. Health Affairs, 2001, 20(6): 64-78.

[30] WALLACE M P, MOODIE E E M. Doubly-robust dynamic treatment regimen estimation via weighted least squares [J]. Biometrics, 2015, 71(3): 636-44.

[31] WALLACE M P, MOODIE E E M, STEPHENS D A. Model assessment in dynamic treatment regimen estimation via double robustness [J]. Biometrics, 2016, 72(3): 855-64.

[32] WALLACE M P, MOODIE E E M, STEPHENS D A. Dynamic Treatment Regimen Estimation via Regression-Based Techniques: Introducing R Package DTRreg [J]. Journal of Statistical Software, 2017, 80(2).

[33] WALLACE M P, MOODIE E E M, STEPHENS D A. Model validation and selection for personalized medicine using dynamic-weighted ordinary least squares [J]. Statistical Methods in Medical Research, 2017, 26(4): 1641-53.

[34] WATKINS C J C H. Learning from delayed rewards [J]. 1989.

[35] ZHANG B, TSIATIS A A, LABER E B, et al. A Robust Method for Estimating Optimal Treatment Regimes [J]. Biometrics, 2012, 68(4): 1010-8.

[36] ZHAO Y, ZENG D, RUSH A J, et al. Estimating Individualized Treatment Rules Using Outcome Weighted Learning [J]. Journal of the American Statistical Association, 2012, 107(499): 1106-18.

ACKNOWLEDGEMENTS

Thanks for all the kind support from my thesis professor Zhang Zhen.

Thanks for the guidance from Duke-NUS professor Bibhas Chakraborty and my supervised Ph.D. Yan Xiaoxi, it was a very nice experience in Singapore.

Thanks for the patient teaching and detailed mails by Waterloo professor Michael Wallace and McGill professor Erica EM Moodie, your selfless codeshare also helps me a lot.

THANKS to everyone I have learned from in past 4 years!