# ADJUSTING FOR TIME VARYING CONFOUNDING IN ADAPTIVE INTERDISCIPLINARY PAIN MANAGEMENT PROGRAM

By

NILABH OHOL [1]

Presented to the Faculty of the Graduate School of

The University of Texas at Arlington in Partial Fulfillment

of the Requirements for the Degree of

DOCTOR OF PHILOSOPHY

UNIVERSITY OF TEXAS AT ARLINGTON

August 2018

[1]Department of Industrial and Manufacturing Systems Engineering The University of Texas at Arlington Arlington, TX 76019 USA nilabh.ohol@mavs.uta.edu

## Acknowledgements

I would like to take this opportunity to thank my advisor, Dr. Victoria Chen, for giving me a platform to conduct this research. Working along with her has been an enriching and fulfilling experience. Needless to say, she has been the single most influential figures in shaping my professional life. I would like to thank my committee members, Dr. Jay Rosenberger, Dr. Aera LeBoulluec, Dr. Robert Gatchel, and Dr. Doyle Hawkins, for guiding me along the way. I wouldn't have been able to finish my work without their support. I would like to thank the support staff at the Department of Industrial Manufactuing and Systems engineering, namely. Julie Estill, Richard Zercher and Kimetha Williams for helping me with providing all the office/ administrative support. I would like to thank Dr. Li Zeng, Dr. Imrhan Sheik and Dr. Paul Componation for providing me academic guidance throughout. I would like to thank my fellow doctoral students at COSMOS who motivated me to work at the highest level. Their feedback and support was invaluable. I would like to thank the physicians and support staff at University of Texas Southwestern Medical Center who gave me the opportunity to work with them. I would like to thank my friends, especially the 'Usual Suspects ', expected nothing less than completion from me. I would like to thank my insanely large/ largely insane family who for taking the blows and giving me a chance to thrive. I would like to thank Aakanksha Jain, Shirish Rao, Gazi Iqbal, Ashkan Farahani, Upa Gupta, Swethaa Jayaraman for all their support. I would also like to thank all my Sulonian school friends, my soccer buddies at UT Arlington and the Arsenal Football club for making life as grad student very comfortable.

Last but not the least I would like to thank my parents and my sister for their support, love and patience. You guys are the best and I couldn't have done it without you!

# Abstract

ADJUSTING FOR TIME VARYING CONFOUNDING IN

ADAPTIVE INTERDISCIPLINARY PAIN MANAGEMENT PROGRAM

Nilabh Ohol

The University of Texas at Arlington, 2017

Supervising Professor : Victoria C. P. Chen

Interdisciplinary pain management combines multiple disciplines of professionals to understand the biological and psychosocial factors causing a patient's pain and to determine the best treatments among many to administer. The Eugene McDermott Center of Pain at University of Texas at Southwestern Medical Center in Dallas runs a two stage adaptive interdisciplinary pain management program with the aim to improve current and future pain outcomes. The sequential treatment regime for the pain and the observational nature of data yield to time varying confounding and a form of endogeneity. This yields biased estimates of the treatment effects which is undesirable. Our adaptive interdisciplinary pain management framework employs state transition and outcome models estimated from actual patient data in the program. This research develops a framework based on the inverse probability of treatment weighting technique to address endogeneity while estimating state transition and outcome models. The parameter estimates obtained from these models are unbiased and can be interpreted as causal effects of the treatments.

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

In 1931 French medical missionary Dr. Albert Schweitzer wrote “Pain is a more terrible lord of mankind than even death itself.” Today, pain has become a universal disorder affecting not only individuals, but also an individual's family, friends and health providers, who provide support to deal with the physical and emotional consequences of pain (MedicineNet 1996). Just about everyone feels pain from time to time. When you cut your finger or pull a muscle, pain is your body's way of indicating that something is wrong. An injury causes acute pain because when the injury heals, you stop hurting. However, chronic pain is different, it stays with your body long after the injury. Doctors often define chronic pain as pain which lasts for 3 to 6 months or longer WebMD (2005). While acute pain is a normal sensation triggered in the nervous system to alert an individual to possible injury and the need to take care, chronic pain persists. Pain signals keep firing in the nervous system for weeks, months, even years. Common chronic pain complaints include headache, low back pain, cancer pain, arthritis pain, neurogenic pain which is pain resulting from damage to the peripheral nerves or to the central nervous system itself of Pain Medicine (2017). Advanced neuroimaging has shown that chronic pain, unlike acute pain, can cause structural changes in the brain that add to the risk of cognitive problems, as well as anxiety and depression SPINE-health (1999).

According to American Academy of Pain Medicine, more than 1.5 billion people in the world are affected by chronic pain. In the United States alone, about 100 million Americans are affected. Many chronic pain conditions affect older adults, among which two thirds are suffering from back pain Iqbal (2017). Approximately 3–4.5% of the global population suffers from neuropathic pain, with incidence rate increasing with age of Pain Medicine (2017). To understand the economic aspect of pain as a public health problem, the Institute of Medicine Report, *Relieving Pain in America: a Blueprint of Transforming Prevention, Care, Education and Research*, estimated costs of $560-

$635 billion annually. This is equivalent to $2000 for everyone living in the U.S. This includes the total incremental cost of health care due to pain, ranging from $261 to $300 billion, and $297-$336 billion due to lost productivity of Pain Medicine (2017). Chronic pain has adverse effects on the lifestyle of an individual. Apart from tremendous health care costs, rehabilitation, and loss of productivity, it creates emotional and financial burden on patients and their families. Hence, developing better pain treatments is the primary goal of all pain research being conducted by institutes.

The goal of pain management is to improve function, enabling individuals to participate in day-to-day activities. However, it is important to recognize that chronic pain usually cannot be cured, but it can be managed. Thus there has been a lot of emphasis on pain management techniques to enable patients to cope with chronic pain problems. The most common treatments for pain include analgesic pain relievers (aspirin, acetaminophen, and ibuprofen), acupuncture, anticonvulsants, antidepressants, migraine headache medicines, biofeedback, capsaicin, chiropractic, cognitive and behavioral therapy, counseling, COX-2 inhibitors, electrical stimulation, exercise, hypnosis, lasers, magnets, nerve blocks, opioids, physical therapy and rehabilitation, R.I.C.E. – Rest, Ice, Compression, and Elevation, and surgery MedicineNet (1996). Patients feel and express pain differently. A patient's chronic pain experience depends on a lot of factors, which motivates the need to construct individually tailored programs. Recent studies have resulted in the evolution of treating chronic pain with not only analgesic medications but also using cognitive behavioral techniques to help patients cope with pain. To further help mitigate the pain,the treatment regime is monitored in stages to help the team of physicians track the pain status of a patient and enable potential modification in the treatment regime. The idea here is to dynamically assess the effects of treatment. In particular, this dissertation research is based on the two-stage adaptive interdisciplinary pain management program at the Eugene McDermott Center for Pain Management at the University of Texas Southwestern Medical Center at Dallas, (https://utswmed.org/conditions-treatments/pain-management/) which offers both interventional and non-interventional treatments. Interdisciplinary pain management involves treatment by a team of physicians, behavioral medicine specialists, physical therapists, nurses and care coordinators. This team works together to provide a variety of treatment plans and strategies to manage pain and improve quality of life (Center 2017). Moreover, in such scenarios pain is difficult to measure since different patients respond differently to pain. Therefore multiple pain outcomes are used to accurately measure a patient's pain. A detailed description of the pain management program

is given in detail in chapter 3. The pain management program uses an interdisciplinary treatment strategy to reduce the effect of chronic pain. This treatment strategy is monitored over two stages and is tailored to adapt to individuals in the program.

Given complex treatment environments like the Eugene McDermott Center for Pain Management at the University of Texas Southwestern Medical Center at Dallas, it has become increasingly important for physicians to develop personalized treatment plans for individual patients. Thus an adaptive treatment decision model would help assist physicians to carve out a treatment regime. Adaptive treatment strategies are a set of decision rules which can be adjusted depending on a patient's state (Murphy 2003). In such an adaptive treatment scenario the patient variables in one stage are influenced by the treatments prescribed in an earlier stage, which in turn influence the treatments that would be prescribed for the next stage. Such type of causal patternbetween patient variable across different stages results in biased estimation of the true effect of the treatments. Thus, it is difficult to clearly identify the causal effect of an individual treatment in alleviating pain, which is the key research question in epidemiology studies. This complication in the healthcare environment is referred to as endogeneity or time varying confounding. Previous work addressing endogeneity considers a single binary treatment and a binary outcome variable (Imai and Ratkovic (2013), Vansteelandt and Daniel (2014), Daniel et al. (2013)), thus, estimating treatment effects in such cases is straightforward using adjustments to logistic regression methods. However, the pain management problem contains a mix of several binary, categorical and continuous variables. The treatment variables are binary or ordinal, and the multiple outcome variables are continuous. The patient variables causing confounding are a mix of binary, categorical and continuous variables. Thus, under a multiple treatment environment and considering the fact that the not many patients are registered in the program makes statistical interpretation more challenging. In this dissertation, a general procedure is proposed to handle the complexity of this problem. Previous work on addressing the pain management case study at Eugene McDermott Center for Pain Management at the University of Texas Southwestern Medical Center at Dallas has been conducted by Lin et al. (2013),Wang (2015),Iqbal (2017). LeBoulluec (LeBoulluec et al. 2013), (LeBoulluec et al. 2018) developed an inverse probability treatment weighted method to address endogeneity. This work specifically addresses multiple treatment variables that are uncorrelated.

This dissertation develops a general framework to adjust for time varying confounding in studies with multiple treatments, while specifically seeking to address the McDermott Center pain management case study. To develop the framework, we look into various simulation studies to gain inference on the different conditions under which a modification to an existing framework developed using inverse probability of treatment weighting (IPTW) is required. Results from the simulation study are used to further develop the framework to address implementation of an IPTW framework the McDermott Center case study. The rest of this thesis is structured as follows. Chapter 2 discusses a paper on simulation study to handle time varying confounding with multiple treatments. It briefly describes the theory and motivation behind developing an IPTW framework with multiple treatments. The case studies begin with by covering an example with independent treatments and then generalizes to discussing the framework with correlated treatments. Chapter 3 presents a paper that develops and implements an IPTW framework for the McDermott Center case study. It focuses on discussing algorithms to estimate the joint probability of treatments and the generation of weights. Chapter 4 concludes the PhD dissertation by discussing conclusions from the two papers and stating future work.

# Bibliography


Center, U. M. (2017). Pain management.

Daniel, R., Cousens, S., De Stavola, B., Kenward, M., and Sterne, J. (2013). Methods for dealing with tine-dependent confounding. *Statistics in Medicine*, 32:1584–1618.

Imai, K. and Ratkovic, M. (2013). Covariate balancing propensity score. *Journal of Royal Statistical Society*, 76:243–263.

Iqbal, G. M. D. (2017). Multi-objective two-stage stochastic programming for adaptive interdisciplinary pain management with piece-wise linear network transition models. Ph.D. Dissertation, Department of Industrial and Manufacturing Systems Engineering, The University of Texas at Arlington, Arlington, Tx.

LeBoulluec, A., Ohol, N., Chen, V., Zeng, L., Rosenberger, J., and Gatchel, R. (2018). Handling time-

varying confounding in state transition models for dynamic optimization of adaptive interdisciplinary pain mangement. *IISE Transactions on Healthcare Systems Engineering*, 0:1–10.

LeBoulluec, A., Zeng, L., Chen, V., Rosenberger, J., and Gatchel, R. (2013). Outcome and state transition modeling for adaptive interdisciplinary pain management. Proceeding of 2013 Industrial and Systems Engineering Research Conference (ISERC).

Lin, C., LeBoulluec, A., Zeng, L., Chen, V., and Gatchel, R. (2013). An adaptive pain management framework. *Health Care Management Science*, 17(3):270–283.

MedicineNet (1996). Chronic pain.

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

of Pain Medicine, A. A. (2017). Aapm facts and figures on pain.

SPINE-health (1999). Chronic pain as a disease: Why does it still hurt?

Vansteelandt, S. and Daniel, R. (2014). On regression adjustment for the propensity score. *Statistics in Medicine*, 33:4053–4072.

Wang, N. (2015). Two-stage stochastic programming for adaptive interdisciplinary pain management. Ph.D. Dissertation, Department of Industrial and Manufacturing Systems Engineering, The University of Texas at Arlington, Arlington, Tx.

WebMD (2005). Do i have chronic pain.

# Chapter 2

# A Study on Time-varying Confounding with Multiple Treatments


## Abstract

Time varying confounding plays a critical role in longitudinal studies, for which data are collected over a period of time. In medical research, estimating an effect of a treatment on an outcome of interest is biased due to presence of time varying confounders. This bias results in inconsistent estimates of treatment effects. Further, in adaptive treatment strategies, not only is an unbiased estimate of the treatment effect desired, but also the outcome of interest is dependent on interactions between the treatment and covariates. Most of the literature on handling time varying confounding demonstrates methods, such as inverse probability of treatment weighting and g-computation, to obtain consistent estimates for a single treatment. This paper extends these methods to multiple treatments and, using a simulation study, highlights the challenges faced in estimating these treatment effects. This paper concludes with suggestions a researcher might consider while implementing these techniques to answer study specific questions.

## 2.1 Introduction

### 2.1.1 Motivation

Longitudinal studies, for which data are collected over time, are common in medical research. These studies are usually observational, meaning that the treatments are not controlled during the study. This is different from randomized controlled studies, in which treatments are specifically manipulated according to an experimental design. Controlled clinical trials are known to be expensive to conduct, highlighting to need to make use of observational studies. An important research question in epidemiology is to estimate a causal effect of treatment on the outcome of interest. While controlled studies are designed to answer this question, longitudinal studies, by contrast, are subject to time varying confounding when a treatment administered changes over time and is affected by other time varying covariates. For example Daniel et al. (2013) explains time varying confounding in a study involving type II diabetic patients recruited to compare two antiglyceamic drugs affecting the risk of a cardiac event in a period of 18 months. Covariates such as $HbA_{1c}$, blood pressure, body mass index, cholesterol level etc. are measured. Daniel explains that an increase in drug dosage or switching to a new drug may be warranted due to high $HbA_{1c}$, but a high value may also indicate a higher risk of cardiac arrest. Thus $HbA_{1c}$ is a confounder and since it changes over time and cannot be forecasted at baseline, it is a time varying confounder. In another example, Robins et al. (2000) highlights time varying confounding in a study of the effect of zidovudine (AZT) treatment on mortality in HIV - infected patients. The time dependent covariate in the study, CD4 lymphocyte count is both an independent predictor of survival and initiation of therapy with AZT and is itself influenced by prior AZT treatment. In the above examples, standard methods such as regression and propensity score fails to estimate unbiased effect of the treatment. Robins and Daniels show the use of inverse probability of treatment weights (IPTW) provide an effective way to address this time varying confounding issue. Other methods such g-computation and g-estimation using structural nested models (SNMs) are also effective in addressing this issue in small dimensional problems and when model is correctly specified. We discuss more on them in the next section.

However both these studies aim at estimating causal effect of a single treatment on an outcome realized

at the end of the study. In some medical research multiple treatment options are available to the physician in every stage of the study. These treatments form a complex relation with each other and with their treatment history while being affected by fixed and time varying covariates. The available literature on IPTW and g-computation does not demonstrate its application under such scenarios. Moreover, the treatments and covariates may be a mixture of types, such as binary, categorical and continuous variables. This motivates us to developing a framework to study methods for time varying confounding with multiple treatments in an adaptive treatment setting.

The remainder of the paper sequentially builds up a foundation to handle the aforementioned problem. Section 2 highlights the existing literature, discusses key concepts relating to time varying confounding, and defines our notation. Robins (1986) was the first to suggest g-computation for estimating the causal effect of a treatment. The paper also discusses two more techniques, inverse probability of treatment weighting using marginal structural models and g-estimation using structural nested models. Albeit this paper focuses on inverse probability of treatment weighting, section two discusses g-computation and g-estimation. We end section two enlisting contributions of this paper.

In sections 3 we explain time varying confounding using mathematical models. Section 4 discusses inverse probability of treatment weighting using different case studies. Section 3 also describes the factorization approach of LeBoulluec et al. (2013) to handle correlated treatments, and this method is used to illustrate our framework. The case study beings with an assumption of independence among the different treatments available in a study. We then increase the complexity of the problem by introducing correlations, increasing the size of the model, adaptive treatment strategy constraint and so on. Case study ends with a comparison of results with some of the existing methods.

The aim of the paper is to introduce a framework to handle time varying confounding using inverse probability of treatment weights for multiple treatments. To show this framework is flexible and can be modified depending on the description of problem and what research question one wishes to answer, the paper highlights different formulations of the IPTW frameworks and applies them on various case studies. Sections 5 and 6 summarizes the results of the case study and draw conclusions to give more insight to study longitudinal data in medical research.

## 2.2 Background and Literature

### 2.2.1 Definitions and Notation

This section defines notation that is used throughout the paper. Consider a multi-stage study involving $n$ patients, and let $i = 1, ..., n$ denote $i^{th}$ patient in the study. Let, $\tilde{T} = \{T_1, ..., T_t\}$ represent a set of treatments available to a physician to administer in each stage. $T_{itk}$ be the $t^{th}$ treatment administered in stage $k$ to patient $i$, where $k = 0, ..., K$. Thus, $\mathbf{T_{tk}}$ is the vector of $t^{th}$ treatment administered in stage $k$. In observational study, unlike controlled clinical trials where treatments are randomized across patient characteristics, the effects of covariates could be significant. Let $\tilde{X} = \{X_1, ..., X_m\}$ be the set of covariates in the study. These covariates can be either time fixed (e.g. age, gender) or time Varying (e.g. number of patient visits, blood glucose level etc.). Time fixed covariates are measured at baseline, stage 0, and remain fixed throughout the study. Time dependent covariates may or may not change in subsequent stages.

Let $Y_{ik}$ be the outcome measure for $i^{th}$ patient in stage $k$, therefore, $\mathbf{Y_k}$ represents vector of patient outcomes in stage $k$. It is also possible that there are multiple outcome variables of interest (Wang (2015), Iqbal (2017), Rawat and Manry (2017)), however in this paper, we assume a single outcome variable is considered per patient. The outcome $Y_{ik}$ could be continuous, categorical or binary and is either measured at the end of every stage or only at the end of the study. In cases where the outcome is measured at the end of every stage, a pre-evaluation outcome in stage 0 is also defined, where $\mathbf{Y_0}$ represents the vector of patient pre-evaluation outcomes measured, and $\gamma = \{\mathbf{Y_0}, \mathbf{Y_1}, ..., \mathbf{Y_K}\}$ is then the outcome history.

Similarly, $t_{i10}$ is the value for treatment 1 administered to patient $i$ in stage 0, which be continuous (e.g., drug dosage in ml), categorical (e.g., specific type of a drug group) or binary (e.g., yes or no). This initial value could either be randomly assigned in cases where a simulation study is designed or, in some longitudinal studies, could be attained based on a patient's physical characteristics. Let $\tau_t = \{\mathbf{T_0}, \mathbf{T_1}, ..., \mathbf{T_K}\}$ the vector of treatment $t$'s history assigned in every stage. Thus, $\bar{\tau}_{tk-1}$ represents the treatment history at the end of stage $k-1$ for treatment $t$. A similar notation is used for covariates, with the covariate history for time fixed covariates taking on same value in all the stages. Let $x_{i10}$ represent covariate 1 in stage 0 for patient $i$. Like treatments, covariates could be take binary, continuous or categorical.

In the paper, we use causal diagrams to graphically represent the potential causal effects between variables. A treatment $T_t$ is said to have a causal effect on an outcome in any stage $k$, if there exists at least one patient $i$ in the population and at least two distinct treatment values $T_t^1$ and $T_t^2$, such that $Y_i(T_t^1) \neq Y_i(T_t^2)$. Alternatively, in any stage $k$, if the outcome variable is not affected by any value of treatment $T_t$ then there exists no causal effect of treatment $T_t$ on outcome $Y$. (Daniel et al. 2013) and (Robins et al. 2000) elaborate more on causal effects and explain direct and total causal effect of a single treatment in structural nested models. In the case of multiple treatments, the definition of causal effects can be modified depending on the between treatment relationships. When the treatments in $\tilde{T}$ are independent of each other, every treatment in the set will have a causal effect on the outcome $Y$ if there exists, for that treatment $T_t$, at least one patient in the population and at least two distinct treatment values that have different effects on the outcome, $Y_i(T_t^1) \neq Y_i(T_t^2)$. If all or some of the treatments in the set $\tilde{T}$ are correlated to each other, then there is a causal effect of that set of correlated treatments, $\tilde{T}_c$, on the outcome $Y$. If there exist at least one patient $i$ in the population and at least two distinct set of treatment values, $\tilde{T}_c^1$ and $\tilde{T}_c^2$ such that $Y_i(\tilde{T}_c^1) \neq Y_i(\tilde{T}_c^2)$. When all the treatments in the study are correlated to each other, then $\tilde{T} = \tilde{T}_c$. In terms of multiple treatments, we define the cumulative effect of treatments on the outcome as the joint effect and their individual effects as marginal effects. This is especially helpful when researchers are interested in estimating treatment effects on outcomes realized in each stage of a multi-stage study. For example, consider two treatments in a study, $\tilde{T} = \tilde{T}_c = \{T_1, T_2\}$ such that $T_1 = \{0, 1\}$ and $T_2 = \{0, 1, 2\}$, that are correlated. If there is a patient $i$ in the population where, hypothetically, two distinct treatment combinations $\tilde{T}_c^1 = \{T_1^1, T_1^2\} = \{1, 1\}$ and $\tilde{T}_c^2 = \{T_2^1, T_2^2\} = \{0, 2\}$ have different outcomes, i.e., $Y_i(\{1, 1\}) \neq Y_i(\{0, 2\})$, then the treatments have a joint causal effect on the outcome. Here, $\tilde{T}_c^1$ and $\tilde{T}_c^2$ are joint effects of the treatments on an outcome in a particular stage. The relation between joint and marginal effects under independence and correlation scenarios is examined using our simulation study framework in sections 3 and 4.

Figure 2.1 shows a two-stage causal diagram under static and dynamic regimes. A static regime is when the treatment administered is not dependent on covariates or outcomes. Dynamic regime are those where treatments that potentially depend on covariates and outcomes occurring over time. For example, treating a

patient until there are signs on the patient becoming anemic and then stopping treatment can be considered as dynamic regime. Robins and Hernan have pioneered the work on dynamic treatment regime and their work along with their collaborators (Robins et al. (2008), Robins (2004), Hernan et al. (2006), Orellana et al. (2010)) The casual diagrams for both static and dynamic treatments consist of nodes representing the variables in the study and arrows representing the potential causal effect of a parent variable on a child in the direction of the arrow.

The casual diagrams also show the joint effects of the treatments on the outcome. We explore their marginal effects in detail later. For static regimes, it can be seen in Figure 2.1 a, that the set of covariates, $\tilde{X}$ does not affect the treatments administered in stages 1 and 2; however, they do have an effect on the outcome. The outcome variable is only measured at the end of the study, i.e., at the end of stage 2. The joint effects of the treatments in every stage affects the outcome $Y_2$ and is affected by the joint effects of treatments in the previous stage. Figure 2.1 b represents a two-stage dynamic regime where the covariates confound the effects of treatments on the outcome. Moreover, the outcomes are measured at the end of every stage. For instance, the baseline treatments $\tilde{T}_0$ (parent), measured at the beginning of the study or in stage 0, affect and outcome $Y_0$ (child) measured in that stage and the outcome $Y_1$ at the end of stage 1. The outcome $Y_0$ also affects the treatment $\tilde{T}_1$ administered in stage 1. $Y_0$ in this case is an intermediate variable since it functions as both an outcome and as a time-varying confounder. The time-fixed covariate $\tilde{X}$ also affects the treatments and outcome in every stage. Thus, causal diagrams help illustrate such complex relationship of an observational study.

$U_0$ is defined as unmeasured confounder (Hernan et al. 2004). In an observational study, it could happen that certain factors affecting the treatments and outcome are not included or might be unknown. The effects due to these factors may skew the estimated effects away from the true effects. These factors are known as unmeasured confounders. One of the key assumptions in observational studies is the presence of no unmeasured confounders. The no unmeasured confounding or conditional exchangeability assumption was first stated by Rubin (1978) for the single time-point setting and then extended by Robins (1986) under time varying confounding. It states that conditional on treatment and covariate history, the treatment received in the current stage is independent of outcome effect. That is, if the treated and untreated patients for a binary

treatment where exchanged, the distribution of the outcome $Y$ will be unaffected since it does not depend on any unmeasured factors. Besides unmeasured confounding, other assumptions are also considered, but not explicitly mentioned while conducting the analysis. For example, while selecting the sample size for the study, it is assumed that the $n$ patients have been selected randomly and independently from an infinite population of $N$ individuals. Also in order to well define the potential outcome $Y_i^{\tilde{\tau}}$ which follows a certain treatment trajectory $\tilde{\tau}$, no inference assumption is made. According to this assumption, setting the treatment trajectory of a different patient $i'$ does not affect potential outcome for patient $i$ Daniel et al. (2013). Finally, we define a consistency assumption which states that the actual outcome is equal to the potential outcome under the treatment administered Rubin (2009). This helps us make sensible inferences about the estimated outcome from the observational data. These definitions and assumptions form the backbone of the analysis under time varying confounding.

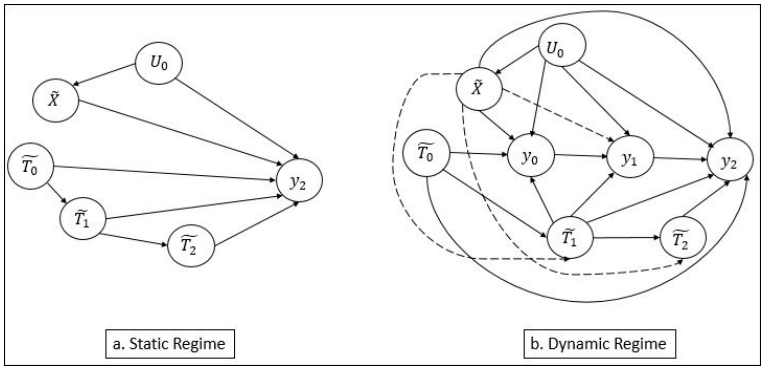



Figure 2.1: Causal Diagrams in Static Regime and Dynamic Regime

### 2.2.2 Time-varying Confounding

Robins, Hernan and others (Robins et al. (2008), Hernan et al. (2006), Hernan et al. (2001), Robins (1999)) in their work define the notion of time varying confounding using mathematical models and causal diagrams. Robins (1999) explains in detail the concept of statistical exogeneity and introduces IPTW to estimate unbiased treatment effects and applies the method to a study of zidovudine (AZT) treatment on mortality HIV - infected patients. Rubin (1978) addresses confounding effects of covariates in an experimental study using the predictive distribution of values under different assignments of treatments. Rubin however estimates

these causal effects under a single time-point setting. Confounding under a single time-point, or in cases where the time factor is controlled, is called time fixed confounding (Greenland 1996). However, the literature explains these concepts using a single treatment and covariate and the outcome variable is measured only at the end of the study. In the case of multiple treatments, the inter-relationship between the treatments in a given stage causes additional confounding that needs to be addressed. Consider a study to measure the depression index in patients suffering from chronic pain. In pre-evaluation stage, i.e. stage $k = 0$,the variables measured at baseline are covariates, $\tilde{X}$ (e.g. age, marital status), treatments $\tilde{T}_0$ (NSAIDs, psychotherapy) and outcome $Y_0$ (depression index). Hypothetically, there could be a case where older patients who are either single or divorced and suffering from high level of pain are more likely to be given pain medication, such as NSAID. However, if NSAIDs are strongly and positively correlated with psychotherapy, then these patients are more likely to be sassigned sessions of psychotherapy. Also, if these patients have also recorded a higher rate of depression and are currently under pain relieving medications, such as NSAID, then there is no causal effect of treatment on outcome. In such a case of time fixed confounding, intuitively it could be perceived that the psychotherapy and NSAIDs have a negative effect on the treatment. However in case of a correctly specified regression model, such as a multinomial logistic model for the depression index as outcome $Y$, estimating the conditional causal effect of $\tilde{T}_0$ on $Y$ would show no association. This is an example of time fixed confounding when treatments are correlated and can be shown using the causal diagram in Figure 2.2.

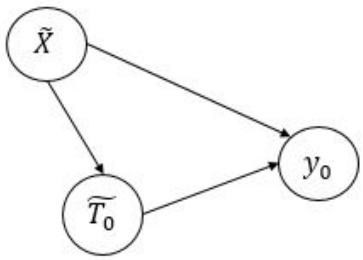



Figure 2.2: Causal diagram for time-fixed confounding

Time varying confounding is illustrated in Figure 2.3. Building on the above example, now consider stage $k > 0$. The treatments in stage $k$ are causally influenced by the treatments in the previous stage $k - 1$ and the covariates $\tilde{X}$. The outcome variable $Y_k$, e.g., depression index, is measured only at the end of the study.

In a naive analysis without adjusting for $\tilde{X}$, the effect of treatment histories $\bar{\tilde{\tau}}$ on $Y_k$ is confounded. It should be noted that in the causal diagram we could have shown arrows from $\tilde{T}_{k-2}$ to $\tilde{T}_k$. These are omitted to make the diagram readable. However the missing arrows from $\tilde{T}_{k-1}$ to $\tilde{X}$ represent no confounding effect of treatment in stage $k-1$ on covariates. By specifying an appropriate regression model with $Y_k$ as outcome and adjusting for $\tilde{X}$, consistent estimates of $\tilde{T}_k$ can be estimated (Daniel et al. 2013).

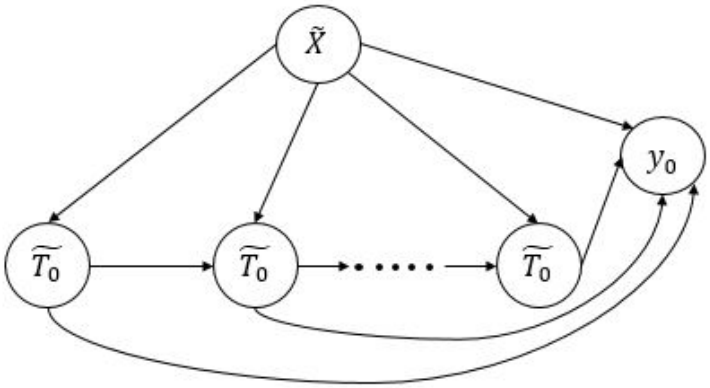



Figure 2.3: Causal diagram for time-dependent confounding

In Figure 2.3, the outcome variable was measured at the end of stage $k$. However, in some studies, the outcome is measured at the end of every stage. The causal diagram in Figure 2.1 b illustrates this case. Thus, the outcome variable, excluding the one measured at the end of stage $k$, can be considered as both confounder and outcome. For example, the outcome variable in stage 1 in Figure 2.1b is an outcome variable when estimating the causal effects of treatments $\tilde{T}_0$ and $\tilde{T}_1$ on $Y_1$. By constrast, an intermediate outcome is a time varying covariate when estimating treatment effects $\bar{\tilde{\tau}}$ on $Y_2$. The arrows from $\tilde{T}_0$ to $Y_1$ shows that it is affected by the past treatment. It must be noted that other time varying confounders may also be present in a study where they are not intermediate variables and serve as only confounders. Adjustment methods for controlling as described above may not work in this case. Controlling for the confounder $Y_1$ may induce serious bias since it lies in the causal pathway between treatment $\tilde{T}_{k-1}$ and outcome $Y_2$.

### 2.2.3 Handling time-varying confounding

A common approach to handle confounding while estimating treatment effects in the time fixed setting in Figure 2.3 is standardization (Vansteelandt and Daniel 2014). Adjusting for the covariates confounding the treatment effects on an outcome using standard approaches in a time varying setting can induce bias. This section describes generalized techniques used when treatments and covariates vary with time (Imai and Ratkovic 2013). G-estimation of structural nested models was first introduced by Robins (Robins (1992), Robins (1993), Rob (1994)) and is closely related to the g-test of the causal null hypothesis, which states that there is no causal effect of treatment in, say, two stages $T_0$ and $T_1$ on an outcome under the assumption of no unmeasured confounding.

$$T_0\{Y^{(t_0,t_1)}\}$$

and

$$T_1\{Y^{(t_0,t_1)}\}|T_0, \tilde{X}_1$$

This null hypothesis is tested by postulating models for $T_0$ and $T_1|T_0, X$. G-estimation is the inversion of the g-test Casella and Berger (2002), which searches for a causal effect for the set of potential outcomes, assumed to be conditionally independent of the time varying treatment, given their histories and covariates. These causal effects are characterized with a set of finite parameters using a class of models known as structural nested models, SNMs. SNMs consists of a set of sub-models, one for each stage k and each kth sub-model compares the potential outcomes (Daniel et al. 2013). Rob (1994) proposed SNMs for continuous outcomes, as follows,

$$E\{Y^{a_0,a_1,\ldots,a_{k-1},a_k,0,\ldots,0}|\bar{T}_{k-1,i} = \bar{t}_{k-1}, \bar{X}_{t,i} = \bar{x}_{t,i}\}$$

$$= E\{Y^{a_0,a_1,\ldots,a_{k-1},0,0,\ldots,0}|\bar{T}_{k-1,i} = \bar{t}_{k-1}, \bar{X}_{t,i} = \bar{x}_{t,i}\} + \psi_k(\bar{t}_k, \bar{x}_k; \phi_{\mathbf{k}}),$$

where $\psi_k(\bar{t}_k, \bar{x}_k; \phi_{\mathbf{k}})$ is the kth blip function (Daniel et al. 2013). Moodie et al. (2009) developed a decision framework using g-estimation for a breast feeding study to measure the effect of breastfeeding patterns in on infant growth. A 12-month period was considered with treatment being to either continue or stop breast feeding. The outcome variable was the height of an infant at the end of study. G-estimation

produces more efficient treatment estimates than other methods, such as inverse probability weighting (IPTW) because it allows incorporation of interactions between treatment and time varying covariates. It is also considered as a good compromise with regards to bias-variance trade off. However, g-estimation is not currently implementable using standard software in all settings. Albeit some software, like STATA and SAS, do have useful macros, they fail to address the whole range of SNMs. If the data are binary, there are no SNMs with estimable parameters to ensure that the probabilities lie between 0 and 1. G-estimation techniques are also computationally intensive, especially in case of multiple treatments, and numerous restrictions must be imposed on SNMs, increasing the possibility of model misspecification.

Another class of structural models is called marginal structural models (MSMs). These are used for parsimonious characterizations of causal effects. When the number of stages and/or treatments increases, the high dimensional nature of characterizations may cause difficulties in both estimation and interpretation. One reason could be because of shortage of patients following a particular treatment trajectory and too many comparisons of potential outcomes (Daniel et al. 2013). The equation below is an example of MSM,

$$E(Y^{\bar{t}}) = \varphi^{-1}\left(\varrho + \varsigma \sum_{k=0}^{K} t_k\right),$$

where $\varphi(\cdot)$ is a link function, e.g. log, logit or identity. $\varrho, \varsigma$ are the parameters of MSM and can be estimated using two techniques, g-computation and inverse probability of treatment weighting, while adjusting for time varying confounding.

G-computation was first introduced by Robins (1986) as an appropriate generalization of standardization, where treatments and covariates vary with time. A formal description of the g-computation formula is given by Robins (1986) and Daniel et al. (2013) to estimate the expected value of outcome based on the treatment history. Daniel et al. (2013) developed a g-computation formula function in Stata to estimate causal effects in the presence of time varying confounding or mediation using g-computation. The function is applied in a couple of simulation studies involving a single treatment. Robins in his work also defines the g-computation formulation for single treatment over multiple stages. We developed a modification to the g-computation formula to incorporate multiple treatments. The parametric and nonparametric versions are shown in the equations below.

The nonparametric form is

$$E\{Y^{(\bar{T}_1,\bar{T}_2)}\} = \sum_{\bar{\tilde{x}}\in\bar{\tilde{X}}} \Big\{ E(Y|\bar{T}_1=\bar{t}_1,\bar{T}_2=\bar{t}_2,\bar{\tilde{X}}=\bar{\tilde{x}}) \prod_{k=0}^{K} P(X_k=x_k|\bar{T}_{1k-1}=\bar{t}_{1k-1},\bar{T}_{2k-1}=\bar{t}_{2k-1},\bar{\tilde{X}}_{k-1}=\bar{\tilde{X}}_{k-1}) \Big\}.$$

The parametric form is

$$E\{Y_k^{(\bar{T}_1,\bar{T}_2)}\} = \int\int_{\bar{\tilde{x}}\in\bar{\tilde{X}}}^{\bar{\tilde{y}}\in\bar{\tilde{Y}}} E(Y|\bar{T}_1=\bar{t}_1,\bar{T}_2=\bar{t}_2,\bar{\tilde{X}}=\bar{\tilde{x}},\bar{Y}_{k-1}=\bar{y}_{k-1}) \prod_{k=0}^{K} f_{Y_{k-1}|\bar{\tilde{T}}=\bar{\tilde{t}},\bar{Y}_{k-2}=\bar{y}_{k-2},\bar{x}}(Y_{k-1},\bar{Y}_{k-2},\bar{\tilde{T}},\tilde{X})d\bar{y} \prod_{k=0}^{K} f_{\tilde{X}_k}(\tilde{X_k})d\bar{x},$$

where $E(Y|\bar{T}_1=\bar{t}_1,\bar{T}_2=\bar{t}_2,\bar{\tilde{X}}=\bar{\tilde{x}})$ e.g. - A linear model with main effects and interactions can be fit to calculate the expected outcome

$f_{Y_{k-1}|\bar{\tilde{T}}=\bar{\tilde{t}},\bar{Y}_{k-2}=\bar{y}_{k-2},\bar{x}}(Y_{k-1},\bar{Y}_{k-2},\bar{\tilde{T}},\tilde{X})$, e.g., if the outcome is continuous, a joint probability density function (p.d.f.) can be obtained by fitting a linear model to obtain expected outcome conditioned on the outcome in previous stage, treatment history and covariates

$f_{\tilde{X}_k}(\tilde{X_k})$ e.g. - if the covariate is binary or categorical, predicted probabilities are obtained, else for continuous covariates p.d.f is calculated.

As can be seen from above equation, the g-computation formula fails to address any correlations that may exist among the treatments. However, these models can be used to estimate treatment effects on the outcome, under the assumption of independence among the treatments. Another drawback of using g-computation is model misspecification. G-computation is sensitive to model specifications and an incorrectly specified model, for example, missing interaction terms or high order polynomials, will result in biased estimates. Finally, the g-computation formula is computationally intensive and when considering continuous or categorical treatments and even binary treatments with many stages, interpretability becomes difficult because there are too many hypothetical interventions to compare.

IPTW has gained popularity in estimating causal effects of treatments in observational studies. IPTW is a probabilistic weighting technique which re-weights the subjects in the analysis mimicking a random assignment of treatments. Equation below defines IPTW for patient $i$. A stabilized version of IPTW is also used in cases where low variation in the weights is desired or when one wishes to study the effect of the weights without adjusting for some covariates.

Unstablized weights:

$$w_i = \frac{1}{\prod_{k=0}^{K} P(T_k|\bar{T}_{k-1}, \tilde{X})}.$$

Stabilized weights:

$$w_i = \frac{\prod_{k=0}^{K} P(T_k|\bar{T}_{k-1})}{\prod_{k=0}^{K} P(T_k|\bar{T}_{k-1}, \tilde{X})}.$$

IPTW was first used by Robins (1999) to study the effect of zidovudine (AZT) treatment on mortality rates in HIV - infected patients with lymphocyte count as a confounder. Apart from epidemiology studies, Theommes and Ong (2016) applied IPTW to a single binary treatment to study the effects of developmental tasks on emerging adulthood. Previous work on IPTW (Robins et al. (2000),Hernan et al. (2001),Joffe et al. (2004),Bodnar et al. (2004),Fewell et al. (2004),Cole and Hernn (2008),Garcia-Aymerich et al. (2008),Rubin (2009)) also assumes a single treatment in a multi-stage setting. LeBoulluec et al. (2018) was the first to implement IPTW for multiple treatments, specifically applying this approach to an adaptive two-stage pain management program. However, this work assumes that the treatments are independent. The current paper develops a framework of simulated case studies to enable a comprehensive study of IPTW for multiple treatments, under the conditions of both independence and correlations. We choose the IPTW framework due to its robustness to model misspecification, versatility in handling binary mix of variable types, computational feasibility, and its ability to adapt to modifications of MSMs. Although it has been observed that there can be inefficiency under extreme weights, we provide recommendations for overcoming these drawbacks to IPTW.

### 2.2.4 Contribution

The major contributions of this research are described as follows:

- This paper generalizes the IPTW method to multiple treatments in multiple stages, when the treatment outcomes are realized at the end of each stage.
- The generalized IPTW process is incorporated within a time varying confounding framework to allow a simulation study of different cases.

- The developed framework is the first to study the correlated treatments case. A step by step approach is provided . Simulation studies under such scenarios are presented and conclusions are drawn from them to help researchers pick the right steps while implementing the framework

## 2.3 Inverse Probability of Treatment Weighting for Multiple Treatments

### 2.3.1 Independent Treatments

Consider a study involving two binary treatments represented by a causal diagram in Figure 2.4 below. The treatments are administered in two stages, 1 and 2. It can be assumed that the initial treatment in stage 0 is either assigned randomly or suggests that the treatment had been prescribed in the past. In either case, the treatment in stage 0 has no association to its assignment, with respect to the ongoing study. The covariates, also known as patient state variables, are a mixture of categorical and binary variables and measured at baseline, i.e., in stage 0. They are assumed to be time-fixed. The outcome variable, in this study is continuous and is measured at the end of each stage. The outcome $Y_0$ is measured at the beginning of the study for the physicians to have a reference value to assist in treatment. We also assume no unmeasured confounders in the study. The covariates $\tilde{X}$ are not dependent on the treatments. However, they confound the treatment effects on the outcome in every stage. The outcome variable $Y_1$ measured at the end of stage 1 is an intermediate variable that acts as an outcome, when estimating stage 1 treatment effects, and acts as a time varying confounder affected by treatments from previous stages, when estimating treatment effects in stage 2. Outcome variable $Y_2$ occurs at the end of stage 2, and outcome variable $Y_0$ is a time varying confounder throughout the study since we do not model to estimate treatment effects in stage 0.

Suppose the two treatments are independent in each stage i.e. assignment of one treatment does not influence the other. Also the joint effect of the treatments on outcome in any stage is the sum of individual treatment effects. Thus, marginal structural model for the two independent treatments in stage 1 can be

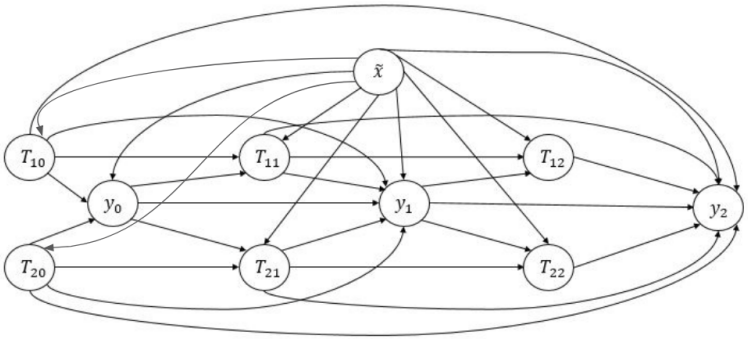



Figure 2.4: Causal Diagrams with Two Treatments

written as,

$$E\{Y^{(t_{10},t_{11},t_{20},t_{21})}\} = \gamma_{int} + \gamma_{10}t_{10} + \gamma_{20}t_{20} + \gamma_{11}t_{11} + \gamma_{21}t_{21} + \gamma_{1(01)}t_{10}t_{11} + \gamma_{2(01)}t_{20}t_{21}$$

The above model is marginal, because it represents the marginal distribution of outcome (unconditional on time varying confounders), and structural, because it is modeled for potential outcomes and not observed outcomes. The parameters of the models can be estimated using the following set of equations.

$$\gamma_{int} = E\{Y^{(0,0,0,0)}\}$$

$$\gamma_{10} = E\{Y^{(1,0,0,0)}\} - E\{Y^{(0,0,0,0)}\}$$

$$\gamma_{20} = E\{Y^{(0,0,1,0)}\} - E\{Y^{(0,0,0,0)}\}$$

$$\gamma_{11} = E\{Y^{(0,1,0,0)}\} - E\{Y^{(0,0,0,0)}\}$$

$$\gamma_{21} = E\{Y^{(0,0,0,1)}\} - E\{Y^{(0,0,0,0)}\}$$

$$\gamma_{1(01)} = E\{Y^{(1,1,0,0)}\} - E\{Y^{(1,0,0,0)}\} - E\{Y^{(0,1,0,0)}\} + E\{Y^{(0,0,0,0)}\}$$

$$\gamma_{2(01)} = E\{Y^{(0,0,1,1)}\} - E\{Y^{(0,0,1,0)}\} - E\{Y^{(0,0,0,1)}\} + E\{Y^{(0,0,0,0)}\}$$

Similarly, we can write MSMs for potential outcomes in stage 2. Generalizing the MSM equation for a distribution of potential outcome associated with a treatment trajectory can be described as a function of the trajectory and some parameters $\gamma$

$$E(Y^{\bar{T}_j}) = h(\bar{T}_j; \gamma)$$

However, it must be noted that these potential outcomes for all combinations of treatment trajectory are possible in the case of a simulated example. However, in observational data these MSMs are different from association models, such as in equation 2.1.

$$E(Y|\bar{T}_j = \bar{t}_j) = h(\bar{t}_j; \alpha) \tag{2.1}$$

Only under the strong assumption of no (measured/unmeasured) confounding, do the parameters of MSM and association models coincide.

The parameter estimates in the MSM are independent of the covariate and treatment history. We estimate these unconfounded parameters by analysis of the observational data after weighting each patient by weights obtained by IPTW framework. The inverse probability of treatment weights for a patient with independent treatments is given by,

$$w_i = \frac{1}{\prod_{j=1}^{t} \prod_{k=1}^{K} f_{T_{jk}|\bar{T}_{jk-1},\tilde{X}}(T_{jk}, \bar{T}_{jk-1}, \tilde{X})}, \tag{2.2}$$

where $k$ is the number of stages, $t$ number of treatments and $f_{T_{jk}|\bar{T}_{jk-1},\tilde{X}}(T_{jk}, \bar{T}_{jk-1}, \tilde{X})$ is the conditional probability mass function for $T_jk$ given $(\bar{T}_{jk-1}, \tilde{X})$ for patient $i$. The joint probability of independent treatments can be decomposed into the product of the marginal probabilities as shown below,

$$P(T_1, ..., T_t|\tilde{X}) = P(T_1|\tilde{X})P(T_2|\tilde{X})....P(T_t|\tilde{X})$$

Thus the IPTW equation with the two treatments can be expanded as,

$$w_i = \prod_{k=1}^{K} \frac{1}{P(T_{1k} = t_{1k}|\bar{T}_{1k-1}, \tilde{X})} \frac{1}{P(T_{2k} = t_{2k}|\bar{T}_{2k-1}, \tilde{X})} \tag{2.3}$$

The marginal probabilities can be estimated from the data by fitting appropriate models such as logistic model for binary treatment, multinomial for categorical and linear models for continuous treatments. Once the observations are 'duplicated'using the weights obtained, the treatments in the reweighted sample are independent of the treatment history and covariates. In terms of causal diagrams, the inward arrows into treatments disappear leaving only the arrows going outwards from the treatments. Intuitively, it can be thought of as the weights supplementing the observations, representing a pseudo-population in which the treatments are unconfounded by the covariates. Under the assumption of no unmeasured confounding, the parameter estimates obtained by the refitted weighted models are unbiased.

### 2.3.2 Correlated Treatments

When treatments are correlated, the parameter estimation using MSMs remains similar to the independent treatment case, where the confounding effects are addressed using the IPTW framework, and weighted MSMs are refit. However, the estimation of treatment weights needs modification to incorporate the correlation structure of the treatments. Consider two causal diagrams representing the same study as in previous section with difference being the correlation among treatments in a given stage. Figure 2.5 (a) shows the confounding factors affecting treatment 1. The dark arrows shows the confounding effects of treatment $T_2$ on $T_1$ in all stages. Similarly Figure 2.5 (b) shows the factors confounding treatment 2.

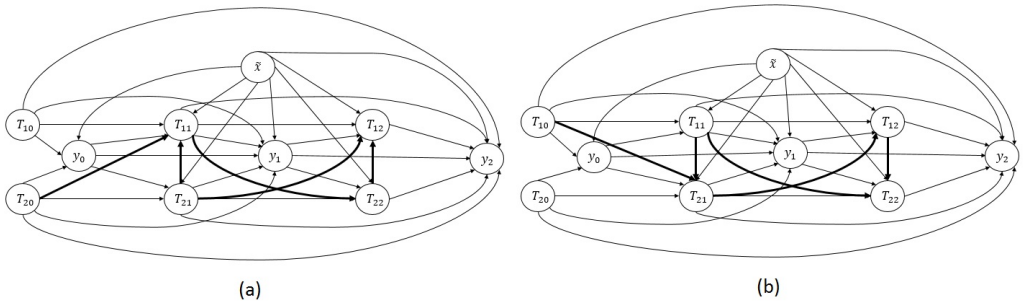



Figure 2.5: Causal Diagrams for Two Correlated Treatments over Two Stages

If one wishes to obtain parameter estimates of MSMs for potential outcomes similar to the independent treatment case, the model can be written as follows with the parameter estimates obtained using the subsequent set of equations.

$$\begin{aligned} E\{Y^{(t_{10},t_{11},t_{20},t_{21})}\} =& \gamma_{int} + \gamma_{10}t_{10} + \gamma_{20}t_{20} + \gamma_{11}t_{11} + \gamma_{21}t_{21} + \gamma_{1(01)}t_{10}t_{11} \\ &+ \gamma_{2(01)}t_{20}t_{21} + \gamma_{12(01)}t_{10}t_{21} + \gamma_{21(01)}t_{20}t_{11} + \gamma_{12(11)}t_{11}t_{21} \end{aligned}$$

Parameter estimates,

$$\gamma_{int} = E\{Y^{(0,0,0,0)}\}$$

$$\gamma_{10} = E\{Y^{(1,0,0,0)}\} - E\{Y^{(0,0,0,0)}\}$$

$$\gamma_{20} = E\{Y^{(0,0,1,0)}\} - E\{Y^{(0,0,0,0)}\}$$

$$\gamma_{11} = E\{Y^{(0,1,0,0)}\} - E\{Y^{(0,0,0,0)}\}$$

$$\gamma_{21} = E\{Y^{(0,0,0,1)}\} - E\{Y^{(0,0,0,0)}\}$$

$$\gamma_{1(01)} = E\{Y^{(1,1,0,0)}\} - E\{Y^{(1,0,0,0)}\} - E\{Y^{(0,1,0,0)}\} + E\{Y^{(0,0,0,0)}\}$$

$$\gamma_{2(01)} = E\{Y^{(0,0,1,1)}\} - E\{Y^{(0,0,1,0)}\} - E\{Y^{(0,0,0,1)}\} + E\{Y^{(0,0,0,0)}\}$$

$$\gamma_{12(01)} = E\{Y^{(1,0,0,1)}\} - E\{Y^{(1,0,0,0)}\} - E\{Y^{(0,0,0,1)}\} + E\{Y^{(0,0,0,0)}\}$$

$$\gamma_{21(01)} = E\{Y^{(0,1,1,0)}\} - E\{Y^{(0,1,0,0)}\} - E\{Y^{(0,0,1,0)}\} + E\{Y^{(0,0,0,0)}\}$$

$$\gamma_{12(11)} = E\{Y^{(0,1,0,1)}\} - E\{Y^{(0,1,0,0)}\} - E\{Y^{(0,0,0,1)}\} + E\{Y^{(0,0,0,0)}\}$$

For modeling observed outcomes and referring back to the IPTW equation 2.2, the conditional joint probability distribution of treatments can no longer be expressed as the product of the probability distribution of individual treatments conditioned on the treatments'past history and covariates. Since the treatments are correlated, equation 2.3 is no longer true. In Figure 2.5 (a), we try to estimate the effect of treatment $\bar{T}_1$ on the outcome variable. The inward arrows represent the causal relationship between treatment 2 and 1. Treatment 1 in any stage is not only affected by T2 in the same stage but also by its previous treatment history. A similar inference about $\bar{T}_2$ can be made from Figure 2.5 (b). It must be noted here that apart from correlation between treatments, these treatments also confound each other's effect on the outcome. For example, $T_{02}$ confounds the effect of treatment $T_{11}$ on stage 1 outcome $Y_1$. Thus while estimating the marginal probability distribution of one treatment, the other treatment is conditioned along with the covariates. Using the chain rule of probability the conditional joint probability of correlated treatments can be decomposed as the product of their conditional marginal probabilities. Equation 2.4 below states the chain rule of probability of $T$ treatments conditioned on a set of covariates $\tilde{X}$.

$$P(T_1, T_2, T_3, ..., T_t|\tilde{X}) = P(T_1|\tilde{X}) \cdot P(T_2|T_1, \tilde{X}) \cdot ... \cdot P(T_t|T_{t-1}, ..., T_1, \tilde{X}) \tag{2.4}$$

Thus the IPTW framework for two correlated treatments for the case study at hand can be expanded as,

$$w_i = \prod_{k=1}^{2} \frac{1}{P(T_{1k} = t_{1k}|\bar{T}_{1k-1}, \tilde{X})P(T_{2k} = t_{2k}|T_{1k}, \bar{T}_{2k-1}, \tilde{X})}$$
$$= \frac{1}{P(T_{11} = t_{11}|\tilde{X})P(T_{21} = t_{21}|T_{11}, \tilde{X})}$$
$$\frac{1}{P(T_{12} = t_{12}|T_{11}, \tilde{X})P(T_{22} = t_{22}|T_{12}, T_{21}\tilde{X})}$$

The marginal probability of treatments can be estimated by fitting appropriate regression models such as logistic, multinomial or linear regression models depending on the data type of the treatment. The weights obtained are used to refit outcome models in a similar way as explained for independent case. A generalized version of the IPTW formulation for multiple treatment (independent or correlated) is given in equation 2.2.

A potential drawback of using the chain rule is the estimated joint probability of correlated treatment is sensitive to ordering. Consider a case with three correlated treatments, the number of ways of decomposing the joint probability into its product of marginal probabilities is $3! = 9$. Thus for T treatments, this number increases to $t!$. A subset of sample decomposition of three correlated treatment is shown below.

$$\begin{aligned} P(T_1, T_2, T_3|\tilde{X}) =& P(T_1)P(T_2|T_1, \tilde{X})P(T_3|T_2, T_1, \tilde{X}) \\ & P(T_2)P(T_1|T_2, \tilde{X})P(T_3|T_1, T_2, \tilde{X}) \\ & P(T_3)P(T_2|T_3, \tilde{X})P(T_1|T_2, T_3, \tilde{X}) \end{aligned}$$

In theory these probability distributions should be equal but due to error in estimation using standard software tools, these estimated probabilities are not same. Thus the estimated conditional joint probability of distribution used in the framework is an averaged value over all possible ordering as shown in equation below.

$$\hat{P}_1(T_1, ...T_t|\tilde{X}) \cdots \hat{P}_{t!}(T_1, ..., Tt|\tilde{X})$$
$$avg(\hat{P}) = \frac{\hat{P}_1 + ... + \hat{P}_{t!}}{t!}$$

**Factorizations**

Calculating average conditional joint probability requires estimating the product of their marginals over all possible ordering. We define a factorizations to be the product of estimated conditional marginal probabilities for one specific treatment ordering. Thus, a study with three correlated treatments has 9 possible factorizations, while one with 10 treatments has $10!$ possible factorizations. The computational complexity increases with the increase in the number of treatments and the factorization runs in time $O(t!)$. This might be undesirable especially in cases where a treatment plan is required be evaluated in real time. The computational complexity can be reduced by randomly sampling $f$ factorizations from the set of $T!$, where $f = \{p\%\ T!\}$ (LeBoulluec et al. 2013). To increase the robustness of the method, we create two groups $G_1$ and $G_2$ and randomly sample $f$ factorizations for each group. For every factorization under each group, the joint probability distribution is estimated, and a performance metric, mean relative percent relative difference (MPRD) shown in equation 2.5 is calculated. The hyper-parameters $f$ and $p$ are tuned by conducting a grid search and using the performance metric as a selection criterion.

$$MPRD = mean(\frac{|avg(\hat{P})_{G_1} - avg(\hat{P})_{G_2}|}{max(avg(\hat{P})_{G_1}, avg(\hat{P})_{G_2})})100. \tag{2.5}$$

The algorithm below explains the factorization approach:

1. Initialize $p$=0,$f_{G1}$=1,$f_{G2}$=1 and $m$=$t$!, where $t$ is the number of treatments
2. Initialize matrix mean percent relative difference, MPRD = $NULL$
3. Obtain a matrix $M$ of $m$ permutations of all correlated treatments
4. Sample($f_{G1}$ and $f_{G2}$ from $M$, repeat = $FALSE$)
5. for each $f$=$\{f_{G1}, f_{G2}\}$:
   (1) estimate $w_i$ $\forall i$=$1, ..., n$
   (2) calculate percent relative difference, $PRD_i$
   (3) $MPRD_f$ = $sum_i$(PRD)/n
   (4) Update matrix MPRD

6. update $p$ = $p + 0.05$

7. update $f_{G1}$=$f_{G2}$= $pm$

8. repeat $c$ to $f$; if($p$=100 or $MPRD_j$-$MPRD_{j+1}/MPRD_j < 0.1$)

Apart from the factorization approach, other techniques can also be used to estimate the joint probability of treatment. Some of them were mentioned earlier in section 3.

**Stabilized Weights**

Sometimes the estimated joint probabilities might have values close to zero or the estimated weights have high error variance. This may happen when a rare observation is present in the study. Under such cases the estimated weights can be very large and cause imprecision in the estimation of the parameters of MSM. Robins et al. (2000) suggested obtaining a stable estimator by re-weighting the samples using a stabilized version of IPTW. The stabilized IPTW for multiple treatments is shown in equation 2.6,

$$sw_i = \frac{\prod_{k=0}^{K} \prod_{j=1}^{T} P(T_{jk} = t_{jk} | \bar{T}_{jk-1})}{\prod_{k=0}^{K} \prod_{j=1}^{T} P(T_{jk} = t_{jk} | \bar{T}_{jk-1}, \tilde{X})}. \tag{2.6}$$

The denominator of the stabilized IPTW equation is same as the unstabilized version. The numerator is same as the denominator without adjusting for covariates. The stability and efficiency of the estimators under stabilized weights might be affected if the models for $T_k|\bar{T}_{k-1}$ are misspecified. However, the consistency is unaffected. It was mentioned earlier that an advantage of using IPTW estimators is their ability to be unaffected by minor model misspecification. This helps in obtaining fairly efficient parameter estimates using stabilized weights. More often, a researcher may be interested in estimating MSMs as a function of baseline covariates along with treatments. In such cases, the MSM defined in equation 2.1 is not applicable. The stabilized IPTW equation can be adjusted for $\tilde{X}_0 \subset \tilde{X}$ in both numerator and denominator. Table 2.1 shows some examples of modifications of MSMs and their respective IPTW formulation to address time varying confounding.

| Model Description | Model Form | Stabilized Weights |
|---|---|---|
| Outcome in stage $k$ given treatments administered | $E(Y^{\bar{T}_j}) = h(\bar{t}_j;\gamma)$ | $sw_i = \frac{\prod_{k=0}^{K}\prod_{j=1}^{T} P(T_{jk}=t_{jk}\|\bar{T}_{jk-1})}{\prod_{k=0}^{K}\prod_{j=1}^{T} P(T_{jk}=t_{jk}\|\bar{T}_{jk-1},\tilde{X})}$ |
| Outcome in stage $k$ given treatments administered and baseline covariates | $E(Y^{\bar{T}_j}\|\tilde{X}_0) = h(\bar{t}_j,\tilde{x}_0;\gamma)$ | $sw_i = \frac{\prod_{k=0}^{K}\prod_{j=1}^{T} P(T_{jk}=t_{jk}\|\bar{T}_{jk-1},\tilde{X}_0)}{\prod_{k=0}^{K}\prod_{j=1}^{T} P(T_{jk}=t_{jk}\|\bar{T}_{jk-1},\tilde{X})}$ |
| Outcome in stage $k$ given treatments administered and previous outcome | $E(Y^{\bar{T}_j}\|Y_{k-1}^{\bar{T}_j}) = h(\bar{t}_j, y_{k-1}^{\bar{T}_j};\gamma)$ | $sw_i = \frac{\prod_{k=0}^{K}\prod_{j=1}^{T} P(T_{jk}=t_{jk}\|\bar{T}_{jk-1},Y_{k-1}^{\bar{T}_j})}{\prod_{k=0}^{K}\prod_{j=1}^{T} P(T_{jk}=t_{jk}\|\bar{T}_{jk-1},\tilde{X})}$ |
| Outcome in stage $k$ given treatments administered, baseline covariates and previous outcome | $E(Y^{\bar{T}_j}\|\tilde{X}_0, Y_{k-1}^{\bar{T}_j}) = h(\bar{t}_j,\tilde{x}_0, y_{k-1}^{\bar{T}_j};\gamma)$ | $sw_i = \frac{\prod_{k=0}^{K}\prod_{j=1}^{T} P(T_{jk}=t_{jk}\|\bar{T}_{jk-1},\tilde{X}_0,Y_{k-1}^{\bar{T}_j})}{\prod_{k=0}^{K}\prod_{j=1}^{T} P(T_{jk}=t_{jk}\|\bar{T}_{jk-1},\tilde{X})}$ |

Table 2.1: Marginal Structural Models and Stablized IPTW under different scenarios

## 2.4 IPTW Simulation Study Framework

Table 2.2 summarizes the IPTW simulation study framework for assessing methods to address time varying confounding with multiple treatments.

This section discusses simulation case studies to demonstrate the functionality of the framework. We begin the section with small and simple cases and sequentially increase the complexity of the problem. At the end of each subsection, we highlight key takeaways from the case study to provide insights. As a

| Steps | Description |
|---|---|
| Step 1 | Identify treatments, patient variables, treatment outcomes in a study. |
| Step 2 | Identify number of stages. |
| Step 3 | Identify causal relationships between treatments, between treatments and time varying patient covariates, and from treatment to outcomes. |
| Step 4 | Check for independence among the treatments in each stage. |
| Step 5 | If the treatments are independent, use the IPTW framework for independent treatments described in section 3 for every treatment based on the indentified causal relationship |
| Step 6 | If the treatments are correlated, use the IPTW framework for correlated treatments described in section 3 for every treatment based on the indentified causal relationship |

Table 2.2: Marginal Structural Models and Stablized IPTW under different scenarios

general approach on the simulation studies, we start with describing the problem, provide guidelines for data generation, implement the framework and show the results. To maintain homoogeneity, for every simulation case study we use a two-stage problem with the outcome variables realized at the end of each stage.

### 2.4.1 Case Study: Independent Treatments

Consider a study with three independent treatments administered across two stages. The study mimics the causal diagram shown in Figure 2.4. There are three covariates confounding the effect of treatments on the outcome. Outcome variables are realized at the end of each stage. From the previous section, it has been established that any treatments administered to the patients before they are in the study are treated as covariates since they have no association with treatments in the present study. It is also assumed there are no unmeasured confounders. The data are generated and the true treatment effects are recorded. Below are the guidelines for data generation.

- A sample size of 1000 is considered for the study.
- Covariate $x_1$ follows a binomial distribution with mean 1 and standard deviation 0.5. In medical research, such a variable could signify, for example, if a patient has had any previous surgeries.
- Covariate $x_2$ and $x_3$ are continuous variables, normally distributed with mean 0 and standard deviation one. These could represent a pre-evaluation treatment score and duration of pain on a standardized scale.
- Treatment $t_1$, in stage 1, follows a binomial distribution and is a function of the three covariates with the probability of treatment defined by the equation,

$$P(t_{11} = 1) = \frac{exp(0.5x_1 + 0.3x_2 + 0.85x_3)}{1 + exp(0.5x_1 + 0.3x_2 + 0.85x_3))}.$$

- Similarly, treatments $t_2$ and $t_3$ in stage 1 are binomial and are functions of the three covariates with probabilities as follows:

$$P(t_{21} = 1) = \frac{exp(0 - 0.25x_1 + 0.3x2 + 0.85x_3)}{(1 + exp(0 - 0.25x1 + 0.3x_2 + 0.85x_3))}$$
$$P(t_{31} = 1) = \frac{exp(0.6x_1 + 0.1x_2 + 0.9x_3)}{(1 + exp(0.6x_1 + 0.1x_2 + 0.9x_3))}.$$

- The outcome variable in stage 1 is defined by the equation below. The coefficients associated with each variable define the effects of the variables on the outcome. For example, treatment 1 has a negative correlation with the pain outcome, intuitively, if a treatment is administered that is the expected to reduce pain. The true treatment effect is 0.6. We induce some noise in the outcome tht follows a normal distribution with mean 0 and standard deviation of 0.9.

$$y_1 = -0.1x_1 + 0.2x_2 - 0.1x_3 - 0.6t_{11} - 0.45t_{21} + \varepsilon(n, 0, 0.9).$$

- Treatments in stage 2 follow the same distributions as in stage 1. However, the probability of a stage 2 treatment is also dependent on the treatment administered in the previous stage and the pain outcome in stage 1. The following set of equations define the probability of treatments in stage 2:

$$P(t_{12} = 1) = \frac{exp(0.5x_1 + 0.3x_2 + 0.85x_3 + 0.85t_{11} + 0.5y_1)}{(1 + exp(0.5x_1 + 0.3x_2 + 0.85x_3 + 0.85t_{11} + 0.5y_1))}$$
$$P(t_{22} = 1) = \frac{exp(0 - 0.25x_1 + 0.3x_2 + 0.85x_3 + 0.65t_{21} + 0.5y_1)}{(1 + exp(0 - 0.25x_1 + 0.3x_2 + 0.85x_3 + 0.65t_{21} + 0.5y_1))}$$
$$P(t_{32} = 1) = \frac{exp(0.6x_1 + 0.1x_2 + 0.9x_3 - 0.45t_{31} + 0.5y_1)}{/(1 + exp(0.6x_1 + 0.1x_2 + 0.9x_3 - 0.45t_{31} + 0.5y_1))}.$$

The data are generated using the following guidelines, and the generalized IPTW process is applied to obtain treatment estimates in stages 1 and 2. We compare the true treatment effects, obtained from the data generation process, to the estimated treatment effects with and without weights. The estimated treatments effects are obtained by fitting a linear model as shown in the equation below.

$$y = f(X, U) + \varepsilon.$$

To obtain the treatment effects of weighted model, a weighted least squares approach is used. The analysis is conducted in R (R Development Core Team 2008). The weights generated using the framework are shown below.

In stage 1, there is no time varying confounding effect due to treatments. The variables confounding the effect of treatment on the outcome in stage 1 are the covariates. For this paper, because the focus is on the estimation of causal treatment effects, we consider the time varying confounding effect in stage 2. It can also be seen from the above plot that the weights are concentrated in the lower values between 1 and 2 with only a few patients weighted higher. When the parameter estimates for the unweighted and weighted

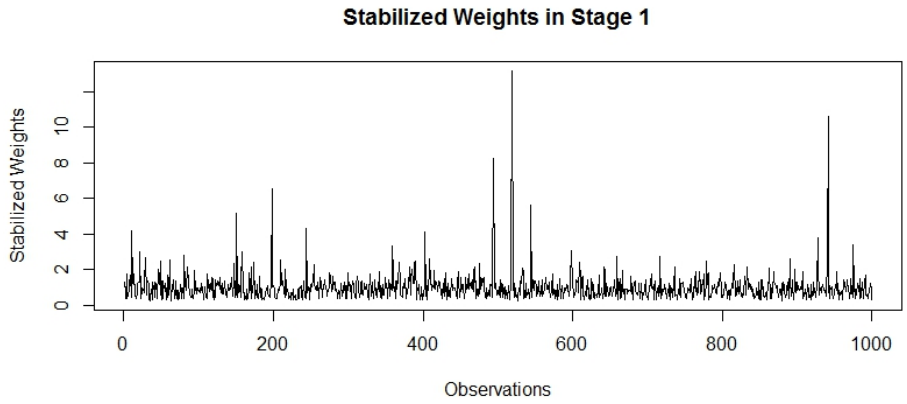



Figure 2.6: Stabilized weights for patients in stage 1 when treatments are independent

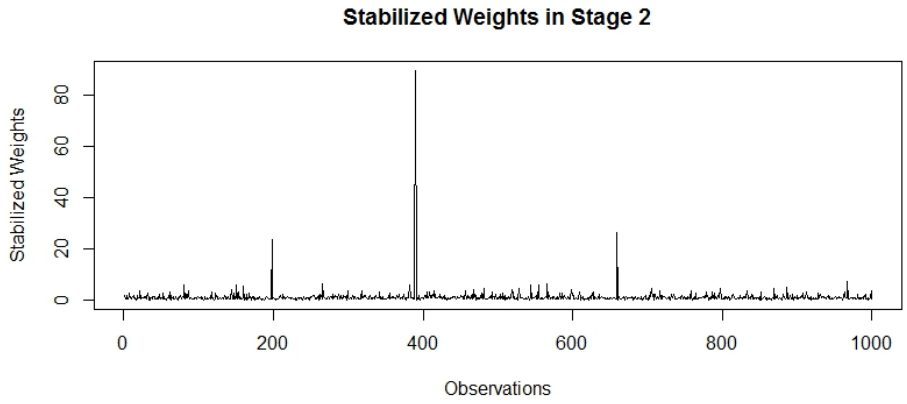



Figure 2.7: Stabilized weights for patients in stage 2 when treatments are independent

models were compared with the true treatment effects, it was observed that the parameter estimates of the unweighted model were close to the true values, as were the estimates of the weighted model. The weighted models had marginally smaller standard errors than the unweighted models. This results captures our intuition of the absence of the time varying confounding effect in stage 1.

In stage 2, the treatments in stage 1 and the covariates confound the effects of treatments on the final outcome. As explained earlier, this results in a time varying confounding effect that biases the effects of treatments in stage 2 on the outcome. The weights in Figure 2.7 are obtained by applying the IPTW framework in stage 2. As can be seen from the figure, some of the patients are assigned very high weights compared to others. This may be due to the probability of treatment assignment associated with a patient

| Treatments | True Value | Un-weighted Model | Weighted Models | Weighted Model with truncation |
|---|---|---|---|---|
| $T_1$ | -0.6 | -0.5529 (0.0483) | -0.5651 (0.04458) | -0.5186 (0.04571) |
| $T_2$ | -0.45 | -0.4280 (0.0464) | -0.4162 (0.04187) | -0.4303 (0.04328) |
| $T_3$ | -0.4 | -0.4951 (0.05025) | -0.3976 (0.03943) | -0.4169 (0.04808) |

Table 2.3: Parameter estimates (and standard errors) for independent treatments

is close to zero, resulting in a very high inverse value. This might cause inefficient treatment estimates with bias towards these high weights. To reduce the variance in this weight vector, a technique known as truncation is used (Bem (2008), Wan (2006), Lee (2011), Xia (2013)). Truncating weights based on a percentile, potentially reduces the effect of outliers on the weights. For example, a 95% truncation will assign the 95th and 5th percentile weights to patients that have weights above the 95th percentile and below the 5th percentile, respectively. This reduces the variation in the weights but subsequently induces error in the parameter estimates. Figure 2.8 below shows a plot of the 95% truncated weights.

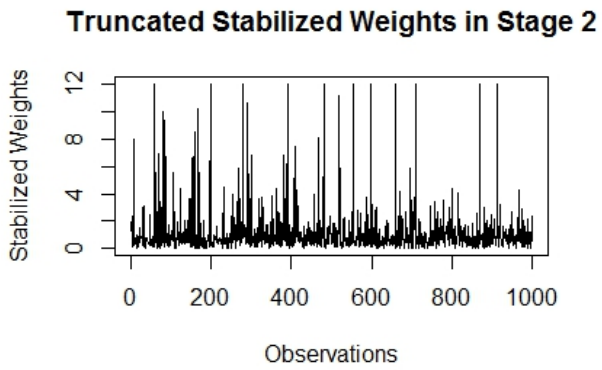



Figure 2.8: $95^{th}$ percentile truncated stabilized weights for patients in stage 2 when treatments are independent

Table 2.3 above summarizes the treatment parameter estimates for weighted (truncated and non-truncated) and unweighted models and compares them to the true treatment effects in stage 2. The values in the brackets are their standard errors. As can be seen from the table, the parameter estimates in the weighted model are approaching the true effects. This is more evident for treatment 3. It can also be seen that the standard errors are lower across all the treatments in the weighted model as compared to unweighted model. Although the variation in weights after truncation is lower, it can be seen that the estimated treatment effects

are further from the true effect. This is more evident for treatment 1. The standard errors are also higher in this truncated weights model compared to the one without truncation. This is because truncation, by reducing variance, fails to capture some rare events in the data, which corresponded to the association of high weight for the rare event. However, the table shows evidence that our generalized IPTW process addresses the time varying confounding effect of treatments on the outcome. A more comprehensive study on a real world problem to demonstrate the time varying effect in the case ofindependent treatments was conducted by LeBoulluec et al. (2018).

### 2.4.2 Case Study: Correlated Treatments

**Example 1**

The above section shows that the generalized IPTW process addresses the problem of time varying confounding in a multi-treatment, multi-stage study under the assumption that the treatments in the study are independent. However, in most medical studies the treatments exhibit some correlation. For example, there may be cases where treatment interactions should be considered. As explained in section 2.3, the joint probability cannot be written as the product of their marginal probabilities. The generalized IPTW process allows for correlated treatments, but requires an approach to estimate the joint probability. In this paper, we implement the factorization approach of LeBoulluec et al. (2013). However, there are some limitations in implementing the framework with factorization approach. This section examines these via different simulation case studies.

Consider a study involving two treatmentsThe data generation process for covariates is similar to the one described for the independent treatments case. However, all the treatments in the study are binary, and the probability assignment of a treatment is represented by equations below.

- Treatment 1 in stage 1:

$$P(t_{11} = 1) = \frac{exp(0.5x_1 + 0.3x_2 + 0.85x_3)}{1 + exp(0.5x_1 + 0.3x_2 + 0.85x_3)}.$$

| Treatments | True Value | Un-weighted Model | Weighted Models with true distribution | Weighted Model with average factorizations |
|---|---|---|---|---|
| $T_1$ | -1.1 | -1.214 (0.0915) | -1.251 (0.0877) | -1.272 (0.0763) |
| $T_2$ | -2.35 | -2.247 (0.0804) | -2.258 (0.0798) | -02.256 (0.0639) |

Table 2.4: Parameter estimates (and standard errors) for correlated treatments in example 1

- Treatment in stage 2:

$$P(t_{12} = 1) = \frac{exp(0.5x_1 + 0.3x_2 + 0.85x_3 + 0.85t_{11} + 0.5y_1)}{1 + exp(0.5x_1 + 0.3x_2 + 0.85x_3 + 0.85t_{11} + 0.5y_1)}.$$

- The outcome variable is generated using the following equations, stage 1:

$$y_1 = -0.1x_1 + 0.2x_2 - 0.1x_3 - 1.1t_{11} - 2.35t_{21} + \varepsilon(n, 0, 0.9).$$

- Outcome in stage 2:

$$y_2 = -0.1x_1 + 0.2x_2 - 0.1x_3 - 1.1t_{12} - 2.35t_{22} - 1.7y_1 + \varepsilon(n, 0, 0.9).$$

Treatments 1 and 2 are correlated with correlation coefficient of 0.8. As can be seen from the equations, the probability of assigning treatment 2 depends on the covariates and treatment 1. The causal diagram in Figure 2.5 shows this relationship. Moreover, the true joint probability distribution of the treatments can be deduced from the set of equations given above. $P(T_1, T_2|X) = P(T_1|X).P(T_2|T_1, X)$. In most cases, these joint probability distributions are not known, which is why the factorization approach is used. In this problem, since there are only two treatments in the study, it is computationally viable to obtain the probabilities by averaging across all the possible factorizations. The table below summarizes the results of estimating treatment effects in stage 2. We use different coefficients in this case to test the robustness of the generalized IPTW process.

It can be seen from the above table that the treatment estimates in the weighted models are close to their true values; moreover, the standard errors are lower than for the unweighted models. It can also be seen that the parameter estimates and standard errors are similar to those obtained under the model in which the joint distribution is estimated using the product of their marginal distribution under the true distribution.

**Example 2**

When calculating average factorizations for two treatments, the joint probabilities can be decomposed in the product of their marginal probabilities in $2! = 2$ ways. However, this count increases with the number of treatments. Consider now a larger study of 5 treatments, with medium to high correlations between them, available in both stages 1 and 2. Add seven covariates that are a mix of continuous and binary variables. These may represent a variety of patient variables, such as age, previous surgical history, duration of pain, pre-evaluation pain score, etc. The covariates are generated using a similar approach as in the previous examples, and all the covariates are assumed independent of each other. Also, the assumption of unmeasured confounders still holds. The outcome variables in the two stages are defined as follows:

$$
\begin{aligned}
y_1 = abs(&0.1x_1 + 0.4x_2 + 0.2x_3 + 0.1x_4 + 0.6x_5 + 0.4x_6 \\
&+ 0.2x_7 - 2t_{11} - 3.25t_{21} - 2.35t_{31} - 1.5t_{41} - 3.5t_{51} \\
&+ y_0 + \varepsilon(n, 0, 0.9)) \\
y_2 = abs(&0.1x_1 + 0.4x_2 + 0.2x_3 + 0.1x_4 + 0.6x_5 + 0.4x_6 \\
&+ 0.2x_7 - 2t_{12} - 3.25t_{22} - 2.35t_{32} - 1.5t_{42} - 3.5t_{52} \\
&+ y_1 + \varepsilon(n, 0, 0.8)).
\end{aligned}
$$

To estimate the joint probabilities of the 5 treatments, the joint probabilities can be decomposed in $5! = 120$ ways. As explained in section 2.3, the factorization approach can be implemented with random factorizations of the decomposed joint probabilities selected. The metric MPRD, also explained in section 2.3, is used to select an appropriate number of factorizations. To make the model more robust, under every factorization count, two groups are randomly created and averaged. Figure 2.9 shows the plot of MPRD for stage 2 treatments under different factorization counts.

It can be seen from the above plot that MPR) decreases with the number of factorizations (count). For more than 90 factorizations, MPRD is less than 10%. The number of factorizations can be selected using this plot. In this case 120 factorizations are selected for fitting the weighted model. The weights are generated using averaging over randomly selected factorizations and compared to using the true

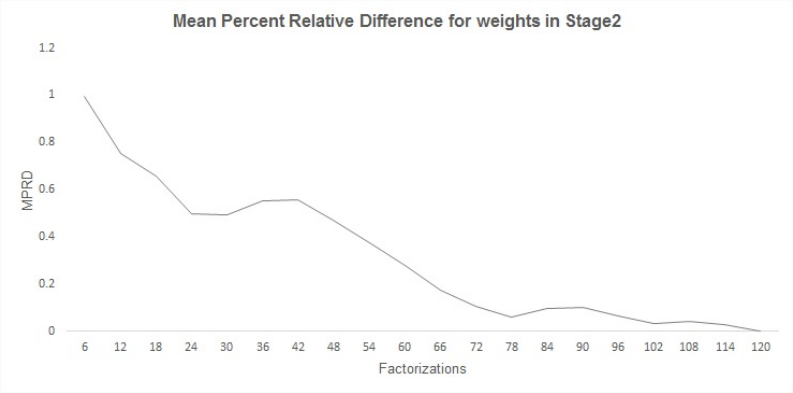



Figure 2.9: Mean percent relative difference for weights in stage 2

distribution. Truncated weights with 99% truncation are also generated for both sets of weights for comparison. The results are summarized in table below.

The table 2.5 shows the parameter estimates for the different models with the standard errors in the brackets. The parameter estimates of the weighted models under the true distribution closely follow the true value with the weighted models having lower standard errors than the unweighted ones. However, the weighted models using the average of factorizations have a larger deviation from the true values with high standard errors. The plot below 2.10 shows the deviation of the parameter estimates of the models from the true values.

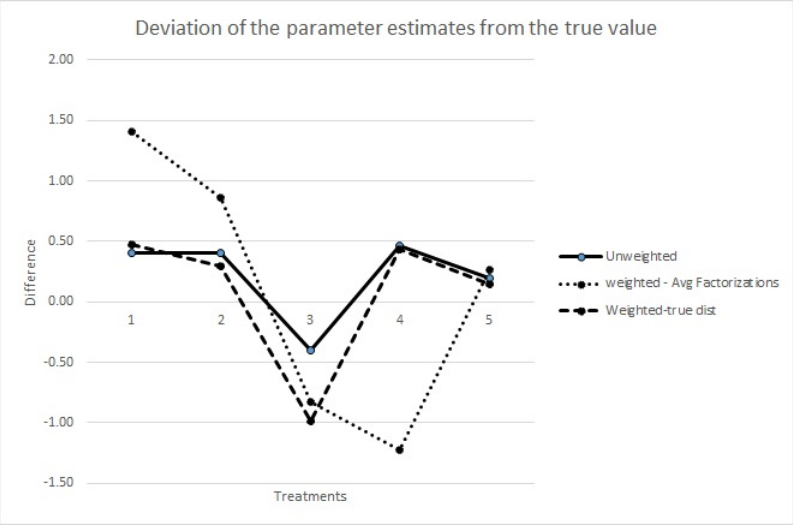



Figure 2.10: Deviation of the parameter estimates from the true value

| **Treatments** | **True Value** | **Unweighted** | **Weighted-average factorization** | **Weighted-average factorizations, 0.99 truncated** | **Weighted-true distribution** | **Weighted-true distribution, 0.99 truncated** |
|---|---|---|---|---|---|---|
| $T_1$ | -2 | -2.4085 (0.429) | -3.4126 (5.6618) | -3.4164 (5.6518) | -2.470 (0.4070) | -2.4021 (0.4391) |
| $T_2$ | -3.25 | -3.657 (0.2897) | -4.1094 (3.3609) | -4.1182 (3.3589) | -3.5424 (0.3655) | -3.5678 (0.3706) |
| $T_3$ | -2.35 | -1.9535 (0.4056) | -1.5260 (3.2962) | -1.5305 (3.2957) | -1.3665 (0.6022) | -1.3679 (0.6021) |
| $T_4$ | -1.5 | -1.9672 (0.4086) | -0.2702 (4.8527) | -0.2678 (4.8469) | -1.9383 (0.3922) | -1.9819 (0.4063) |
| $T_5$ | -3.5 | -3.6980 (0.4066) | -3.7680 (3.3098) | -3.7683 (3.3092) | -3.6447 (0.6000) | -3.6448 (0.6000) |

Table 2.5: Parameter estimates (and standard errors) for different models under correlated treatments in example 2

In theory, these chain rule for the calculating the joint probability should yield same value regardless of the ordering of treatments. However, since the source of variation in estimating these probabilities comes from the modeling approach and the data generation techniques, these were further investigated. The model is fit in R and the error induced in the result should be consistent, not only across all the different models under different factorizations, but also across different datasets. This then narrows the search to the data generation process. It was noticed that correlation coefficients for certain pairs of treatments were very high. For example, the correlation between treatments 3 and 5 was 0.96. This caused a problem of data separation which resulted in perfect separation of the simulated data. The resultant predicted probabilities for these treatments were either 1 or 0, which in turn caused high weights and subsequently biased treatment estimates. The problem of data separation is generally resolved by either eliminating one the correlated treatments or by merging the two treatments together to create a new variable. Since this exercise is beyond the scope of this paper, we move ahead with our analysis by creating a new example addressing this problem.

| Treatment 1 | Treatment 2 |
|---|---|
| 0 | 0 |
| 1 | 1 |

Table 2.6: High Postive Correlations

| Treatment 1 | Treatment 2 |
|---|---|
| 1 | 0 |
| 0 | 1 |

Table 2.7: High Negative Correlations

**Example 3**

The aim of this example is to demonstrate the generation of joint probabilities using the average factorizations approach. In the earlier examples, we have demonstrated the generalized IPTW process in estimating unbiased treatment effects, mitigating time varying confounding. However, it was noticed that these treatment effects are biased when a randomized factorization approach was used to estimate the joint probability. This was mainly due to the correlation structure of the treatments. When treatments are binary, a high correlation structure will generate data as described below.

- In the case of high positive correlations ($> 0.7$) between treatments, a majority of the data will have a structure as shown in the table below. That is, treatment combinations of (0,1) and (1,0) will have fewer instances, and, in the case of low sample size, they could even be missing. This causes data separation and inefficient parameter estimates.
- In the case of high negative correlations ($< -0.7$) between treatments, the data appear as in the table below. Fewwer instances of (0,0) or (1,1) reduces the predictive power of the model resulting in more misclassification and inefficient parameter estimates.

In the case of low to moderate correlations, logistic regression performs better than most techniques to obtain prediction probabilities, especially in cases with low sample size. Consider a study with three binary treatments and two covariates, one binary and one continuous. Let $n$ be the sample size, where $n = 200$. The data are generated using the following set of equations:

- Covariate $x_1$ follows normal distribution with mean 0 and standard deviation 1.

- Covariate $x_2$ follows binomial distribution with probability of success 0.5.

- Treatment 1 follows a binomial distribution with probability in stage 1 and stage 2 as,

$$P(T_{11} = 1) = \frac{exp(0.4x_1 + 0.3x_2)}{1 + exp(0.4x_1 + 0.3x_2)}$$
$$P(T_{12} = 1) = \frac{exp(0.4x_1 + 0.3x_2 + 0.35t_{11} + 0.15y_1)}{1 + exp(0.4x_1 + 0.3x_2 + 0.35t_{11} + 0.15y_1)}.$$

- Treatment 2 follows a binomial distribution with probability in stage 1 and stage 2 as,

$$P(T_{21} = 1) = \frac{exp(0.4x_1 + 0.3x_2 + t_{11})}{1 + exp(0.4x_1 + 0.3x_2 + t_{11})}$$
$$P(T22 = 1) = \frac{exp(0.4x_1 + 0.3x_2 + t_{12} + 0.35t_{21} + 0.15y_1)}{1 + exp(0.4x_1 + 0.3x_2 + t_{12} + 0.35t_{21} + 0.15y_1)}.$$

- Similarly, treatment 3 follows a binomial distribution with probability in stage 1 and stage 2 as,

$$P(T_{31} = 1) = \frac{exp(0.4x_1 + 0.3x_2 + t_{11} + t_{21})}{1 + exp(0.4x_1 + 0.3x_2 + t_{11} + t_{21})}$$
$$P(T_{32} = 1) = \frac{exp(0.4x_1 + 0.3x_2 + t_{12} + t_{22} + 0.15y_1 + 0.35t_{31})}{1 + exp(0.4x_1 + 0.3x_2 + t_{12} + t_{22} + 0.15y_1 + 0.35t_{31})}$$

- Thus, the true joint probability of distribution can be decomposed as the product of the conditional probabilities as $P(T_1, T_2, T_3|X) = P(T_1|X)P(T_2|T_1, X)P(T_3|T_1, T_2, X)$.

- The outcome variables in stage 1 and stage 2, with signal to noise ratio as 10, are defined by the following set of equations. The probability density plots of the outcome variable are shown in figure below.

$$y_1 = 3 + 0.4x_1 + 0.3x_2 + t_{11} + t_{21} + t_{31} + \varepsilon(n, 0, 0.9).$$

$$y2 = 3 + 0.4x_1 + 0.3x_2 + t_{12} + t_{22} + t_{32} - 0.5y_1 + \varepsilon(n, 0, 0.55).$$

The correlation structure of the treatments generated are summarized in the tables below.

The data generated using the above method is then used to estimate average joint probabilities. These estimated average joint probabilities are then compared with the true probabilities obtained by multiplying the product of the conditional probabilities generated using the equations above and shown in the plot below.

It can be seen from the plots that the estimated joint probabilities closely follow the true joint probabilities. A similar inference was observed from the probability plots in stage 1. These probabilities

| **Treatments** | $t_{11}$ | $t_{21}$ | $t_{31}$ |
|---|---|---|---|
| $t_{11}$ | 1 | 0.35 | 0.25 |
| $t_{21}$ | 0.35 | 1 | 0.28 |
| $t_{31}$ | 0.25 | 0.28 | 1 |

Stage 1

| **Treatments** | $t_{12}$ | $t_{22}$ | $t_{32}$ |
|---|---|---|---|
| $t_{12}$ | 1 | 0.38 | 0.18 |
| $t_{22}$ | 0.38 | 1 | 0.3 |
| $t_{32}$ | 0.18 | 0.3 | 1 |

Stage 2

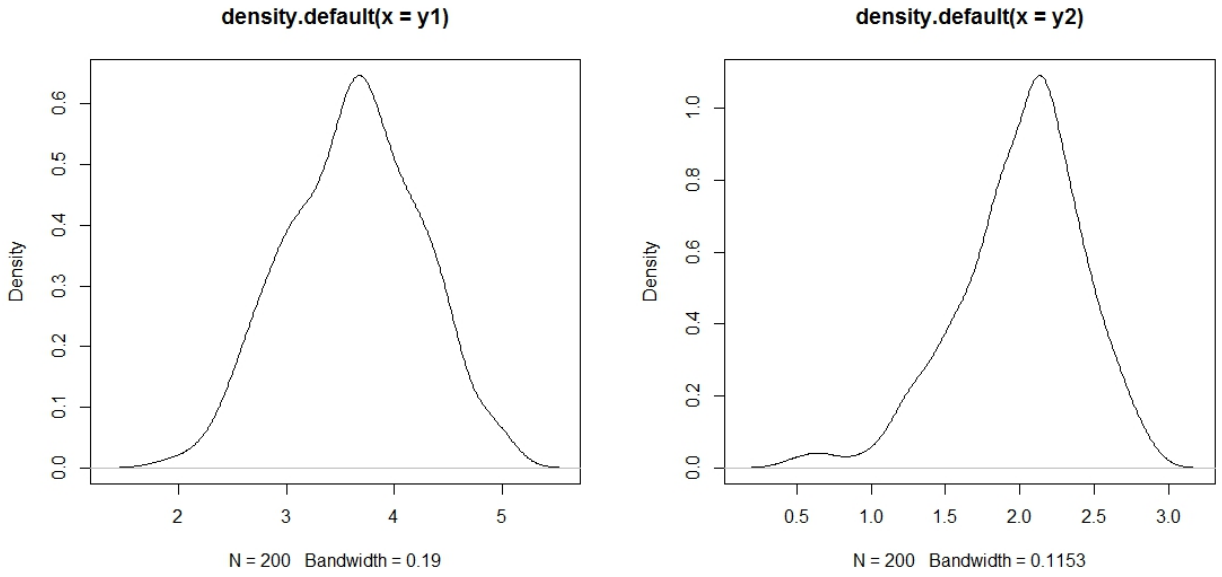
Table 2.9: Correlation Structure among Treatments

| | $t_{12}$ | | $t_{22}$ | | $t_{32}$ |
|---|---|---|---|---|---|
| $t_{11}$ | 0.37 | $t_{21}$ | 0.27 | $t_{31}$ | 0.19 |

Table 2.10: Treatment correlations between treatments in different stages

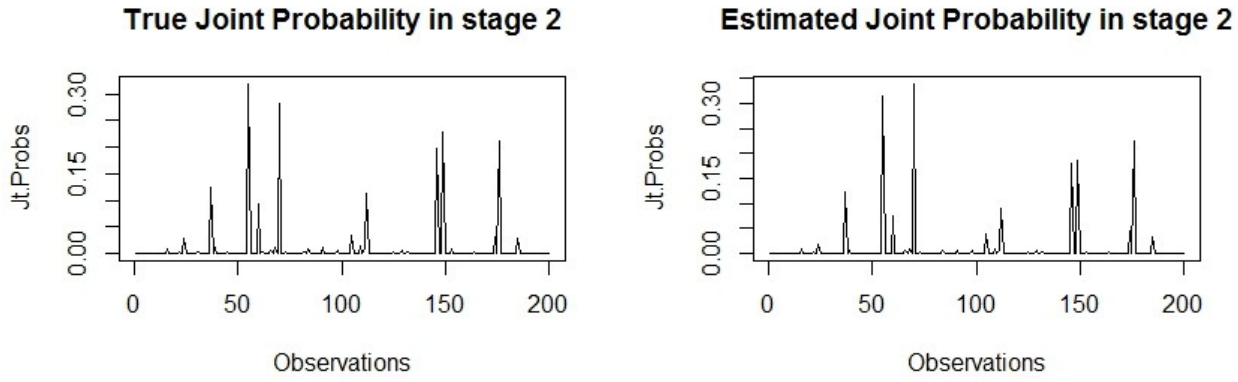
| **Treatments** | **True Effects** | **Un-weighted** | **Weighted** |
|---|---|---|---|
| $t_{12}$ | 1 | 1.07204 (0.05269) | 1.01052 (0.04686) |
| $t_{22}$ | 1 | 1.00642 (0.06456) | 1.02681 (0.05925) |
| $t_{32}$ | 1 | 1.0919 (0.07204) | 0.84141 (0.05641) |

Table 2.11: Parameter estimates (and standard errors) for treatments in example 3




Stage 1 Stage 2

Figure 2.12: Distribution of outcome variables




True Estimated

Figure 2.14: Joint Probabilities in Stage 2

were then used to fit the weighted models and the parameter estimates and standard errors were compared. The results are shown in the table below.

The weighted treatment estimates are closer to the true effect with lower standard errors than the unweighted models, as expected. In order to understand the effects of sample size on factorizations

approach, average factorizations were generated for the same example with varying sample sizes and compared with their true probabilities. The plot below shows the boxplot of the absolute deviation of the estimated probabilities under average factorizations from the true probabilities.

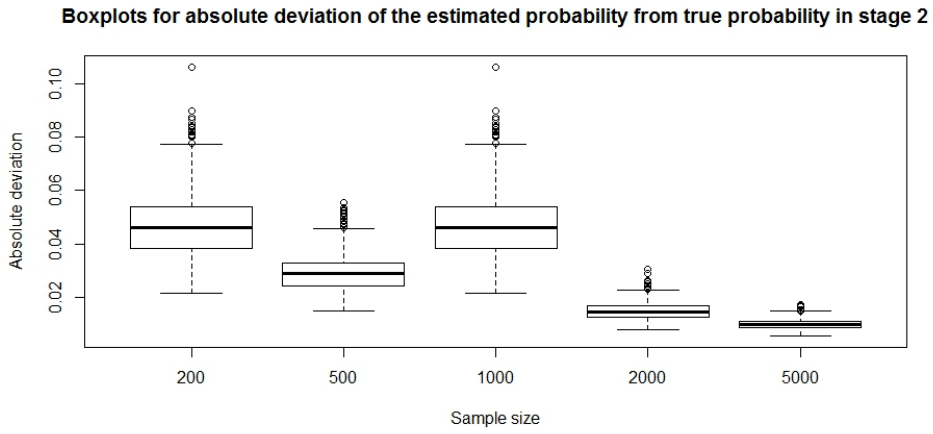



Figure 2.15: Boxplots for absolute deviation of estimated probability from the truth in stage 2

As the sample size increases, the interquartile range becomes tighter with fewer outliers. The effect on the factorization approach seems to reduce as the sample size increases. This can be tested by generating similar plots for a random decomposition of the joint probability. Figure 2.16 shows the boxplot of absolute deviation of estimated probability from the true probability under the following decomposition, $P(T_1, T_2, T_3|X) = P(T_2|X) \cdot P(T_3|T_2, X) \cdot P(T_1|T_3, T_2, X)$. The plot shows that with a larger sample size ($> 2000$) the absolute deviation is less than 0.02.

## 2.5 Concluding Remarks

The aim of this study is to demonstrate how to mitigate time varying confounding in observation studies through a framework of simulation case studies using a generalized IPTW process. This is the first paper to develop a generalized IPTW process. Further, our framework is used to study the factorization approach (LeBoulluec et al. 2013)) for calculating the critical joint probability in IPTW. In case of independent treatments, the generalized IPTW process simplifies to the approach in LeBoulluec et al.

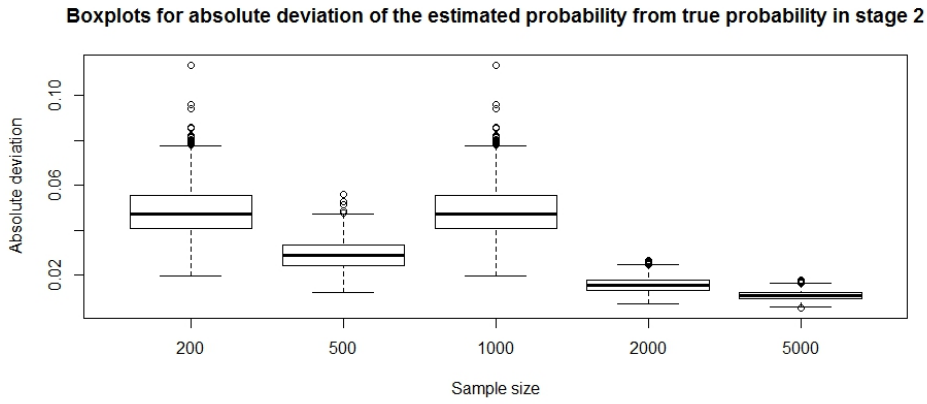



Figure 2.16: Boxplots for absolute deviation of estimated probability from the truth in stage 2 under random decomposition

(2018). This approach is straightforward since the joint probabilities of treatments can be decomposed as the product of their marginals conditioned on the covariates. This is demonstrated in the current work using an example with three binary treatments.

Correlated treatments are estimated by decomposing them using the chain rule of probability. The conditional joint probabilities can be decomposed using multiple orderings, and, thus, without a knowledge of the true ordering, these joint conditional probabilities can be estimated by averaging across factorizations using all the possible orderings. However, the number of possible factorizations increases with the number of treatments, so an approach is used in which a set of random factorizations are generated, and the probabilities generated using them are averaged to obtain the weights. The number of factorizations are selected using the proposed MPRD metric.

The estimation of joint probability under the scenario of correlated treatments is demonstrated using three examples. The first example shows the implementation of IPTW framework for two correlated treatments. The estimation of joint probabilities under a small set of treatments is computationally easy, and the time varying confounding seems to be addressed, as the treatment estimates for the weighted model are closer to the true values and have lower standard errors than the unweighted model. The second example demonstrates the high computational complexity for a five treatment scenario. The MPRD plot is used to select the number of factorizations necessary. The framework falls

short when the correlations between treatments are very high. A correlation coefficient of 0.7 or higher results in data separation and should be checked before implementing the generalized IPTW process. The third example shows the case when the correlations between the treatments are moderate and do not cause data separation. The generalized IPTW process successfully demonstrates that the average factorizations approach addresses the time varying confounding. The sample size effect is also tested, and it can be concluded that the error due to the sampling of factorizations reduces with the increase in sample size.

This simulation study framework successfully demonstrates that the generalized IPTW process can address time varying confounding. The IPTW formulation under different problem definitions is highlighted. However, there are some caveats such as the type of causal effect a researcher wants to study, the correlation structure of the covariates and the size of the problem, that a researcher should keep in mind. The IPTW formulation under different problem definitions is also highlighted.

# Acknowledgement

This material is based upon work partially supported by the National Science Foundation under Grant CMII-1434401.

# Bibliography

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

(2006). Diagnosing bias in inverse probability of treatment weighted estimator resulting from violation of experimental treatment assignment. *U.C. Berkeley Division of Biostatistics Working Paper Series*, 211:1–25.

(2008). Data-adaptive selection of the truncation level for inverse-probability-of-treatment-weighted estimators. *U.C. Berkeley Division of Biostatistics Working Paper Series*, 230:1–21.

(2011). Weight trimming and propensity score weighting. *PLoS ONE*, 3:1–6.

(2013). Comparison of approaches to weight truncation for marginal structural cox models. *Epidemiologic Methods*, 2:1–20.

Bodnar, L., Davidian, M., Siega-Riz, A., and Tsiatis, A. (2004). Marginal structural models for analyzing causal effects of time-dependent treatments: An application in perinatal epidemiology. *American Journal of Epidemiology*, 159(10):926–934.

Casella, G. and Berger, R., editors (2002). *Statistical Inference*. Wadsworth and Brooks, Pacific Grove, CA.

Cole, S. and Hernn, M. (2008). Constructing inverse probability weights for marginal structural models. *American Journal of Epidemiology*, 168(6):656–664.

Daniel, R., Cousens, S., De Stavola, B., Kenward, M., and Sterne, J. (2013). Methods for dealing with tine-dependent confounding. *Statistics in Medicine*, 32:1584–1618.

Fewell, Z., Hernn, M., Wolfe, F., Tilling, K., Choi, H., and Sterne, J. (2004). Controlling for time-dependent confounding using marginal structural models. *The Stata Journal*, 4(4):402–420.

Garcia-Aymerich, J., Lange, P., Serra, I., Schnohr, P., and Ant, J. (2008). Time-dependent confounding

in the study of the effects of regular physical activity in chronic obstructive pulmonary disease: An application of the marginal structural model. *Annals of Epidemiology*, 18(10):775–783.

Greenland, S. (1996). Confounding and exposure trends in case-crossover and case-time-control designs. *Epidemiology*, 7(3):231–239.

Hernan, M., Brumback, B., and Robins, J. (2001). Marginal structural models to estimate the joint causal effect of nonrandomized treatments. *Journal of the American Statistical Association*, 96(454):440–448.

Hernan, M., Hernandez-Diaz, S., and Robins, J. (2004). A structural approach to selection bias. *Epidemiology*, 15:615–625.

Hernan, M., Lanoy, E., Costagliola, D., and Robins, J. (2006). Comparison of dyamic treatment regimes via inverse probability weighting. *Basic and Clinical Pharmacology and Toxicology*, 98(3):237–242.

Imai, K. and Ratkovic, M. (2013). Covariate balancing propensity score. *Journal of Royal Statistical Society*, 76:243–263.

Iqbal, G. M. D. (2017). Multi-objective two-stage stochastic programming for adaptive interdisciplinary pain management with piece-wise linear network transition models. Ph.D. Dissertation, Department of Industrial and Manufacturing Systems Engineering, The University of Texas at Arlington, Arlington, Tx.

Joffe, M., Ten Have, T., Feldman, H., and Kimmel, S. (2004). Model selection, confounder control, and marginal structural models: Review and new applications. *American Statistician*, 58(4):272–279.

LeBoulluec, A., Ohol, N., Chen, V., Zeng, L., Rosenberger, J., and Gatchel, R. (2018). Handling time-varying confounding in state transition models for dynamic optimization of adaptive interdisciplinary pain mangement. *IISE Transactions on Healthcare Systems Engineering*, 0:1–10.

LeBoulluec, A., Zeng, L., Chen, V., Rosenberger, J., and Gatchel, R. (2013). Outcome and state

transition modeling for adaptive interdisciplinary pain management. Proceeding of 2013 Industrial and Systems Engineering Research Conference (ISERC).

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

Orellana, L., Rotnitzky, A., and Robins, J. (2010). Dynamic regime marginal structural mean models for estimation of optimal dynamic treatment regimes, part i: main content. *The International Journal of Biostatistics*, 6(2).

R Development Core Team (2008). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Rawat, R. and Manry, M. (2017). Second order training of a smoothed piecewise linear network. Neural Processing Letters.

Robins, J. (1986). A new approach to causal inference in mortality studies with a sustained exposure period - application to control of the healthy worker survivor effect. *Mathematical Modelling*, 7:319–336.

Robins, J. (1992). Estimation of time-dependent accelerated failure time model in the presence of confounding factors. *Biometrika*, 79(2):321–334.

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

Robins, J., Hernan, M., and Brumback, B. (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology*, 11(5):550–560.

Robins, J., Orellana, L., and Rotnitzky, A. (2008). Estimation and exploration of optimal treatment and testing strategies. *Statistics in Medicine*, 27:4678–4721.

Robins, J. M. (1993). *Analytic Methods for Estimating HIV-Treatment and Cofactor Effects*, pages 213–288. Springer US, Boston, MA.

Robins, J. M. (2004). *Optimal Structural Nested Models for Optimal Sequential Decisions*, pages 189–326. Springer New York, New York, NY.

Rubin, D. (1978). Bayesian inference for causal effects: the role of randomization. *The Annals of Statistics*, 6:34–58.

Rubin, D. (2009). Concerning the consistency assumption in causal interference. *Epidemiology*, 20(6):880–883.

Theommes, F. and Ong, A. (2016). A primer on inverse porobability of treatment weighting and marginal structural models. *Emerging Adulthood*, 4(1):40–59.

Vansteelandt, S. and Daniel, R. (2014). On regression adjustment for the propensity score. *Statistics in Medicine*, 33:4053–4072.

Wang, N. (2015). Two-stage stochastic programming for adaptive interdisciplinary pain management. Ph.D. Dissertation, Department of Industrial and Manufacturing Systems Engineering, The University of Texas at Arlington, Arlington, Tx.

# Chapter 3

# Handling time varying confounding in an adaptive interdisciplinary pain management program with multiple correlated treatment options


## Abstract

Time varying confounding is a common effect observed in medical research where the problem involves analyzing observational longitudinal data. In such setups, the effect of a treatment on an outcome is confounded by other factors, which causes a bias in treatment estimates. An important research topic is to understand the causal effects of treatments in such observational studies. Moreover, this problem becomes increasingly difficult in situations where multiple treatments are administered. One such case is an adaptive interdisciplinary pain management study that combines multiple disciplines of professionals to understand the biological and psychosocial factors causing a patientpain and to determine the best combination of treatments to administer. The sequential nature of the treatment induces time varying confounding. LeBoulluec et al. (2018), addresses this time varying confounding when treatments are independent; however, in most cases these treatments exhibit some correlation. This paper extends the IPTW framework of LeBoulluec et al. (2018) to address time varying confounding in a two stage adaptive interdisciplinary pain management program when treatments exhibit correlation. This paper simulates an example to study such a case to gain understanding, and then applies the approach to a pain management program at the Eugene McDermott Center for Pain Management at the University of Texas Southwestern Medical Center at Dallas. This approach can be

employed in other health care applications, for which treatments are observational over time stages.

## 3.1 Introduction

Today, pain has become a universal disorder affecting not only the individuals, but also the individual's family, friends and health providers, who provide support to help deal with the physical and emotional consequences of pain (MedicineNet 1996). Doctors often define chronic pain as pain that lasts for 3 to 6 months or longer (WebMD 2005). While acute pain is a normal sensation triggered in the nervous system to alert an individual to possible injury and the need to take care, chronic pain persists. Pain signals continue firing in the nervous system for weeks, months, and even years. Advanced neuroimaging has shown that chronic pain, unlike acute pain, can cause structural changes in the brain that add to the risk of cognitive problems, including anxiety and depression (SPINE-health 1999).

According to American Academy of Pain Medicine, more than 1.5 billion people in the world are affected by chronic pain. In the U.S. alone, about 100 million Americans are affected. Many chronic pain conditions affect older adults, among which two thirds suffer from back pain (Iqbal 2017). Approximately 3–4.5% of the global population suffers from neuropathic pain, with incidence rates increasing with age (of Pain Medicine 2017). To emphasize just how prevalent chronic pain is in the United States, Table 3.1 below shows the comparison of chronic pain with other major conditions. It can be seen that pain affects more Americans than diabetes, heart disease and cancer combined. To understand the economic aspect of pain as a public health problem, refer to Institute of Medicine Report, “Relieving Pain in America: a Blueprint of Transforming Prevention, Care, Education and Research, that documents costs of $560-$635 billion annually. This is equivalent to $2000 for everyone living in the U.S., which includes the total incremental cost of health care due to pain from $261–$300 billion, and from $297–$336 billion due to lost productivity (of Pain Medicine 2017). Chronic pain has adverse effects on the lifestyle of an individual. Apart from tremendous health care costs, rehabilitation, and loss of productivity, it bodes emotional and financial burden on the patient and their families. Hence, developing better pain treatments is the primary goal.

The goal of pain management is to improve function, enabling the individuals to participate in day-to-day activities. However, it is important to remember that chronic pain usually cannot be cured, but can

| Condition | Number of Sufferers in United States | Source |
|---|---|---|
| Chronic Pain | 100 million | Institute of Medicine of National Academics |
| Diabetes | 25.8 million | American Diabetes Association |
| Coronary Heart Disease | 16.3 million | American Heart Association |
| Stroke | 7 million | |
| Cancer | 11.9 million | American Cancer Society |

Table 3.1: Incidence of Pain as Compared to Major Conditions (of Pain Medicine 2017)

be managed. Thus, there has been an emphasis on pain management techniques to enable patients to cope with chronic pain problems. The most common treatments for pain include analgesic pain relievers (aspirin, acetaminophen, and ibuprofen), antidepressants, migraine headache medicines, and similar medications. However, studies have indicated that the integration of interdisciplinary and multidisciplinary pain management programs has promising effects on pain management. These programs promote non-invasive techniques, such as physical therapy and cognitive behavioral therapies. Multidisciplinary pain management involves a variety of specialists with independent goals. By contrast, interdisciplinary pain management involves treatment by a team of physicians, behavioral medicine specialists, physical therapists, nurses and care coordinators working towards one goal (Schatman and Campbell 2007). With a growing number of treatment options and new medications, formulating an evidence-based, individually-tailored treatment plan has become increasingly complex (LeBoulluec et al. 2018). Patients feel and express pain differently. Thus, there is a need to construct an individually tailored program for patients suffering from chronic pain. To further mitigate the pain, the treatment regime can be monitored in stages, which potentially leads to modifications in the treatment regime. The idea here is to dynamically assess the effects of treatment, so as to expedite improvement in patient status. The McDermott Center is one such adaptive interdisciplinary pain management program. The team works together to provide a variety of treatment plans and strategies to manage pain and improve quality of life (Center 2017). Interdisciplinary pain management involves treatment by a team of physicians, behavioral medicine specialists, physical therapists, nurses and care coordinators. The team works together to provide variety of treatment plans and strategies to manage pain and improve quality of life Center (2017). Figure 3.1 shows the layout of a two stage interdisciplinary pain management program at the McDermott Center. The larger goal of the collaboration is

to enable evidence-based treatment decisions using patient data analysis with multi-stage stochastic optimization.

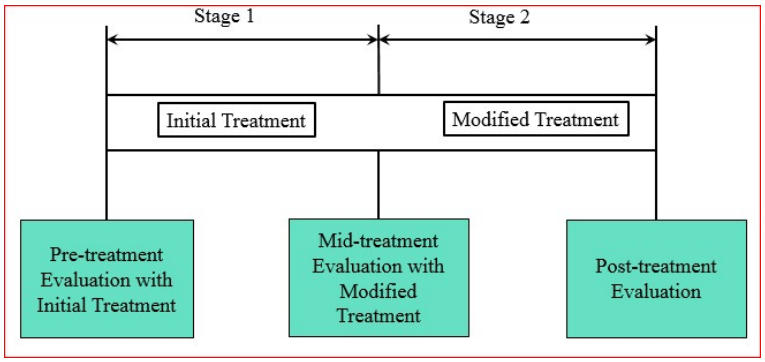



Figure 3.1: Two-Stage Interdisciplinary Pain Management Program (Iqbal 2017)

In the McDermott adaptive interdisciplinary pain management program, a patient complaining of chronic pain arrives at the center. The patients'medical history, including their demographic information, past diagnosis and surgical history, is recorded. Based on these data, a pre-treatment evaluation pain score is generated at the beginning of stage 1 which is used by a physician to prescribe a treatment plan for the patient. After a few weeks, depending on an individual case, the patient is evaluated on what is called a mid-treatment evaluation at the beginning of stage 2. At this point a new pain score is generated based on the patient/textquotesingles response to the treatment. Based on this score, a modified treatment regime is then prescribed. The post treatment evaluation happens at the end of stage 2 where a final pain outcome is measured concluding the two-stage program. Another evaluation occurs one year later, but this evaluation is not considered for this study.

It is known that pain is perceived differently by patients, so to help physicians evaluate each patient, specific pain outcomes are defined by the McDermott Center. These include: Beck Depression Inventory (BDI), Dallas Pain Questionnaire (DPQ), Oswestry Pain Disability Index (OSW), Pain Drawing Analogue (PDA), Multidimensional Pain Inventory (mpi), 36-item, Short Form Survey Physical Component Score (SF36pcs) and 36-item Short Form Survey Mental Component Score (SF36mcs). In this paper, only one pain outcome, Oswestry Pain Disability Index (OSW), is studied. Figure 3.2 shows the description of OSW outcome. OSW measures perceived functional disability and is most widely

used for assessing the disability level of back pain. This is essentially a questionnaire consisting of 10 sections with each section having a score range of 0–5. Thus, a maximum attainable score is 50. Based on the responses by a patient, a raw score is generated, which is interpreted as follows: 0–10 reflects minimal disability, and hence no treatment is necessary; 11–20 signifies mild disability, with a need for some conservative treatment; 21–30 signifies severe disability, and detailed investigation is required; 31–40 signifies crippling disability, with severe intervention required and 41–50 signifies that the patient is bed bound.

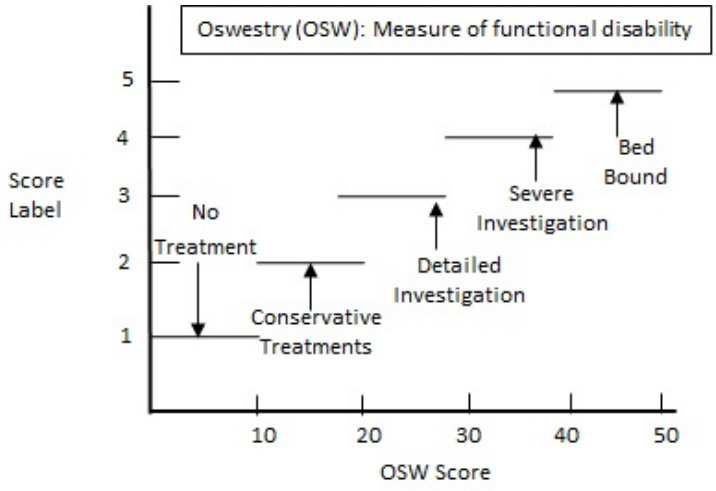



Figure 3.2: Oswestry Pain Disability Index (Iqbal 2017)

Given this complex treatment plan, it has become increasingly important for physicians to administer the most effective treatment plan for individual patients. An adaptive treatment decision model could assist physicians to identify a treatment regime tailored for each patient. Adaptive treatment strategies are a set of decision rules which can be adjusted depending on a patientstate (Murphy 2003). In such an adaptive treatment scenario, the patient's state variables are influenced by treatments, which in turn are influenced by the state variables and treatments in a previous stage. In other words, the treatment effect on a pain outcome is confounded by the patient's state variable. This phenomenon is called time varying confounding or endogeneity. In epidemiology studies, a key question is to understand the true or causal effect of treatment on an outcome. However, the time varying confounding results in biased estimate of treatment effects. It is thus necessary to adjust for these effects while de-

veloping causal models. Previous work on handling endogeneity considers a scenario where there is a single binary treatment and a binary outcome variable. Estimating treatment effects in such cases is performed by using adjustments to logistic regression methods (LeBoulluec 2013). However, the pain management problem contains a mix of binary and continuous data where traditional methods tend to fail. In this paper, a general procedure to handle such a complex problem is presented. Previous work on this pain management problem has been conducted by Lin et al. (2013),Wang (2015),Iqbal (2017), LeBoulluec et al. (2013). In particular, LeBoulluec et al. (2018) developed an Inverse Probability Treatment Weighting (IPTW) framework to address endogeneity when the treatment variables available to a physician are assumed to be independent. In related work, LeBoulluec et al. (2013) introduced a framework using a factorization approach when the treatments are correlated. The factorization approach is examined further by Ohol et al. (2018), and the current work extends the development of the IPTW framework for correlated treatments. The paper also encapsulates some common pitfalls and problems faced while implementing such a framework.

The remainder of this paper is organized as follows. Section 2 provides a literature review on pain management, adaptive treatment strategies, and time varying confounding. Section 3 discusses the bigger picture of the pain management problem and where this research fits in. Section 4 discusses the IPTW framework, encapsulating the approach, problems faced and techniques suggested to handle them. Section 5 presents a simulation case study to demonstrate the framework, and then describes the application of the framework to the McDermott adaptive interdisciplinary pain management case study. Finally, Section 6 discusses future work.

## 3.2 Literature Review

### 3.2.1 Pain Management Program

Pain managementgoal is to improve the quality of life for patients suffering from chronic pain. In the past, this was met with little success because pain management focused mainly on the physical side,

and patients were treated only with analgesic medications. With the growing addiction to opioid abuse, mainly in patients suffering from severe pain, it has become increasingly important to identify alternatives for treating chronic pain. Rosenblum et al. (2008) in his paper discusses the controversies, current status and future directions regarding the use of opioids and the treatment of chronic pain. Improvements were made as theories eventually evolved from single-cause to multi-cause explanations. As a result, adjuvant therapies that were designed for other medical conditions, in addition to the primary analgesic treatment were introduced to treat pain.

Melzack and Wall (Melzack and Wall 1965) were the first to propose a gate control theory that discussed the inclusion of physical and psychosocial factors. Conway and Higgins (Conway and Higgins 2011) in their literature review discuss the need to develop new treatment strategies. They provide recommendations for a model of care for pain management and also discuss important factors, such as legislations and other socio-economic factors to pain. Due to the complexities of pain, an interdisciplinary team of professionals, including anesthesiologists, physical therapists, psychologists, etc., are assembled to create pain management programs that build individualized treatment regimes for each patient. These programs offer broad forms of treatment and utilize multiple disciplinary components, depending on the type of pain and a patientresponse to the treatment. Application of interdisciplinary and multidisciplinary pain management has introduced novel approaches, such as cognitive behavioral treatment and other non-pharmacological treatments, for cases where medication does not alleviate the pain to a desired level. In addition, some medications have been discovered to provide better pain relief than analgesics (Schatman and Campbell 2007), (Gould 2007). There are a growing number of studies that indicate integration of interdisciplinary or multidisciplinary pain management programs has promising effectiveness on pain (Flor et al. 1992). For example, Flor et al. (1992) reviewed the result of sixty-five studies which supports the efficacy of multidisciplinary pain management centers. In another example, a study by Olason (2004) applied an interdisciplinary pain management program to focus more on increasing a patientability to function and eliminate analgesics in a rehabilitation clinic. Applying physiotherapies within a cognitive behavioral framework was shown to be successful by Eccleston and Eccleston (2004). Consequently, multidisciplinary and interdisci-

plinary pain management practices are being developed widely (Main and Spanswick (2000), Gatchel (2005), Schatman and Campbell (2007), Gould (2007), Gatchel et al. (2007)). This has led to the use of cognitive behavioral or non-pharmacological treatments, which are prescribed when analgesic medications cannot manage pain or provide a desired level of pain relief (Schatman and Campbell (2007), Gould (2007), Gatchel (2005), DArcy (2007), Gatchel and Okifuji (2006)).

A statistical modeling and optimization approach to adaptive interdisciplinary pain management was previously presented by Lin (2010), for which a stochastic dynamic programming approach was employed as a decision making tool. LeBoulluec (2013) first looked at the inherent endogeneity in data from the McDermott adaptive interdisciplinary pain management program. She developed a framework based on IPTW to mitigate the estimation of biased treatment estimates (LeBoulluec et al. 2018). This preliminary work showed promising results and shed light on the problem of endogeneity in multi-stage pain management data. Outcome models created by LeBoulluec et al. (2013) were incorporated in an two-stage stochastic programming model by Wang (2015). Iqbal (2017) developed a two-stage stochastic programming model in which the state transition and outcome models were piecewise linear network models created by Rawat and Manry (2017) using weights developed by LeBoulluec et al. (2018). Iqbal (2017) considers multiple pain outcomes in his model, which are essential in understanding the multi-dimensional aspect of a patientpain.

### 3.2.2 Adaptive Treatment Strategies

An adaptive treatment strategy is a dynamic framework which prescribes treatment to a patient depending on its varying needs (Murphy 2003), (Dawson and Lavori 2003), (Collins et al. 2007), (Murphy et al. 2007), (Pineau et al. 2007). It adapts the treatment type and level by sequentially modifying treatment based on the patient's changing state. To identify a treatment regime, an adaptive treatment strategy uses patient information, such as a patient's risk factors, outcome measures, irregularity (e.g., missing doctor appointments and failure to follow medicine regimen) in following the treatment plan, and outcomes as inputs to decision rules. This personalized approach is useful over the one-size-fits all treatment strategy. Based on this notion, adaptive treatment strategies are gaining momentum in

treating some re-occurring disorders, including depression, chronic pain, drug addiction, etc.(Dawson and Lavori (2003), Lavori and Dawson (1998), Lavori et al. (2001), Collins et al. (2005), Murphy (2005)).

Developing useful adaptive treatment strategies requires data that can be collected either from a randomized experiment or from an observational study. It is usually preferred to have data generated from randomized experiments, such as clinical trials, where useful inferences can be made in a controlled environment. Traditionally, random confirmatory trials, also known as RCT, is the most accepted method to evaluate a treatment program. However, one of the major drawbacks is that RCT only analyzes the overall treatment effect (Howard and Jacobs 2016). With the introduction of adaptive treatment strategies, experimental designs must be modified to consider a patients response to treatment, including multiphase optimization strategies (MOST) (Collins et al. 2007) and sequential multiple assignment randomized trail (SMART). Murphy Murphy (2003) published the first attempt at developing adaptive treatment strategies using a reinforcement learning based dynamic programming approach (Si et al. (2004), Werbos (1974), Kaelbling et al. (1996), Sutton and Barto (1998), Lee and Lee (2004), Werbos (2007)). They focused on sequential randomized clinical trials, which yielded ideal data for optimizing adaptive treatment strategies (Murphy et al. (2007), Collins et al. (2007), Pineau et al. (2007), Guez et al. (2008), Murphy and Bingham (2009), Shortreed et al. (2011)). In a more recent study, Nahum-Shani et al. (2016) used Q-learning regression, a data analysis method with SMART, to construct an adaptive treatment strategy employing naltrexone, behavioral intervention and telephone disease management to reduce alcohol consumption over 24 hours in alcohol dependent individuals. Gunlicks-Stoessel (2016) constructed a SMART design to develop an adaptive treatment strategy for adolescent depression. Shavitt (2016) also used SMART for children and adolescents with obsessive-compulsive disorder. Kugler Kugler et al. (2016) used MOST to build a remotely delivered responsive parenting intervention to prevent obesity among children of low income mothers, with and without depressive symptoms. Other notable work implementing MOST was done by Gwadz et al. (2017) and Collins et al. (2016). Medical professionals utilize many methods today for various purposes, such as clinical experience, trial and error, behavioral, and psychosocial and

biological theories. These methods can be utilized to create decision rules for adaptive treatment strategies (Lin et al. (2013), LeBoulluec et al. (2013), Robins (1986), Robins and Berkane (1997), Robins and Greenland (1994), Murphy (2003), Si et al. (2004), Murphy et al. (2007), Collins et al. (2007), Pineau et al. (2007)). Adaptive treatment strategies were first applied to pain management by Lin (Lin (2010), Lin et al. (2013)) who developed a framework for adaptive pain management based on the adaptive treatment strategy concept.

### 3.2.3 Time Varying Confounding

In epidemiology studies, it is desired to estimate the causal effect of a treatment. In a non-experimental longitudinal study, conventional methods of parameter estimation fail to address the inherent confounding caused by covariates on treatments that are themselves correlated over time. This means the probability of assigning a treatment at time, $t$ , depends on its past history. In an experimental setup, where the treatments are randomly assigned, this time varying confounding does not occur. Conventional methods, such as propensity scoring, could be used to obtain unbiased estimates of the average effect of a treatment on the outcome, but fails to estimate unbiased causal effects (Daniel et al. 2013). Robins (1999) explains the dilemma of causal interpretation by introducing two terms, causal and statistical exogeneity. According to Robins, a treatment is statistically exogenous if the probability of assigning that treatment given its assignment history does not depend on the history of time varying prognostic factors, like patient variables. If a treatment is statistically exogenous, it does not necessarily means that it is causally exogenous. Causal exogeneity is different from statistical exogeneity because of the presence of unmeasured confounders that influence in predicting the probability of treatment assignment given its past treatment history. The opposite of exogeniety is endogeneity. It must be noted that the econometric time series literature has a different definition of endogeneity. However, the IPTW approach the In the next two subsections, an explanation of endogeneity is presented, and the IPTW approach for addressing endogeneity is described.

**Endogeneity**

Endogeneity, which can be considered as opposite of exogeneity, is defined when an independent variable is correlated with its error term in a regression model $y = f(X) + \epsilon$. Endogenous variables share a form of functional relationship that influences the outcome variable. The functional relationship is the correlation between the predictor and the error term. For example, a time varying treatment $\mathbf{U_t}$ is called endogenous if its probability depends on the history of time-dependent patient variables $\mathbf{X_t} = \{\mathbf{X_1}, ..., \mathbf{X_t}\}$, conditional on the treatment history in stage $t-1$. $\mathbf{U_{t-1}} = \{\mathbf{U_1}, ..., \mathbf{U_{t-1}}\}$. This can be expressed mathematically as,

$$Corr(\mathbf{X_2}\{\mathbf{U_2}|\mathbf{U_1}\}) \neq 0, \tag{3.1}$$

where $Corr(A, B|C)$ denotes the correlation of A and B given C. Equation 3.1 states that treatment on each day depends on the history of both treatment and patient variables. In the presence of endogeneity, the estimation of the treatment effect will be biased. More specifically, the main concern in epidemiology studies is the causal effect of the treatment on an outcome of interest. Here a causal effect means a direct effect from the treatment to the outcome, not from any other variable, or through any other variable. Correspondingly, the bias caused by endogeneity is with respect to the true causal effect. In other words, with endogeneity, an unbiased estimate of the causal effect of treatment on the outcome cannot be obtained (Robins (1999), Robins et al. (2000)). Exogeneity ensures independence between the patient variables and treatment in a particular stage, given the treatment history in the previous stages, i.e., the correlation in above equation is equal to zero in the presence of exogeneity. Robins, also defines the degree of statistical non-exogeneity, or endogeneity using equation 3.2, which is eventually developed as an equation to estimate IPTW weights.

$$w^{-1}(t) = \prod_{t=0}^{T} \frac{P(U(t)|\bar{U}(t-1), X(t))}{P(U(t)|\bar{U}(t-1))}, \tag{3.2}$$

where the numerator is the probability of treatment received given the treatment and prognostic factor history, and the denominator represents the probability of treatment received in a stage given its treatment history only. If $w^{-1}(t) = 1$, then it implies statistical exogeneity, i.e., no endogeneity.

**Inverse Probability of Treatment Weighting (IPTW) Method**

In an adaptive treatment strategy, endogeneity is a challenging problem because conventional methods for confounder adjustment, such as stratification, matching and propensity score methods, regression adjustment, and standardization do not work (Weitzen et al. (2004), DAgostino (2007), Klungel et al. (2004)). Instrumental variable methods presented by Hogan and Lancaster (2004) obtain unbiased estimation of treatment effects by making use of some instruments or additional information. However, the reliance on the availability of instruments limits their applicability. A plethora of literature in the field of econometrics handles endogeneity using the generalized method of moments (Windmeijer and Santos Silva 1997) in count data models with endogenous regressors. Moodie et al. (2009) in her work proposes structural mean models using Gg-estimation to obtain unbiased parameter estimates for a set of decision rules. She analyzes the duration of breastfeeding with a view to maximizing infant growth. However, the decision variable in her work is binary, and the intervention is randomized, unlike the non-randomized nature of the McDermott adaptive interdisciplinary pain management case study. The IPTW method has been gaining popularity in this type of research and provides a viable solution to the endogeneity problem (Robins et al. (2000), Hernan et al. (2001), Joffe et al. (2004), Bodnar et al. (2004), Fewell et al. (2004), Cole and Hernn (2008), Garcia-Aymerich et al. (2008), VanderWeele (2009)).

More recent work on IPTW has been done by Hong and Raudenbush (2008), Austin and Schuster (2016), Austin and Stuart (2015), Austin (2013), Theommes and Ong (2016), Huber (2014). All the referenced work assumes the model form to be known, where a baseline model is built using the treatment and all covariates in the study. A weighted model is then fit using weighted linear model to compare it with the baseline model to check for reduction in confounding effects. For example, Austin and Stuart (2015) provide qualitative and quantitative methods to assess whether measured baseline covariates are balanced between treatment groups in weighted samples. They also propose a set of diagnostics when using IPTW to estimate causal treatment effects using observational data. Hong and Raudenbush (Hong and Raudenbush 2008) propose a strategy for studying the effects of time varying instructional treatments in repeatedly observed student achievement. The students are

monitors in two stages, grades 4 and 5, and the treatment variable is a binary variable that is the assignment of intensive math instruction. Theommes and Ong (2016), in a primer on IPTW and marginal structural models, also introduces the application of treatment weights in multiple stages for a single binary treatment. They study on the effects of developmental tasks on emerging adulthood. While Raudenbush considers weighting approach in the final stage, Theommes discusses the potential for weighting in multiple stages in a study.

### 3.2.4 Contribution

The major contributions made in this research are as follows:

- This research generalizes the IPTW framework by **?**, under the condition of correlated treatments, by addressing issues such as data separation and treatment interactions and co-usages.
- In problems where the true model form is unknown and the treatments are correlated, estimating joint conditional probability distribution is critical. This paper presents some greedy techniques and compares them with an average factorization approach examined in an earlier work (Ohol et al. 2018).

## 3.3 Overview of Adaptive Interdisciplinary Pain Management

This section presents an overview of adaptive interdisciplinary pain management research and the issues addressed. It is important to highlight the overall research perspective of the problem to establish different branches of the research. Figure 3.3 shows an integration on adaptive interdisciplinary pain management research objectives within the pain management interface.

The nodes under the interface forms the scope of the research for this problem. The medical records and patient information are collected by researchers at the Eugene McDermott Center for Pain Management at the University of Texas Southwestern Medical Center at Dallas. Thus, the raw data from the McDermott Center directly interact with the pain interface and form the primary input. The raw data

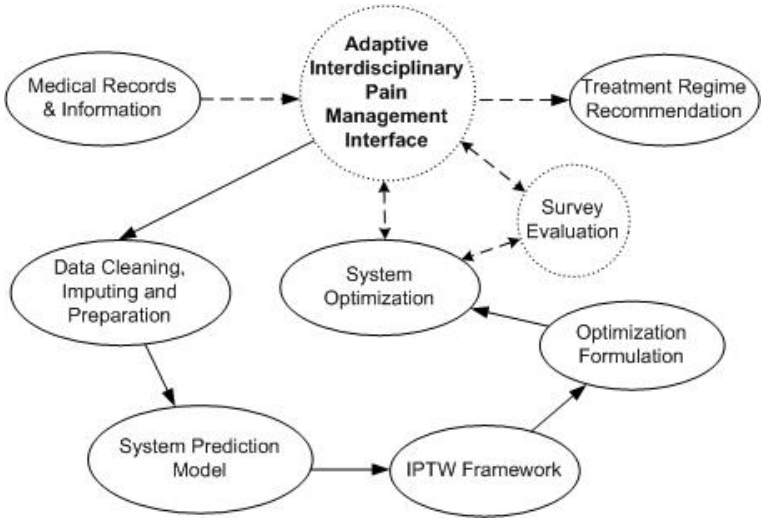



Figure 3.3: Integration of Adaptive Interdisciplinary Pain Management Research Objectives

includes patient information, treatment plans administered across two stages, and their pain outcome measures. Raw data contains many missing or invalid values that can be addressing via imputation of missing values, so as to preserve as much data as possible. Besides imputation, many patient variables and treatments are grouped together to reduce the sparsity of data. Treatment grouping is done in consultation with a physician at the center. This process forms the first step of the interface and is labeled as Data Cleaning, Imputing and Processing. The next step involves building state transition and outcome models for patients using linear regression or piecewise linear network methods. These are prediction models which are used in an optimization framework for treating patients with an objective to minimize cost and reduce pain. Since the McDermott adaptive interdisciplinary pain management program is two stages and adaptive in nature, there is inherent time varying confounding causing biased treatment estimates. This needs to be addressed before fitting the state transition and outcome models. Thus, the next step in the research is to develop an IPT process to mitigate endogeneity. The weights obtained from this step are used to create weighted predictive models that are formulated as constraints in the optimization model. The formulation and development of the optimization model is the final objective of this larger project. Different versions of the optimization have been studied, including two-stage stochastic programming, stochastic dynamic programming using mixed integer linear programming, mixed integer quadratically-constrained programming. As a part of

future work, survey evaluations would be conducted and incorporated in the pain management interface. The output of the interface is the recommended treatment regime. This paper focuses on the third step of the interface, which is the development of the IPTW framework to mitigate endogeneity. More specifically, this paper discusses solutions when the treatments in a given stage are correlated with each other.

## 3.4 Inverse Probability of Treatment Weighting (IPTW) for correlated treatments

As mentioned earlier, LeBoulluec et al. (2018) demonstrates the implementation of an IPTW framework for multiple independent treatment variables. The generalized IPTW framework used for correlated treatments follows a similar structure, but must address the key challenge of calculating the joint probability of treatments. This generalized framework also encapsulates some other common problems that could be encountered while working with observational longitudinal data. Ohol et al. (2018) explains in detail the parameter estimation using marginal structural models when treatments are correlated. Causal diagrams are used to explain the confounding effects of patient variables on treatments. The estimation of the joint probability is derived via the chain rule, using an average factorization approach. This section reviews the average factorization approach, while discussing other greedy algorithms to estimate the joint probability of treatments.

Figure 3.4 below shows a causal diagram for the two-stage pain management program. The nodes represent various parameters considered in the study. Let $X_i$ denote the vector of patient variables such as age, gender, marital status, past medical history, etc. Let $U_i$ represent the decision variables which is a vector of treatment options available for the physician to administer. Let $Y$ denote the pain outcome measure, in this case, the Oswestry Pain Disability measure (OSW). The arrows represent an effect of one variable on another. It can be seen from the figure that the structure of this causal diagram is appropriate for the McDermott pain management program. The state or patient variables in the first stage affect the decision or treatment variables of the current stage and then influence the

state and decision variables in the next stage, which in turn affect the outcome measure. It is also noted that the state variables in the first and the second stages confound the effect of the decision variables on the pain outcome.

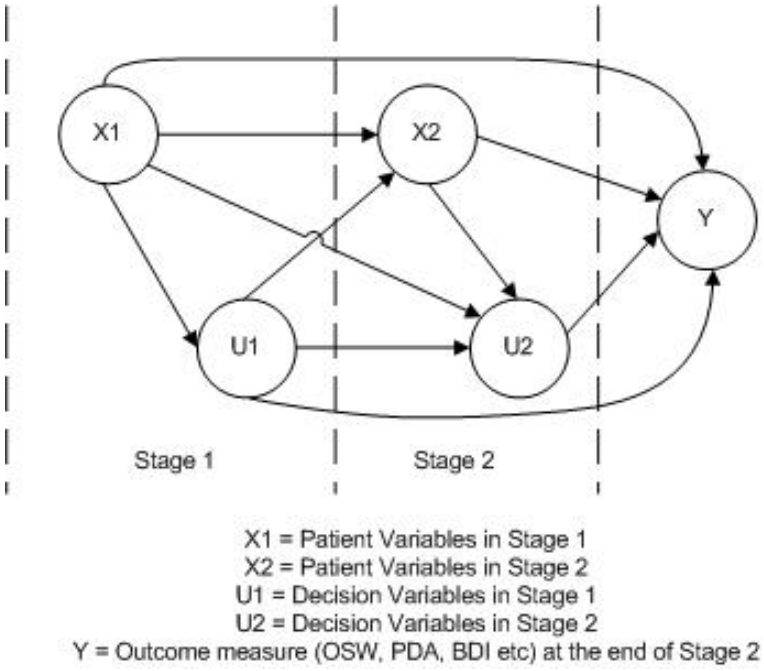



Figure 3.4: Causal diagram for two stage pain management program

Under the assumption of no unmeasured confounders, unbiased parameter estimates can be obtained by performing a weighted analysis, where each patient is assigned a weight depending on the inverse of the conditional probability of receiving a treatment. Thus, for a single treatment in a single stage, the weights can be represented as, $w_i = 1(P(U_1 = u_1|\mathbf{X_1} = \mathbf{x_1}))$, which is the inverse of the probability that a patient receives treatment $U_1$ at level $u_1$, given his or her patient history $X_1$. The treatment can be either binary or categorical. For continuous treatments, a probability density function could be considered instead of the conditional probability. In this research, however, we handle treatments which are binary. The weights are estimated by fitting a preliminary logistic regression, for example $U_1$ on $X_1$ in stage 1. The estimators obtained using these weights are called the IPTW estimators Robins et al. (2000). The effect of weighting is to create a pseudo-population consisting of $w_i$ copies of each patient reducing the effect of the confounding patient variables. Referring back to Figure 3.4, the

McDermott adaptive interdisciplinary pain management problem comprises two stages with multiple treatment options available in each stage. The IPTW process can incorporate multiple treatments by decomposing the joint probability of the treatments being observed into the product of their conditional marginal probabilities in every stage.

$$w = \prod_{t=1}^{T} \prod_{k=1}^{K} \frac{1}{P(U_{kt} = u_{kt} | \mathbf{U_{kt-1}} = \mathbf{u_{kt-1}}, \mathbf{X_t} = \mathbf{x_t})} \tag{3.3}$$

Equation 3.3 calculates the IPTW weights for patients registered in the pain management program where

$t = 1, 2$; number of stages

$k =$ number of treatments

$u_{kt}$ is the observed $k^{th}$ treatment in current stage $t$

$x_t$ is the vector of patient or state variables in current stage $t$

$u_{kt-1}$ is the value of $k^{th}$ treatment observed in the previous stages Thus, for example, for two stages and two treatment options available in both stages, the weights in stage 1 can be calculated as,

$$w_{stage1} = \prod_{t=1}^{1} \prod_{k=1}^{2} \frac{1}{P(U_{kt} = u_{kt} | \mathbf{U_{kt-1}} = \mathbf{u_{kt-1}}, \mathbf{X_t} = \mathbf{x_t})}$$

$$w_{stage1} = \frac{1}{P(U_{11} = u_{11} | \mathbf{X_1} = \mathbf{x_1})} \frac{1}{P(U_{21} = u_{21} | \mathbf{X_1} = \mathbf{x_1})}$$

and the weights in stage 2 as,

$$w_{stage1} = \prod_{t=1}^{2} \prod_{k=1}^{2} \frac{1}{P(U_{kt} = u_{kt} | \mathbf{U_{kt-1}} = \mathbf{u_{kt-1}}, \mathbf{X_t} = \mathbf{x_t})}$$

$$w_{stage1} = \frac{1}{P(U_{11} = u_{11} | \mathbf{X_1} = \mathbf{x_1})} \frac{1}{P(U_{21} = u_{21} | \mathbf{X_1} = \mathbf{x_1})} \frac{1}{P(U_{12} = u_{12} | \mathbf{U_{11}} = \mathbf{u_{11}}, \mathbf{X_2} = \mathbf{x_2})} \frac{1}{P(U_{22} = u_{22} | \mathbf{U_{21}} = \mathbf{u_{21}}, \mathbf{X_2} = \mathbf{x_2})}$$

These formulations can be extended to multiple treatments and multiple stages without the loss of generality. The predicted probabilities in the denominator are obtained using appropriate regression models. In the pain management study, treatments are assumed to be binary where logistic regression models are fitted to obtain the predicted conditional probabilities. It should also be noted that the

probability, $P(U_{kt} = u_{kt}|U_{kt-1} = u_{kt-1}, X_t = x_t)$, may vary greatly between patients when components of the state variables are strongly associated with treatment variables. This may result in some patients being weighted extremely high compared to others. As a result the IPTW estimator may have large variance and may deviate from normality. In order to avoid this, the weights are replaced with stabilized weights, which are calculated as, $s_w = P(U_k = u_k)/P(U_k = u_k|\mathbf{X_1} = \mathbf{x_1})$ for a particular treatment k in single stage. If the treatment is un-confounded with the state variable, this stabilized weight $s_w$ is equal to 1 indicating no association between treatment and state variable. In the case of confounding, the weight would be near 1 and less variable than the un-stabilized weight $w$ Robins et al. (2000). Thus, for the McDermott adaptive interdisciplinary pain management case study, the stabilized weights are calculated using equation 3.4.

$$w_{stage1} = \prod_{t=1}^{T}\prod_{k=1}^{K} \frac{P(U_{kt} = u_{kt}|\mathbf{U_{kt-1}} = \mathbf{u_{kt-1}})}{P(U_{kt} = u_{kt}|\mathbf{U_{kt-1}} = \mathbf{u_{kt-1}}, \mathbf{X_t} = \mathbf{x_t})}, \tag{3.4}$$

where the probability in the numerator can be calculated as the predicted probability in a univariate logistic regression model in stage 1, and as a logistic regression model conditioned on the treatment observed in previous stages for stages greater than 1. The joint conditional probability of treatments is estimated using the average factorization approach, as explained by LeBoulluec (2013) and Ohol et al. (2018). However, the average factorization approach becomes computationally expensive with the increase in the number of treatments. Moreover, this approach tends to induce more noise since the effect of true distribution is averaged along with multiple other decompositions that do not represent the true distribution. This section describes other algorithms that can be considered as an alternative to the average factorization approach. This paper lays a foundation for the next steps in the outcome and state transition modeling approach. The rest of this section explains the steps in the generalized IPTW framework, and Section 4 discusses a simulated case study and the McDermott case study.

### 3.4.1 Generalized IPTW Framework

**Step 1: Data Investigation**

This step is different from the initial phase of the McDermott adaptive interdisciplinary pain management interface which deals with data imputing, cleaning and preparation. Once the dataset is pre-processed,further investigation should be done based on expert knowledge on the data. Any anomalies should be consulted and taken out of the analysis. For example, in the McDermott adaptive interdisciplinary pain management case study, there are certain dosage limitation on certain medications which might be prescribed. There are severe interactions between medications which have been specified at the pain center. Any observations which violates these conditions should be first consulted with an expert and if deemed incorrrect, should be removed from the analysis.

**Step 2: Handling Data Issues**

This step involves investigation of variables before obtaining weights. Initial exploratory data analysis can help obtain insights on some of the common data issues embedded in the data. These require attention at a fairly early stage of model building, in order to avoid issues in model fitting. This step is similar to the feature engineering step in machine learning problems, where an engineer has to check for data issues, such as sparsity, correlations, and the presence of similar features. However, in observational study with a small the sample size, issues of data separation may arise when a predictor variable or a set of predictor variables completely or partially separate an outcome variable. This happens in particular when the data comprise of binary or categorical variables, like in the pain management case study. Under data separation, the modeling fitting process will fail for a logistic regression model. This is because the maximum likelihood estimates do not exist for the variables that separates the outcome variable. Complete and quasi-complete separation can be identified by forming $22$ contingency tables to observe the cell frequencies of the outcome variables against each predictor variable. If either of the non-diagonal elements have zero cell frequencies, there is quasi-complete separation and if all the non-diagonal cells have zero frequencies, there is complete separation. Fur-

ther investigation is done by observing the estimated coefficients and standard errors obtained from fitting logistic regression models with treatments as response variables and the patient state variables as predictors. For the software-generated fitted model, variables that contribute to data separation have high estimated coefficients and high standard errors. Predicted probabilities obtained under complete or quasi-complete separation are asymptotically close to 1. This means there is perfect or near perfect prediction and the weights obtained have very high magnitude. Thus, for the McDermott adaptive interdisciplinary pain management case study, the data should be checked for separation and if it exists, the contributing variables must be handled.

Data separation can be handled by either grouping levels of categorical variables or by checking if another version of outcome or predictor variable is present. In the proposed framework, grouping conducted by considering various constraints and rules that exist for the specific case study or application. Treatments are grouped separately from patient variables. In medical applications, treatments can be grouped based on their co-usage and interactions. Expert knowledge along with correlation analysis and variance inflation (VIF) values can be used to form groups.

#### Step 3: Building Preliminary Models

This step involves building initial state transition and outcome models to identify significant variables in the study. The dataset in observational study can consist of large number of independent variables and fewer observations. In previous work by LeBoulluec et al. (2018) the number of insignificant variables are eliminated by fitting a stepwise linear regression model or piecewise linear network on the data. Although this step helps reduce the computational complexity of a problem with large feature space, care must be taken before eliminating them since some of the features might have confounding effects on the treatments. Eliminating such variables would violate the assumption of unmeasured confounders resulting in inefficient treatment estimates obtained from the IPTW framework. The McDermott adaptive interdisciplinary pain management data consist of patient variables, referred to as state variables in the optimization, and treatment variables referred to as decision variables in the

optimization. The preliminary model is formulated as,

$$y = f(X, U) + \varepsilon,$$

where $y$ is the patient's outcome measure, $X$ is the vector of patient variables, $U$ is the vector of treatments, and $f$ is a linear function. For example, to estimate effects of three treatments, such as psychotherapy, physical therapy and NSAIDS, on an Oswestry Pain Disability Measure (OSW), a linear model could be used. The linear model is built with other covariates such as age, surgical history and previous recorded pain measure,If there is enough evidence that a state variable, such as surgical history, confounds the effect of the treatments prescribed on the pain outcome, these variables cannot be removed from the analysis.

**Step 4: Estimating Joint Probabilities**

The IPTW framework uses Equation 3.3 to estimate the weights. The denominator of the equation requires estimating the conditional probability of a treatment administered to a patient given the patient's treatment and covariate history. In the case of multiple treatments, this estimation comes from decomposing the joint conditional probabilities of treatments given a patient's history. LeBoulluec et al. (2018) discusses the decomposition of three treatments, which can be done in 3!=6 ways. Under the assumption of independence these different decomposition are identical. However, this is not the case when the treatments are correlated. Ohol et al. (2018) discusses a factorization approach developed by LeBoulluec et al. (2018) to estimate joint conditional probabilities of treatments. A need to develop such a method arises when the true joint distribution of treatments is unknown. The average factorization approach forms as a good approximation to the estimation. However, this approach is computationally intensive and induces noise in the estimation. As an alternative, greedy search algorithms using a minimization and maximization approach and another algorithm based on a greedy approach are employed. The greedy search algorithms search for a treatment that minimizes or maximizes the estimated probability conditioned on the treatments history, the remaining treatments in that stage, and the covariates. The algorithms run until all the treatments have been selected. The order

in which they are selected forms the order of the chain rule to calculate the joint conditional probability. This can be represented using the following equations and Algorithm 1.

Let, $\pi = i_1, i_2, ..., i_t$ be a given permutation of a number between $1$ and $t$ s.t. $\hat{p}(T_{i_1}, ..., T_{i_n}|\tilde{X}) = \hat{p}(T_{i_1}|\tilde{X}) * ... * \hat{p}(T_{i_n}|T_{i_1}, ..., T_{i_{n-1}}\tilde{X})$ Let, $\{T_t\}$ be the set of treatments

**Algorithm 1** Greedy Search Minimization Approach

0: **procedure** GREEDYMIN

0: $\mathrm{i}_k$ =argmin$_j \hat{p}(T_j|\{T_{tnj}\}, \tilde{X}),\ k = n, n-1, ..., 1$

0: Update $\{T_t\ = (T_t n T_{ik})$

0: Repeat 1 and 2 till $\{T_{tnj}\} = \emptyset$

0: **procedure** GREEDYMAX

0: $\mathrm{i}_k$ =argmax$_j \hat{p}(T_j|\{T_{tnj}\}, \tilde{X}),\ k = n, n-1, ..., 1$

0: Update $\{T_t\ = (T_t n T_{ik})$

0: Repeat 1 and 2 till $\{T_{tnj}\} = \emptyset$
=0

The third algorithm that operates on a greedy search technique was developed by DeBonet et al. (1997) and is called Mutual Information Maximizing Input Clustering (MIMIC). Let, $\pi = i_1, i_2, ..., i_t$ be a given permutation of a number between $1$ and $t$.

$\hat{p}_\pi(T) = p(T_{i_1}|T_{i_2})p(T_{i_2}|T_{i_3})...p(T_{i_{n-1}}|T_{i_n})p(T_{i_n})$.

The distribution of $\hat{p}_\pi(T)$ uses $\pi$ as an ordering for the pairwise conditional probabilities. The goal is to choose the permutation (chain rule ordering) that maximizes the agreement between $\hat{p}_\pi(T)$ and $p(T)$ using Kullback-Liebler divergence given by,

$$D(p||\hat{p}_\pi) = \int_t p[log(p) - log(\hat{p}_\pi)]dT = E_p[log(p)] - E_p[log(\hat{p}_\pi]$$

$$= -h(p) - E_p[log(p(T_{i_1}|T_{i_2})p(T_{i_2}|T_{i_3})...p(T_{i_{n-1}}|T_{i_n})p(T_{i_n}))]$$

$$= -h(p) + h(T_{i_1}|T_{i_2}) + h(T_{i_2}|T_{i_3}) + ... + h(T_{i_{n-1}}|T_{i_n}) + h(T_{i_n})$$

Because $h(p)$ does not depend on $\pi$, the cost function seeks to minimize,

$J_\pi(T) = h(T_{i_1}|T_{i_2}) + h(T_{i_2}|T_{i_3}) + ... + h(T_{i_{n-1}}|T_{i_n}) + h(T_{i_n})$.

Algorithm 2 summarizes the MIMIC approach. The computational complexity of the algorithm is $O(n^2)$.

**Algorithm 2** MIMIC Approach

0: **procedure** MIMIC

0: $i_n = \text{argmin}_j \hat{h}(T_j)$

0: $i_k = \text{argmin}_j \hat{h}(T_j|T_{i_{k+1}}),$ where $j \neq i_{k+1}...i_k$ and $k = n-1, n-2, ...2, 1$ =0

**Step 5: Calculating weights**

Once the joint conditional probabilities of treatment are estimated, the weights can be calculated using Equation 3.4. As mentioned earlier, logistic regression models are used to estimate the predicted probabilities of Equation 3.4. The weights are created for both stages and are used to create new state transition and outcome models.

**Step 6: Rebuild Models**

Once the weights are obtained, the initial models built in step 2 are refitted using the weights to obtain the weighted estimates of the treatment effects as shown below,

$$\hat{\beta_W} = (X^T W X)^{-1} X^T W Y,$$

where $W$ is the matrix of weights obtained using the IPTW method, and $\hat{\beta_W}$ holds the coefficient estimates of the weighted model.

These weighted state transition and outcome models are essentially weighted linear regression models or weighted piecewise linear models. As stated earlier, these weights represent duplicate copies of patient observations in a pseudo-population meaning. A weight for a particular patient is equivalent to the number of observations that the particular case represents. Building the weighted models is the final step in the framework. Performance of the framework is analyzed by comparing the estimated coefficients and standard errors for the weighted and un-weighted models. In addition, these models could be used in an optimization routine to assess their effect on decision making.

## 3.5 Case Study

The case studies presented in this section demonstrate the implementation of the generalized IPTW framework on cases with correlated treatments. The first case study is a simulation study to demonstrate the need for properly estimating the joint probabilities of treatment. If the true joint probability distribution of treatments is known, then the weights generated mitigate time varying confounding. The second case study implements the generalized IPTW framework on the McDermott pain management data. The weights are generated using MIMIC to estimate the joint probabilities of treatments.

### 3.5.1 Simulation study

The simulation study presented here is a continuation of the case study presented in the earlier work by Ohol et al. (2018). Their paper demonstrated the generation of joint probabilities of treatments using the average factorization approach and compared against the probabilities generated using the true joint distribution. This study will demonstrate the effectiveness of the estimation of joint probabilities in mitigating time varying confounding. The data generation process for the simulation study for sample size of 200 is explained using the equations below.

- Covariate $x_1$ follows a normal distribution with mean 0 and standard deviation 1.
- Covariate $x_2$ follows a binomial distribution with probability of success 0.5.
- Treatment 1 follows a binomial distribution with probabilities in stage 1 and 2 given below,

$$P(T_{11} = 1) = \frac{exp(0.4x_1 + 0.3x_2)}{1 + exp(0.4x_1 + 0.3x_2)}$$
$$P(T_{12} = 1) = \frac{exp(0.4x_1 + 0.3x_2 + 0.35t_{11} + 0.15y_1)}{1 + exp(0.4x_1 + 0.3x_2 + 0.35t_{11} + 0.15y_1)}.$$

- Treatment 2 follows a binomial distribution with probabilities in stage 1 and 2 given below,

$$P(T_{21} = 1) = \frac{exp(0.4x_1 + 0.3x_2 + t_{11})}{1 + exp(0.4x_1 + 0.3x_2 + t_{11})}$$
$$P(T_{22} = 1) = \frac{exp(0.4x_1 + 0.3x_2 + t_{12} + 0.35t_{21} + 0.15y_1)}{1 + exp(0.4x_1 + 0.3x_2 + t_{12} + 0.35t_{21} + 0.15y_1)}.$$

– Treatment 3 follows a binomial distribution with probabilities in stage 1 and 2 given below,

$$P(T_{31} = 1) = \frac{exp(0.4x_1 + 0.3x_2 + t_{11} + t_{21})}{1 + exp(0.4x_1 + 0.3x_2 + t_{11} + t_{21})}$$

$$P(T_{32} = 1) = \frac{exp(0.4x_1 + 0.3x_2 + t_{12} + t_{22} + 0.15y_1 + 0.35t_{31})}{1 + exp(0.4x_1 + 0.3x_2 + t_{12} + t_{22} + 0.15y_1 + 0.35t_{31})}.$$

– Thus, the true joint probability distribution can be decomposed as the product as follows: $P(T_1, T_2, T_3|X) = P(T_1|X)P(T_2|T_1, X)P(T_3|T_1, T_2, X)$.

– The outcome variable in stage 1 and stage 2 is generated by the following code. The signal to noise ratio is 10.

$$y_1 = 3 + 0.4x_1 + 0.3x_2 + t_{11} + t_{21} + t_{31} + \varepsilon(n, 0, 0.9).$$

$$y_2 = 3 + 0.4x_1 + 0.3x_2 + t_{12} + t_{22} + t_{32} - 0.5y_1 + \varepsilon(n, 0, 0.55).$$

The data generated using the above method is then used to estimate average joint probabilities, which are then compared against the true probabilities obtained by multiplying the product of the conditional probabilities generated using the equations above and shown in Figure 3.5.

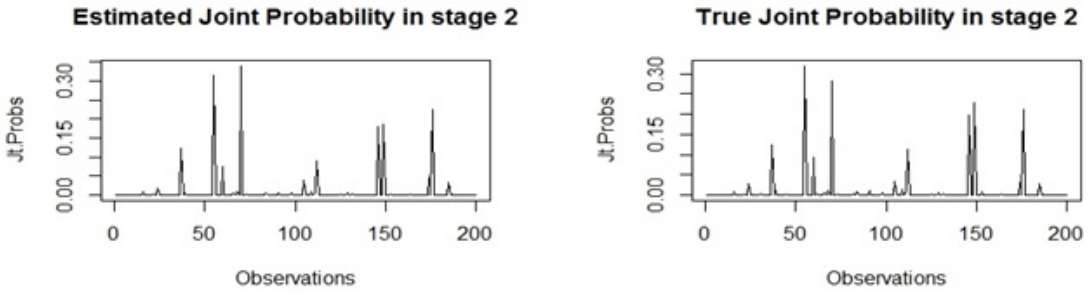



Figure 3.5: Estimated Joint Probabilities in Stage 2

It can be seen from the plots that the estimated joint probabilities closely follow the true joint probabilities. A similar inference was observed from the probability plots in stage 1. These probabilities were then used to fit the weighted models, and the parameter estimates and standard errors were compared. The results are shown in the Table 3.2.

The treatment coefficient estimates from the weighted model are closer to the true effect with lower standard errors than using unweighted model. An important modeling note here is that additive models are used for the analysis. Second order interactions are tested using a likelihood ratio test, and

| Treatments | True Effect | Unweighted | Weighted |
|---|---|---|---|
| $t_{12}$ | 1 | 1.07204 (0.05269) | 1.01052 (0.04686) |
| $t_{22}$ | 1 | 1.00642 (0.06456) | 1.02681 (0.05925) |
| $t_{32}$ | 1 | 1.0919 (0.07204) | 0.84141 (0.05641) |

Table 3.2: Parameter estimates (and standard errors) for simulated case with correlated treatments

| $t_{12}$ **- model type** | $x_1$ | $x_2$ |
|---|---|---|
| Unweighted | 0.3884 | 0.4015 |
| Weighted | 0.00473 | 0.01442 |

Table 3.3: Testing differences for treatment 1 in stage 2

the interactions are concluded as insignificant. Based on the true distribution,the joint probability of treatments are given in equation $P(T_1, T_2, T_3|X) = P(T_1|X)P(T_2|T_1, X)P(T_3|T_1, T_2, X)$ and aids the data generation process.

It can be seen that in stage 2, treatment 1's effect on the outcome variable $y_2$ is confounded by the covariates $x_1, x_2$ and previous treatment history, $t_{11}$. Treatment 2's effect on $y_2$ is confounded by the covariates, treatment 2's history in stage 1 and also by treatment $t_{12}$ in second stage, as per the decomposition of the true joint probability distribution. Finally, the variables confounding treatment 3's effect on $y_2$ are the covariates, treatment 3's history on previous stage $t_{31}$ and the other two treatments in stage 2 $t_{12}$ and $t_{22}$. The IPTW approach should mitigate the effect of these confounding variables. To test them, the stabilized weights for stage 2 are divided by treatments and the stabilized weights for each treatment in stage 2 are then used to test this difference effect. Tables 3.3, 3.4, 3.5 below summarize these results.

It can be seen from the tables above that the time varying confounding is significantly reduced among all the confounders. The effect is more prominent when the true joint conditional probability distribution is used since we can correctly identify the confounding factors. This effect is otherwise hard to notice in a problem with a huge feature space. Thus, it becomes increasingly important to identify the structure

| $t_{22}$ **- model type** | $x_1 - t_{21}$ | $x_2 - t_{21}$ | $x_1 - t_{12}$ | $x_1 - t_{12}$ |
|---|---|---|---|---|
| Unweighted | 0.418 | 0.578 | 0.6955 | 1.3524 |
| Weighted | 0.10976 | -0.1921 | 0.01763 | -0.043 |

Table 3.4: Testing differences for treatment 2 in stage 2

| $t_{32}$ **- model type** | $x_1 - t_{31}$ | $x_1 - t_{12}$ | $x_1 - t_{22}$ | $x_2 - t_{31}$ | $x_2 - t_{12}$ | $x_2 - t_{22}$ |
|---|---|---|---|---|---|---|
| Unweighted | 0.5051 | 0.51 | 0.6955 | 1.3524 | 0.7651 | 1.2114 |
| Weighted | -0.1584 | -0.0038 | -0.0834 | -0.1052 | 0.4678 | 0.5115 |

Table 3.5: Testing differences for treatment 3 in stage 2

of the true distribution of the treatment probability. The next case study shows implementation of the generalized IPTW framework on the McDermott adaptive interdisciplinary pain management data. In step 4 we estimate the joint probability of treatments using the methods explained in the previous section.

### 3.5.2 McDermott Adaptive Interdisciplinary Pain Management Case Study

This section discusses the application of the generalized IPTW framework to data from the Eugene McDermott Center for Pain Management at the University of Texas Southwestern Medical Center at Dallas. These data consist of patient information and treatments administered along with their observed outcome. A total of 294 observations were recorded between January 1998 and June 2007 LeBoulluec et al. (2013). LeBoulluec divided the dataset into training and testing sets using K-means clustering. The training dataset consists of 235 observations, and the testing dataset has the remaining 59 observations. It must be noted that since the data set is small, there are cases of rare observations that are included in the training data set. The complete dataset contains 53 patient variables, including as age, gender, litigation, marital status, race, physical history, past diagnosis, surgical history, etc. In addition, there are 14 stage 1 treatments and 13 stage 2 treatment variables. The outcome measure is used in this study is OSW and is recorded for pre-, mid-, and post-evaluation. The treatment variables are coded for this study and are summarized in Table 3.6 below.

$ProcGr11$ is only available as a treatment option in stage 1. Due to the interdisciplinary nature of the study, treatments are prescribed from both the groups and patients must receive at least one treatment in the course of the study. All the procedural treatments are binary variables (0: no, 1: yes), and the pharmaceutical treatments are categorical variables with two or three levels (coded as 0: no, 1, 2, 3). The patient variables are a mix of binary, categorical, and continuous variables The OSW

| Code | Procedural Treatments | Code | Pharmaceutical Treatments |
|---|---|---|---|
| ProcGr1 | Injections | RxGr1 | Tramadol |
| ProcGr2 | Block Procedures | RxGr2 | NSAIDs |
| ProcGr4 | Stimulation Procedure | RxGr3 | Narcotic |
| ProcGr9 | Cognitive Therapy | RxGr4 | Muscle Relaxant |
| ProcGr10 | Physical Therapy | RxGr5 | Antidepressant |
| ProcGr11 | Number of Additional Procedures | RxGr6 | Tranquilizer |
| | | RxGr7 | Sleeping Pills |
| | | RxGr8 | Others |

Table 3.6: Treatment Variables in Pain Management Data

score is defined in the earlier section and, as described, is treated as a continuous variable that can attain a maximum score of 50, indicating severe bedridden pain. The training dataset is used in the generalized IPTW framework, and the process is explained below.

**Step 1: Data investigation**

The raw dataset was obtained from the Eugene McDermott Center for Pain Management at the University of Texas Southwestern Medical Center at Dallas which contained missing values and other inconsistencies. A data processing tool was developed for the larger project to handle imputation and cleaning (refer to Figure 3.3). In this step the processed dataset was further investigated for any anomalies. It was noticed that six patients were not prescribed any treatment in any stage of the study. For these patients, there was no consistent pattern with the pain outcome, i.e., the OSW score increased first and then decreased across the stages for two patients, and otherwise for others. Thus, these observations were removed from the study after consulting with physician associated with the McDermott Center. The number of observations in the training data were reduced to 229.

**Step 2: Building preliminary models**

An additive model form is used in step. The preliminary model used contains all the variables in the study. Models for stage 1 and stage 2 are built using a linear model approach with the outcome variable $Post\_OSW$ being treated as continuous. The covariates are a mix of binary and continuous variables, and the treatments are binary.

**Step 3: Handling data issues - Data Separation**

A variety of data issues, such as sparse data, the presence of linear combinations of predictors, data separation, may exist for this case study. These problems need to be addressed before moving forward. Failing to address these problems could result in extremely high weights. As mentioned in the previous section, a small dataset with a large number of predictor variables will potentially give rise to data separation, especially when the dependent variables are binary or categorical, and thus needs to be addressed. The procedural and pharmaceutical treatments for the pain management data are binary variables. Data separation was a major problem identified in the pain management data. This step is thus further divided into subsections to explain how data separation was handled.

Description

Data separation in the pain management data was first encountered when logistic regression models were fit in step 3. It was observed that a few patients were assigned extremely high weights, in the magnitude of thousands, as compared to the rest. On further investigation, of the parameter estimates of these models, it was seen that the standard errors and coefficients of some of the predictors in the logistic models were very high indicating data separation. As a result of data separation, incorrect predicted probabilities are obtained. To understand the mathematical explanation behind the maximum likelihood estimates which are used to generate parameter estimates and standard errors. Let the log likelihood for Logit model be,

$$l(\beta) = \beta \sum_i \tilde{x}_i y_i - \sum_i ln[1 + exp(\beta(\tilde{x}_i)]$$

$$\begin{aligned} \frac{\partial l(\beta)}{\partial \beta} &= \sum_i \tilde{x}_i y_i - \sum_i \tilde{x}_i \hat{y}_i \\ where, \hat{y}_i &= \frac{1}{1 + exp(-\beta x_i)} \end{aligned} \tag{3.5}$$

$$\begin{aligned} \frac{\partial l(\beta)}{\partial \beta} &= 0 \\ \sum_i \tilde{x}_i y_i - \sum_i \tilde{x}_i \hat{y}_i &= 0 \end{aligned} \tag{3.6}$$

Since $\hat{\beta}$ is a vector, Equation 3.6 is a set of equations, one for each parameter to be estimated, identical to 'normal' equations for least square estimates in linear regression, estimated $y_i$ is nonlinear function of $x_i$. The explicit solution for $\hat{\beta}$ is,

$$\begin{aligned} &\hat{\beta} = ln(\frac{f_{11}f_{22}}{f_{12}f_{21}}) \\ &where, f_{ij} = observed\ cell\ frequency\ in\ 22\ table\ for\ a\ dichotomous\ x\ variable; \\ &i = 1, 2\ and\ j = 1, 2 \end{aligned} \tag{3.7}$$

When there is no explicit solution, numerical methods like Newton-Rhapson method are used to calculate parameter estimates and standard errors using an iterative algorithm. From Equation 3.7, it is evident that if at least one of the non-diagonal elements of the $22$ table is zero the maximum likelihood estimates does not exist. Thus the above explained approach was used to identify the treatment and patient variable pairs which cause data separation. It was also identified that the treatment variables are quasi-completely separated by the predictors.

Literature There are multiple ways to handle data separation. Quasi-complete separation, where only one of the non-diagonal frequencies in the $22$ table is zero, is mainly caused when an explanatory variable is a dummy variable. A first step is to check for such inconsistencies. If no such inconsistencies are found, other techniques should be employed. On further examination of the $22$ tables for the treatments, it was observed that some of the factor levels of the categorical treatments had very few observations. This was mainly due to the small size of the data. This indicated a potential to merge the categories of the treatments into binary, where 0: no treatment received and 1: treatment received. The binary treatments are used in generation of weights when logistic regression models are fit. While developing the state transition and outcome models, these categories are not collapsed. It would be appropriate at this juncture to highlight the importance of checking for data separation, even though the weights seemed to be adequate.

Proposed Solution: Elimination by grouping variables Grouping technique is considered to be a viable option to eliminate data separation and can be incorporated as a general step in the framework. However, the initial inception of the technique arises from the fact that there exists certain treatment prescription patterns in the data which needs further exploration under the guidance of a domain ex-

| **Treatments** | RxGr1 | RxGr2 | RxGr3 | RxGr4 | RxGr5 | RxGr6 | RxGr7 | RxGr8 |
|---|---|---|---|---|---|---|---|---|
| ProcGr1 | Moderate | Moderate | | | Mild | | | Moderate |
| ProcGr2 | | Moderate | | | Mild | | | Moderate |
| ProcGr4 | | | | | | | | |
| ProcGr9 | | | | | | | | |
| ProcGr10 | | | | | | | | |
| ProcGr11 | | | | | | | | |
| RxGr1 | | | Mild | Moderate | Moderate | Moderate | Moderate | Moderate |
| RxGr2 | | | | | | Moderate | | Moderate |
| RxGr3 | | | | Moderate | Moderate | Moderate | | Severe |
| RxGr4 | | | | | | Mild | Mild | Moderate |
| RxGr5 | | | | | | Severe | Mild | Moderate |
| RxGr6 | | | | | | | Mild | |
| RxGr7 | | | | | | | | Moderate |
| RxGr8 | | | | | | | | |

Table 3.7: Summary of treatment interactions

pert. Previous work done by Lin et al. (2013), LeBoulluec (2013), Wang (2015), Rawat and Manry (2017) and Iqbal (2017) considers certain predetermined interactions between treatments which acts as a catalyst for further analysis. Table3.7 below shows an interaction matrix highlighting mild moderate and severe interactions among treatments. These are the restrictions placed on the prescription of the treatments by the physician.

For example, there is a severe interaction between RxGr4 (Muscle Relaxant) and RxGr6 (Tranquilizer); thus, they are never prescribed together. Moderate interactions imply that the treatments are prescribed in rare cases. Mild interactions relax this notion to a higher degree.

Based on the previous information about treatment interactions, the nature of the pain management program and the available data, certain rules were instituted in the grouping blueprint. First, the treatments to be grouped would be similar across all the stages, since the adaptive treatment strategy is to devise a treatment regime, homogeneity across the stages is helpful for interpretation. For instance, if ProcGr9 (physical therapy) and ProcGr10 (cognitive behavioral therapy) are combined, they would be combined across all the stages. Hence, the correlation analysis of both stage 1 and 2 are considered in creating new groups. However, it must be noted that the IPTW framework can handle presence of

| Treatments - stage 1 | Interacting terms | Treatments - stage 1 | Interacting terms |
|---|---|---|---|
| ProcGr1_1 | ProcGr2_1 | ProcGr1_2 | ProcGr2_2 |
| | ProcGr4_1 | | ProcGr4_2 |
| | ProcGr11_1 | ProcGr2_2 | ProcGr4_2 |
| ProcGr2_1 | ProcGr4_1 | RxGr1_2 | RxGr3_2 |
| | ProcGr11_1 | RxGr3_2 | RxGr7_2 |
| ProcGr4_1 | ProcGr11_1 | | RxGr8_2 |
| RxGr1_1 | RxGr3_1 | | |
| | RxGr6_1 | | |
| | RxGr8_1 | | |
| RxGr3_1 | RxGr8_1 | | |
| RxGr7_1 | RxGr8_1 | | |

Table 3.8: Summary of treatment interactions

different decision variables in different stages. Next, the treatment variables considered for grouping are similar in characteristics. Since the pain management program is interdisciplinary in nature, a key element is to consider both pharmaceutical and procedural treatments. Hence, to maintain this attribute of the study, procedural treatments will not be considered for grouping with pharmaceutical variables. This same approach is applied while checking for the grouping of state variables.

Finally, the type of correlation and Variance Inflation (VIF) analysis for checking co-usages and interactions uses subsets of observations from the data. It was identified from further analysis that each patient is administered very few treatments. Thus there is sparseness in the treatment data. An exception was observed for cognitive behavioral techniques and physical therapies. To check for co-usages, all the available observations in the training data were used for the correlation and VIF analyses. However, to check for interactions, only the observations where at least one instance of a treatment in the pair of treatments under consideration were used. This mainly arises from the fact that pharmaceutical treatments are not administered as much and thus the data contain high number of '0' instances. Table 3.8 below summarizes the potential interactions among treatments to be considered for grouping.

To check for co-usages, a correlation analysis was performed on the treatments in stage 1 and stage 2 with all the observations. It was observed that the correlations between ProcGr9 and ProcGr10 were high in both stages, 0.8387 and 0.8418, respectively. All the other treatment pairs had low correlation

| Type | Parent Variables | New Variable | Type | Parent Variables | New Variable |
|---|---|---|---|---|---|
| Treatment | ProcGr9 | ProcGr12 | Patient Variables | race_1 | race |
| | ProcGr10 | | | race_2 | |
| | ProcGr2 | ProcGr13 | | phydx3 | phydx32 |
| | ProcGr4 | | | phydx8 | |
| | ProcGr11 | | | Phydx4 | phydx33 |
| | RxGr1 | RxGr9 | | phydx15 | |
| | RxGr3 | | | phydx20 | phydx34 |
| | RxGr7 | RxGr10 | | phydx31 | |
| | RxGr8 | | | pastdx4 | pastdx33 |
| | | | | pastdx14 | |
| | | | | SghxGr6 | SghxGr12 |
| | | | | SghxGr11 | |

Table 3.9: List of grouped treatments and patient variables

coefficients. On further consultation with the McDermott Center, these two variables were grouped together. VIFallow us to consider the treatment interactions in a multidimensional space. The correlation analysis helps identify multiple pairs of treatments that might be considered for grouping. VIFs are calculated for that particular treatment and all the other treatments with which had a high pairwise correlation. VIFs are calculated for each case when the three treatments are added in the model one by one. For example, it was observed that VIF values were high when ProcGr2_1, ProcGr4_1 and ProcGr11_1 are together, suggesting high multicollinearity. Table 3.9 summarizes the groups. It was observed that these new variables do not cause separation.

**Step 4: Estimating conditional joint probability of treatments**

The heat maps in Figures 3.6 and 3.7 show the correlation structure of the treatments in stages 1 and 2. It can be seen that there is mild to low correlation among the treatments in each stage after they are grouped based on interactions and co-usages.

With the new treatment groups, the correlation between the treatment variables has been reduced. As discussed in LeBoulluec et al. (2018) under the assumption of independence, the treatment order in the decomposition is irrelevant. To test independence for a treatment, the treatment is regressed with the other treatments and covariates and their significance is tested at a 0.05 alpha level. Based

on the tests the pairs of correlated treatments identified are summarized in Figures 3.8 and 3.9.

Based on this final correlation structure, the joint probabilities of treatments are estimated using the GreedyMin, GreedyMax and MIMIC algorithms. The decomposition of the joint probabilities of treatments is given in Table 3.10.

It can be seen from the tables that the treatment order using GreedyMin and GreedyMax is the opposite of each other. MIMIC generates yet another different ordering. However, MIMIC performed better in terms of computational time and the variation in the weights. Simulation studies were also developed to test the performance of the three algorithms. It was noticed that when the number of treatments increases, and when more covariates are included in the study, MIMIC successfully identifies the true distribution of joint probability of treatmens. with the increase in sample size. Hence, we decided to estimate weights based on the treatment order obtained by MIMIC.

**Step 5: Calculating weights**

The weights for stage 1 and stage 2 are calculated using the decomposition obtained from the MIMIC algorithm. The stabilized weights are calculated using Equation 3.4. An additive logistic regression model is fit to obtain predicted probabilities for every treatment. Figures 3.10, 3.11 show the weights obtained for stages 1 and 2. Weights were also obtained using the average factorization approach. The computational time to obtain average probabilities for all possible factorizations in a single stage was approximately 2.5 days on an Intel i7 processor with 12 gigabytes of RAM as compared to MIMIC which took several seconds.

It can be seen that the joint weights using the IPTW framework are lower in stage 1 than in stage 2. This is expected since more confounding is induced in stage 2 than stage 1. Moreover, the stabilized weights in stage 2 are a product of the weights in stage 1 and in stage 2. Thus, it is expected that some patients will have higher weights than others. Truncation is not used in this case since the weights seem reasonable and the current model helps identify some rare patient cases where the weights are high. For example, patient number 50 has a lower weight of 8 in stage 1; however, in stage 2, the patient is assigned a higher weight, approximately 70. This means that the probability of assigning the

| | |
|---|---|
| GreedyMax | |
| **Stage** | **Joint probability of treatments** |
| Stage 1 | P(ProcGr1_1, ProcGr12_1, ProcGr13_1, RxGr2_1, RxGr4_1, RxGr5_1, RxGr6_1, RxGr9_1, RxGr10_1\| X) = P(RxGr2_1\|X) P(RxGr9_1\|X) P(RxGr5_1\|RxGr2_1, X) P(ProcGr12_1\|X) P(ProcGr1_1\|RxGr2_1, X) P(RxGr4_1\|RxGr2_1, RxGr5_1, RxGr9_1,X) P(ProcGr13_1\|X) P(RxGr6_1\|ProcGr13_1, X) P(RxGr10_1\|RxGr5_1, X) |
| Stage 2 | P(ProcGr1_2,ProcGr12_2,ProcGr13_2,RxGr2_2,RxGr4_2,RxGr5_2,RxGr6_2,RxGr9_2,RxGr10_2\|X) = P(RxGr4_2\|X) P(ProcGr12_2\|X) P(RxGr2_2\|RxGr4_2,X) P(RxGr9_2\|RxGr4_2,X) P(ProcGr1_2\|RxGr4_2,RxGr9_2,X) P(RxGr5_2\|X) P(RxGr6_2\|RxGr4_2,RxGr5_2,RxGr9_2,X) P(ProcGr13_2\|RxGr4_2,ProcGr12_2,RxGr5_2,X) P(RxGr10_2\|RxGr2_2,RxGr5_2) |
| GreedyMin | |
| **Stage** | **Joint probability of treatments** |
| Stage 1 | P(ProcGr1_1,ProcGr12_1,ProcGr13_1,RxGr2_1,RxGr4_1,RxGr5_1,RxGr6_1,RxGr9_1,RxGr10_1\|X) = P(RxGr9_1\|X) P(RxGr5_1\|X) P(RxGr2_1\|RxGr5_1,X) P(RxGr4_1\| RxGr2_1,RxGr5_1,RxGr9_1,X) P(ProcGr12_1\|X) P(ProcGr1_1\|RxGr2_1,X) P(ProcGr13_1\|X) P(RxGr6_1\| ProcGr13_1,X) P(RxGr10_1\|RxGr5_1,X) |
| Stage 2 | P(ProcGr1_2,ProcGr12_2,ProcGr13_2,RxGr2_2,RxGr4_2,RxGr5_2,RxGr6_2,RxGr9_2,RxGr10_2\|X) = P(ProcGr12_2\|X) P(RxGr2_2\|X) P(RxGr4_2\|RxGr2_2,X) P(RxGr9_2\|RxGr4_2,X) P(ProcGr1_2\|RxGr4_2,RxGr9_2,X) P(RxGr5_2\|X) P(RxGr6_2\|RxGr4_2,RxGr5_2,RxGr9_2,X) P(RxGr10_2\|RxGr2_2,RxGr5_2) P(ProcGr13_2\|RxGr4_2,ProcGr12_2,RxGr5_2,X) |
| MIMIC | |
| **Stage** | **Joint probability of treatments** |
| Stage 1 | P(ProcGr1_1,ProcGr12_1,ProcGr13_1,RxGr2_1,RxGr4_1,RxGr5_1,RxGr6_1,RxGr9_1,RxGr10_1\|X) = P(RxGr6_1\|X) P(RxGr10_1\|X) P(ProcGr12_1\|X) P(RxGr9_1\|X) P(RxGr5_1\|X) P(RxGr4_1\|RxGr5_1,RxGr9_1,X) P(ProcGr13_1\| RxGr6_1,X) P(ProcGr1_1\|X) P(RxGr2_1\|ProcGr1_1,RxGr4_1,RxGr5_1X) |
| Stage 2 | P(ProcGr1_2,ProcGr12_2,ProcGr13_2,RxGr2_2,RxGr4_2,RxGr5_2,RxGr6_2,RxGr9_2,RxGr10_2\|X) = P(RxGr6_2\|X) P(RxGr9_2\|RxGr6_2,X) P(ProcGr1_2\|RxGr6_2,RxGr9_2,X) P(ProcGr13_2\|X) P(RxGr5_2\|ProcGr13_2,RxGr6_2,X) P(ProcGr12_2\|ProcGr13_2,X) P(RxGr4_2\|ProcGr13_2,ProcGr1_2,RxGr6_2,RxGr9_2,X) P(RxGr2_2\|RxGr4_2,X) P(RxGr10_2\|RxGr2_2,RxGr5_2,X) |

Table 3.10: List of grouped treatments and patient variables

treatment regime when the patient was in stage 2 given the patient's previous treatment assignment, physical characteristics and the patient's response to the previous treatment is low. This tells us that these types of patients are rare cases in the study and need to be further investigated. Thus, this method not only helps us to mitigate time varying confounding by assisting in estimation of causal treatment effects, but also helps identify some rare patient cases in the study. This could prove to be a useful piece of information for physicians and also critical input to the subsequent optimization step in the larger McDermott adaptive interdisciplinary pain management project (refer to Figure X3.3).

**Step 6: Rebuild models**

The final step involves rebuilding weighted models and comparing their performance against the unweighted models. The weighted models, similar to unweighted ones, have an additive linear model form and are built for both stage 1 and stage 2. Since the true parameter estimates are unknown, the method of comparison used in for case study is by comparing their standard errors. Figures 3.12, 3.13 compares the standard errors of treatments in stages 1 and 2.

It can be seen from the figures that the standard errors of the treatments are consistently lower in all but one treatment in stage 1 and for all the treatments in stage 2. It can be concluded that the generalized IPTW framework successfully mitigates time varying confounding in a two-stage observational study when the treatments are correlated. Under correlation it also stresses the importance of properly estimating the joint probabilities of treatments. Figures 3.14 and 3.15 compare the results of this study for all the variables with an initial study performed on McDermott adaptive interdisciplinary pain management data using an average factorization approach. It can be seen that the standard errors are higher in magnitude for the average factorization approach with standard errors higher for some variables for the weighted case than the unweighted. However, with the MIMIC approach not only are the standard errors reduced in magnitude, but also most of the variables in the weighted models have lower standard errors than their unweighted counterparts. This demonstrates the value of properly estimating joint probabilities of treatments.

## 3.6 Discussion and Future Work

This paper demonstrates the application of a generalized IPTW framework to address time varying confounding in an adaptive interdisciplinary pain management case study. The longitudinal and observational nature of the problem created confounding effects of treatments on patient's pain outcome, which create biased estimates. Earlier work has shown its implementation on multiple treatments under the assumption of independence. This paper discusses modifications to the framework for correlated treatments. This paper discussed approaches to estimate joint probabilities of treatments, namely GreedyMin, GreedyMax and MIMIC. All three show promising results, and are computationally efficient. The emphasis on the correct ordering in the chain rule of the joint probability calculation is demonstrated using a simulation study. pain management case study uses the generalized IPTW framework to successfully address the time varying confounding.

Although this approach shows promising results, additional work is currently underway to develop state transition and outcome models where the patients state variables are independent of each other. A correlation plot of the covariates showed small to mild correlation between some of the variables. With higher correlation, this could cause severe mulitcollinearity issues which might produces inefficient parameter estimates. Moreover, with increase in the number of variables and stages the size of the feature space becomes very large thus motivating to move towards a parsimonious model, especially if higher order interactions are included. However parsimony can inadvertently induce unmeasured confounder effect and thus more work is needed to develop these models. In some cases the treatment variables are highly imbalanced. The logistic regression models can produce biased results towards the majority class. This problem us different from the data separation and thus various sampling methods such as up-sampling, down-sampling or even synthetic minority oversampling SMOTE technique could be used to obtain balanced class. A tree based approach could also be used to handle the large feature space with correlated variables.

## Acknowledgement

This material is based upon work partially supported by the National Science Foundation under Grant CMII-1434401.

# Bibliography

Austin, P. (2013). The use of propensity score methods with survival or time-to-event outcomes: reporting measures of effect similar to those used in randomized experiments. *Statistics in Medicine*, 33:1242–1258.

Austin, P. and Schuster, T. (2016). The performance of different propensity score methods for estimating absolute effects of treatments on survival outcomes: A simulation study. *Statistical Methods in Medical Research*, 25(5):2214–2237.

Austin, P. and Stuart, E. (2015). Moving towards best practice when using inverse probability of treatment weighting (iptw) using the propensity score to estimate causal treatment effects in observational studies. *Statistics in Medicine*, 34:3661–3679.

Bodnar, L., Davidian, M., Siega-Riz, A., and Tsiatis, A. (2004). Marginal structural models for analyzing causal effects of time-dependent treatments: An application in perinatal epidemiology. *American Journal of Epidemiology*, 159(10):926–934.

Center, U. M. (2017). Pain management.

Cole, S. and Hernn, M. (2008). Constructing inverse probability weights for marginal structural models. *American Journal of Epidemiology*, 168(6):656–664.

Collins, L., Kugler, K., and Gwadz, M. (2016). Optimization of multicomponent behavioral and biobehavioral interventions for the prevention and treatment of hiv/aids. *AIDS Behav*, 20:197–214.

Collins, L., Murphy, S., Nair, V., and Strecher, V. (2005). A strategy for optimizing and evaluating behavioral interventions. *Annals of Behavioral Medicine*, 30(1):65–73.

Collins, L., Murphy, S., and Strecher, V. (2007). The multiphase optimization strategy (most) and the sequential multiple assignment randomized trial (smart): 162 new methods for more potent ehealth interventions. *American journal of preventive medicine*, 32(5):S112–S118.

Conway, J. and Higgins, I. (2011). Literature review: Models of care for pain management. Literature review report.

DAgostino, R. (2007). Estimating treatment effects using observational data. *Journal of the American Medical Association*, 297(3):314–316.

Daniel, R., Cousens, S., De Stavola, B., Kenward, M., and Sterne, J. (2013). Methods for dealing with tine-dependent confounding. *Statistics in Medicine*, 32:1584–1618.

Dawson, R. and Lavori, P. (2003). Comparison of designs for adaptive treatment strategies: Baseline vs. adaptive randomization. *Journal of Statistical Planning and Inference*, 117(2):356–385.

DeBonet, J., Isbell, C., and Viola, P. (1997). Mimic: Finding optima by estimating probability densities. Massachusetts Insititute Of Technology, Artificial Intelligence Laboratory Report.

DArcy, Y. (2007). Pain management: Evidence-based tools and techniques for nursing professionals. Marblehead, Mass.: HCPro.

Eccleston, Z. and Eccleston, C. (2004). Interdisciplinary management of adolescent chronic pain: developing the role of physiotherapy. *Physiotherapy*, 90(2):77–81.

Fewell, Z., Hernn, M., Wolfe, F., Tilling, K., Choi, H., and Sterne, J. (2004). Controlling for time-dependent confounding using marginal structural models. *The Stata Journal*, 4(4):402–420.

Flor, H., Fydrich, T., and Turk, D. (1992). Efficacy of multidisciplinary pain treatment centers: a meta-analytic review. *Pain*, 49(2):221–230.

Garcia-Aymerich, J., Lange, P., Serra, I., Schnohr, P., and Ant, J. (2008). Time-dependent confounding in the study of the effects of regular physical activity in chronic obstructive pulmonary disease: An application of the marginal structural model. *Annals of Epidemiology*, 18(10):775–783.

Gatchel, R., editor (2005). *Clinical essentials of pain management*. American Psychological Association, Washington, DC.

Gatchel, R. and Okifuji, A. (2006). Evidence-based scientific data documenting the treatment and cost-effectiveness of comprehensive pain programs for chronic nonmalignant pain. *Journal of Pain*, 7(11):779–793.

Gatchel, R., Peng, Y., Peters, M., Fuchs, P., and Turk, D. (2007). The biopsychosocial approach to chronic pain: Scientific advances and future directions. *Psychological Bulletin*, 133(4):581–624.

Gould, H., editor (2007). *Understanding Pain: What It Is, Why It Happens, and How Its Managed*. Demos Medical, New York, NY.

Guez, A., Vincent, R., Avoli, M., and Pineau, J. (2008). Adaptive treatment of epilepsy via batch-mode reinforcement learning. Proceedings of the 20th Conference on Innovative Applications of Artificial Intelligence (IAAI).

Gunlicks-Stoessel, M. (2016). Constructing smart to develop an adaptive treatment strategy for adolescent depression. Symposium: Journal of teh American Academy of Child and ADolescent Psychiatry.

Gwadz, M., Collins, L., Cleland, C., Leonard, N., Wilton, L., Gandhi, M., Braithwaite, S., Perlman, D., Kutnick, A., and Ritchie, A. (2017). Using the multiphase optimization strategy (most) to optimize an hiv care continuum intervention for vulnerable populations: a study protocol. *BMC Public Health*, 17(383).

Hernan, M., Brumback, B., and Robins, J. (2001). Marginal structural models to estimate the joint causal effect of nonrandomized treatments. *Journal of the American Statistical Association*, 96(454):440–448.

Hogan, J. and Lancaster, T. (2004). Instrumental variables and inverse probability weighted for causal inference from longitudinal observational studies. *Statistical Methods in Medical Research*, 13(1):17–48.

Hong, G. and Raudenbush, S. (2008). Causal inference for time-varying instructional treatments. *Journal of Education and Behavioral Statistics*, 33(3):333–362.

Howard, M. and Jacobs, R. (2016). The multiphase optimization strategy (most) and the sequential multiple assignment randomized trial (smart): two novel evaluation methods for developing optimal training program. *Journal of American Statistical Association*, 37:1246–1270.

Huber, M. (2014). Identifying causal mechanisms (primarily) based on inverse probability weighting. *Journal of Applied Economics*, 29:920–943.

Iqbal, G. M. D. (2017). Multi-objective two-stage stochastic programming for adaptive interdisciplinary pain management with piece-wise linear network transition models. Ph.D. Dissertation, Department of Industrial and Manufacturing Systems Engineering, The University of Texas at Arlington, Arlington, Tx.

Joffe, M., Ten Have, T., Feldman, H., and Kimmel, S. (2004). Model selection, confounder control, and marginal structural models: Review and new applications. *American Statistician*, 58(4):272–279.

Kaelbling, L., Littman, M., and Moore, A. (1996). Reinforcement learning: A survey. *Journal of Artificial Intelligence Research*, 4:237–285.

Klungel, O., Martens, E., Psaty, B., Grobbee, D., Sullivan, S., Stricker, B., Leufkens, H., and de Boer, A. (2004). Methods to assess intended effects of drug treatment in observational studies are reviewed. *Journal of Clinical Epidemiology*, 57:1223–1231.

Kugler, K., Balantekin, K., Birch, L., and Savage, J. (2016). Application of the multiphase optimization strategy to a pilot study: an empirical example targeting obesity among children of low-income mothers. *BMC Public Health*, 16(1181).

Lavori, P. and Dawson, R. (1998). Developing and comparing treatment strategies: An annotated portfolio of designs. *Psychopharm Bull*, 34(1):1318.

Lavori, P., Rush, A., Wisniewski, S., Alpert, J., Fava, M., Kupfer, D., Nierenberg, A.and Quitkin, F., Sackeim, H., Thase, M., and Trivedi, M. (2001). Strengthening clinical effectiveness trials: equipoise-stratified randomization. *Biological Psychiatry*, 50(10):792801.

LeBoulluec, A. (2013). Outcome and state transition modeling for adaptive interdisciplinary pain management. Ph.D. Dissertation, Department of Industrial and Manufacturing Systems Engineering, The University of Texas at Arlington, Arlington, Tx.

LeBoulluec, A., Ohol, N., Chen, V., Zeng, L., Rosenberger, J., and Gatchel, R. (2018). Handling time-varying confounding in state transition models for dynamic optimization of adaptive interdisciplinary pain mangement. *IISE Transactions on Healthcare Systems Engineering*, 0:1–10.

LeBoulluec, A., Zeng, L., Chen, V., Rosenberger, J., and Gatchel, R. (2013). Outcome and state transition modeling for adaptive interdisciplinary pain management. Proceeding of 2013 Industrial and Systems Engineering Research Conference (ISERC).

Lee, J. and Lee, J. (2004). Approximate dynamic programming strategies and their applicability for process control: A review and future directions. *International Journal of Control, Automation, and System*, 2(3):267–278.

Lin, C. (2010). Adaptive pain management decision support system. Ph.D. Dissertation, The University of Texas at Arlington.

Lin, C., LeBoulluec, A., Zeng, L., Chen, V., and Gatchel, R. (2013). An adaptive pain management framework. *Health Care Management Science*, 17(3):270–283.

Main, C. and Spanswick, C., editors (2000). *Pain Management: An Interdisciplinary Approach*. Churchill Livingstone, London, UK.

MedicineNet (1996). Chronic pain.

Melzack, R. and Wall, P. (1965). Pain mechanisms a new theory. *Science*, 150(3699):971–979.

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

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

Murphy, S. (2005). An experimental design for the development of adaptive treatment strategies. *Stat Med*, 24:14551481.

Murphy, S. and Bingham, D. (2009). Screening experiments for developing dynamic treatment regimes. *Journal of American Statistical Association*, 184:391–408.

Murphy, S., Collins, L., and Rush, A. (2007). Customizing treatment to the patient: Adaptive treatment strategies. *Drug and Alcohol Dependence*, 88:S1–S3.

Nahum-Shani, I., Ertefaie, A., Lu, X., Lynch, K., McKay, J., D.W., O., and Almirall, D. (2016). A smart data analysis method for constructing adaptive treatment strategies for substance use disorders. *Society for the Study of Addiction*, (112):901–909.

of Pain Medicine, A. A. (2017). Aapm facts and figures on pain.

Ohol, N., Chen, V., LeBoulluec, A., Rosenberger, J., and Gatchel, R. (2018). A study on time-varying confounding with multiple treatments. UT Arlington Technical Report.

Olason, M. (2004). Outcome of an interdisciplinary pain management program in a rehabilitation clinic. *Work*, 22:9–15.

Pineau, J., Bellemare, M., Rush, A., Ghizaru, A., and Murphy, S. (2007). Constructing evidence-based treatment strategies using methods from computer science. *Drug and Alcohol Dependence*, 88(2):S52–S60.

Rawat, R. and Manry, M. (2017). Second order training of a smoothed piecewise linear network. Neural Processing Letters.

Robins, J. (1986). A new approach to causal inference in mortality studies with sustained exposure periods-application to control of the health worker survivor effect. *Computers and Mathematics with Applications*, 14:1393–1512.

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

Robins, J. and Berkane, M., editors (1997). *Causal Inference from complex longitudinal data. Latent Variable Modeling and Applications to Causality. Lecture Notes in Statistics (120)*. Springer- Verlag Inc, New York, NY.

Robins, J. and Greenland, S. (1994). Adjusting for differential rates of pcp prophylaxis in high- versus low-dose azt treatment arms in an aids randomized trial. *Journal of the American Statistical Association*, 89:737–749.

Robins, J., Hernan, M., and Brumback, B. (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology*, 11(5):550–560.

Rosenblum, A., Marsch, L., Joseph, H., and Portenoy, R. (2008). Opiods and treatment of chronic pain: Controversies, current status, and future directions. *HHS Public Access*, pages 405–416.

Schatman, M. and Campbell, A., editors (2007). *Chronic Pain Management Guidelines for Multidisciplinary Program Development*. Informal Healthcare, New York, NY.

Shavitt, R. (2016). Adaptive treatment strategies for children and adolescents with obsessive-compulsive disorder. Symposium: Journal of teh American Academy of Child and ADolescent Psychiatry.

Shortreed, S., Laber, E., Lizotte, D., Stroup, T., Pineau, J., and Murphy, S. (2011). Informing sequential clinical decision-making through reinforcement learning: An empirical study. *Machine Learning*, 84(1):109–136.

Si, J., Barto, A., Powell, W., and Wunsch, D., editors (2004). *Handbook of Learning and Approximate Dynamic Programming*. Wiley, New York, NY.

SPINE-health (1999). Chronic pain as a disease: Why does it still hurt?

Sutton, R. and Barto, A., editors (1998). *Reinforcement Learning: An Introduction*. MIT Press, Cambridge, MA.

Theommes, F. and Ong, A. (2016). A primer on inverse porobability of treatment weighting and marginal structural models. *Emerging Adulthood*, 4(1):40–59.

VanderWeele, T. (2009). Marginal structural models for the estimation of direct and indirect effects. *Epidemiology*, 20(1):18–26.

Wang, N. (2015). Two-stage stochastic programming for adaptive interdisciplinary pain management. Ph.D. Dissertation, Department of Industrial and Manufacturing Systems Engineering, The University of Texas at Arlington, Arlington, Tx.

WebMD (2005). Do i have chronic pain.

Weitzen, S., Lapane, K., Toledano, A., Hume, A., and Mor, V. (2004). Principles for modeling propensity scores in medical research: A systematic literature review. *Pharmacoepidemiology and Drug Safety*, 13:841–853.

Werbos, P. (1974). Beyond regression: New tools for prediction and analysis in the behavioral sciences. Ph.D. Dissertation, Committee on Applied Mathematics, Harvard University.

Werbos, P. (2007). Using adp to understand and replicate brain intelligence: The next level design. Proceedings of the 2007 IEEE Symposium on Approximate Dynamic Programming and Reinforcement Learning (ADPRL 2007).

Windmeijer, F. and Santos Silva, J. (1997). Endogeneity in count data models: An application to demand for health care. *Journal Of Applied Econometrics*, 12:281–294.

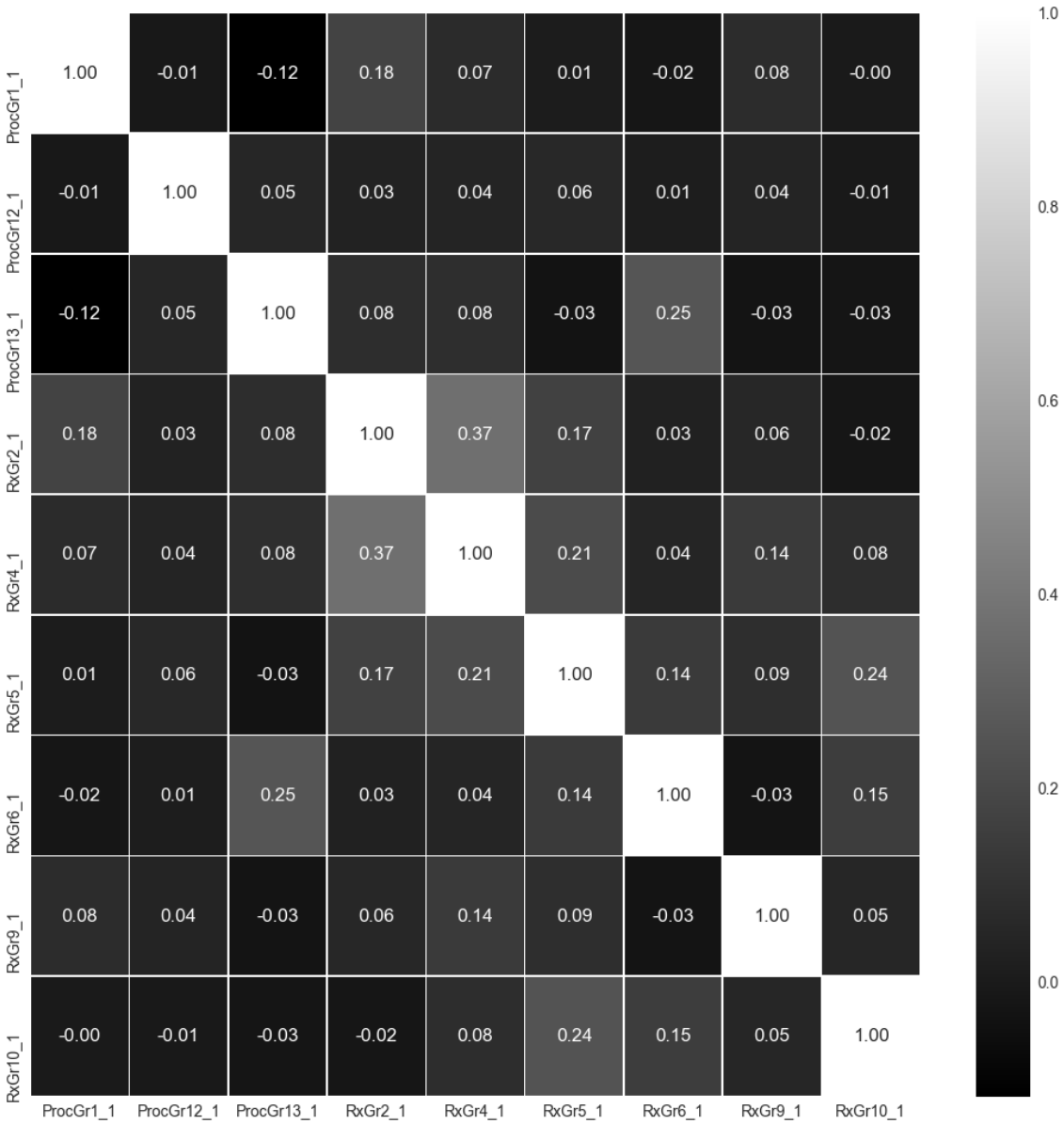



Figure 3.6: Heatmap for Treatment Correlation in Stage 1

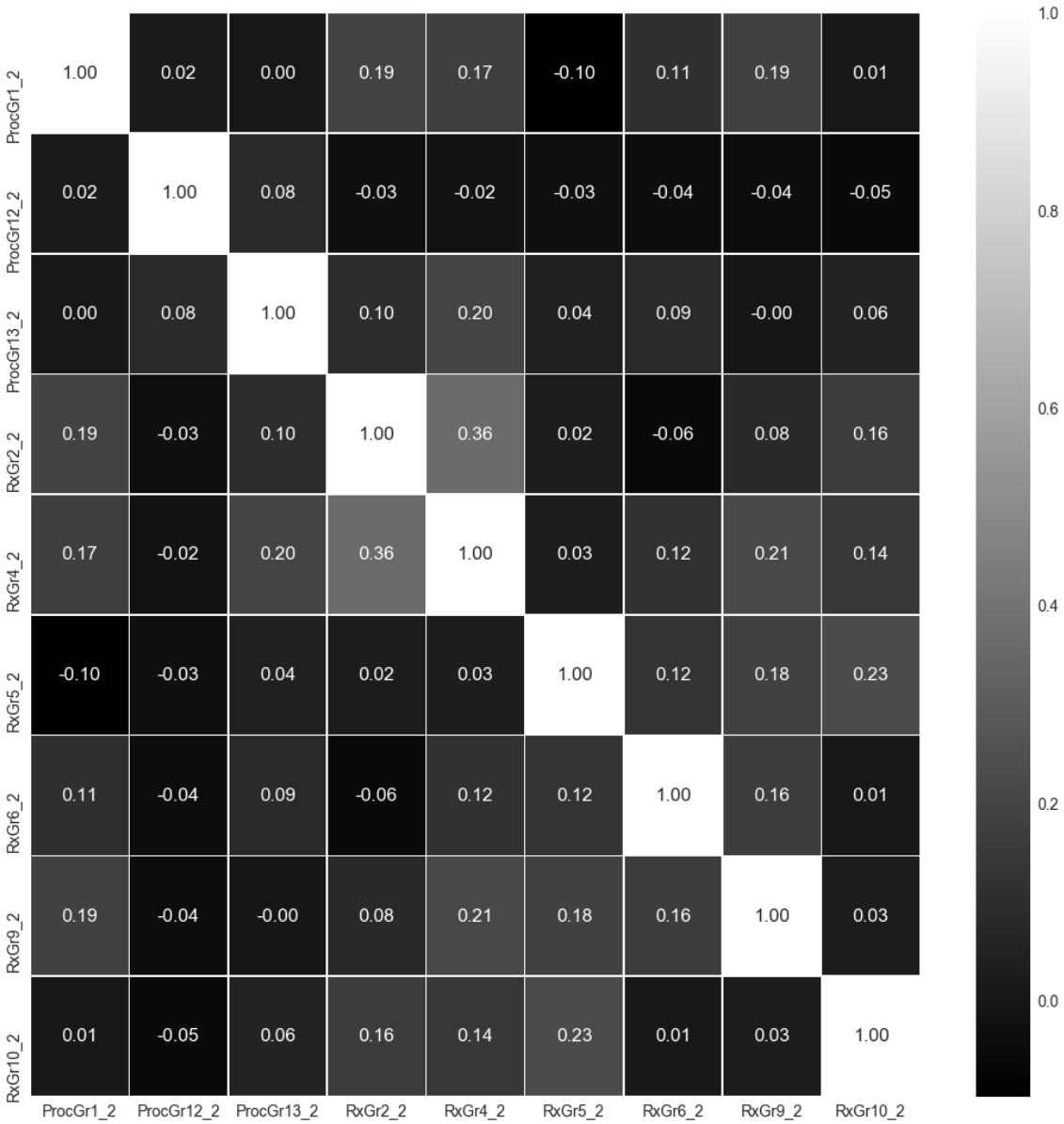



Figure 3.7: Heatmap for Treatment Correlation in Stage 2

| Stage 1 | ProcGr1_1 | ProcGr12_ | ProcGr13_ | RxGr2_1 | RxGr4_1 | RxGr5_1 | RxGr6_1 | RxGr9_1 | RxGr10_1 |
|---|---|---|---|---|---|---|---|---|---|
| ProcGr1_1 | | | | | | | | | |
| ProcGr12_1 | | | | | | | | | |
| ProcGr13_1 | | | | | | | | | |
| RxGr2_1 | | | | | | | | | |
| RxGr4_1 | | | | | | | | | |
| RxGr5_1 | | | | | | | | | |
| RxGr6_1 | | | | | | | | | |
| RxGr9_1 | | | | | | | | | |
| RxGr10_1 | | | | | | | | | |

Figure 3.8: Final Treatment Independence Structure in Stage 1

| Stage 2 | ProcGr1_2 | ProcGr12_ | ProcGr13_ | RxGr2_2 | RxGr4_2 | RxGr5_2 | RxGr6_2 | RxGr9_2 | RxGr10_2 |
|---|---|---|---|---|---|---|---|---|---|
| ProcGr1_2 | | | | | | | | | |
| ProcGr12_2 | | | | | | | | | |
| ProcGr13_2 | | | | | | | | | |
| RxGr2_2 | | | | | | | | | |
| RxGr4_2 | | | | | | | | | |
| RxGr5_2 | | | | | | | | | |
| RxGr6_2 | | | | | | | | | |
| RxGr9_2 | | | | | | | | | |
| RxGr10_2 | | | | | | | | | |

Figure 3.9: Final Treatment Independence Structure in Stage 2

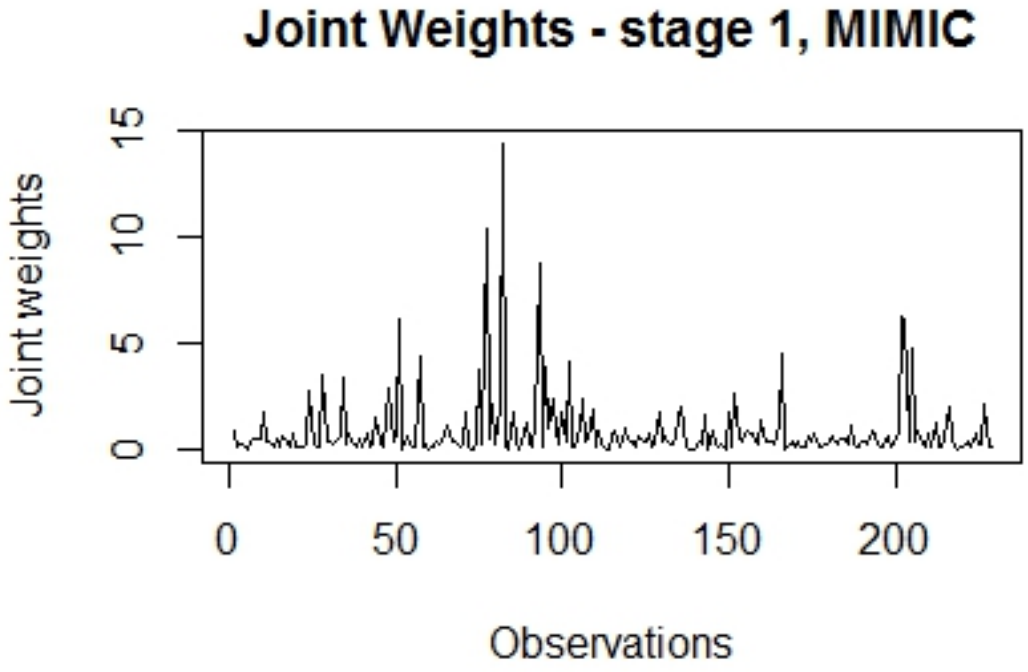



Figure 3.10: Estimated Weights in Stage 1

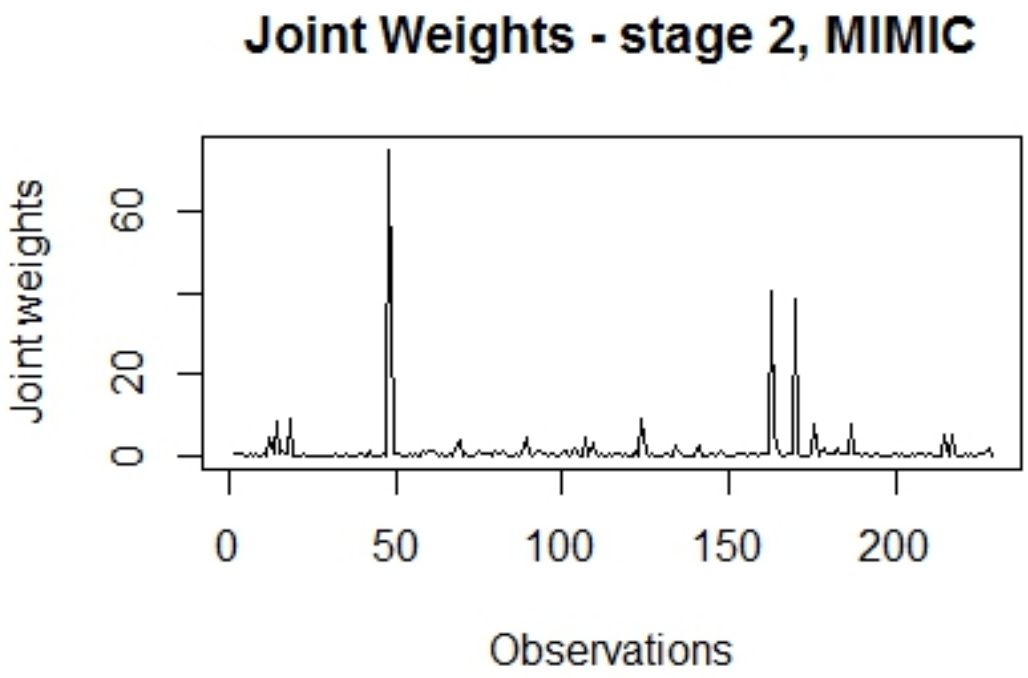



Figure 3.11: Estimated Weights in Stage 2

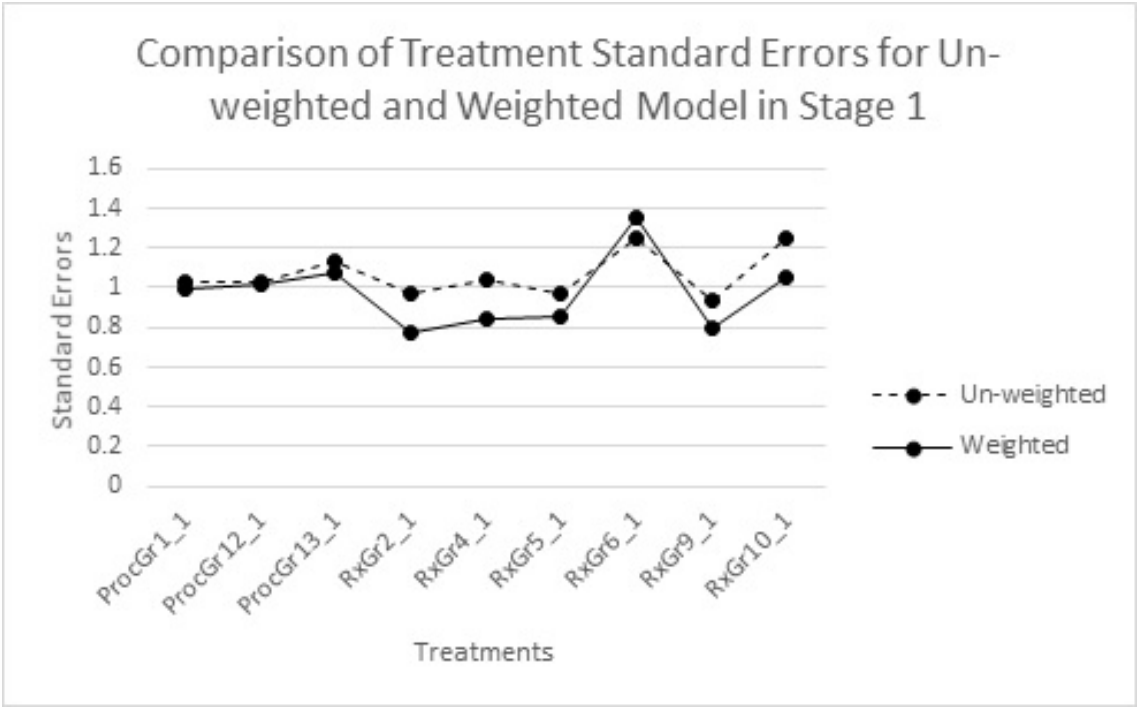



Figure 3.12: Standard Error Comparison in Stage 1

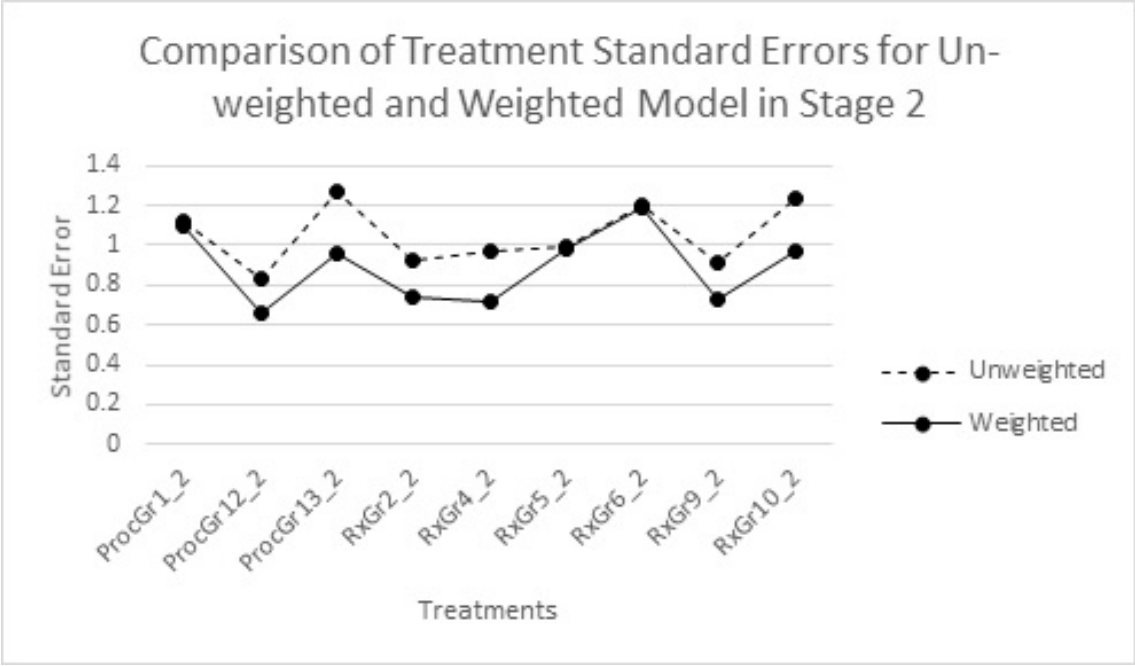



Figure 3.13: Standard Error Comparison in Stage 2

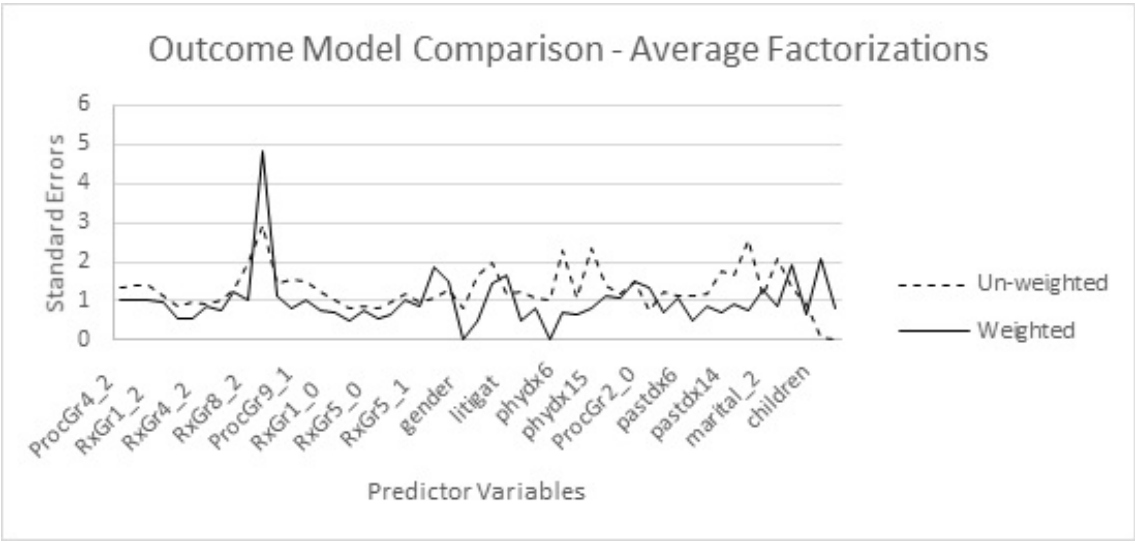



Figure 3.14: Comparison of Standard Errors in Outcome Models - Average Factorizations

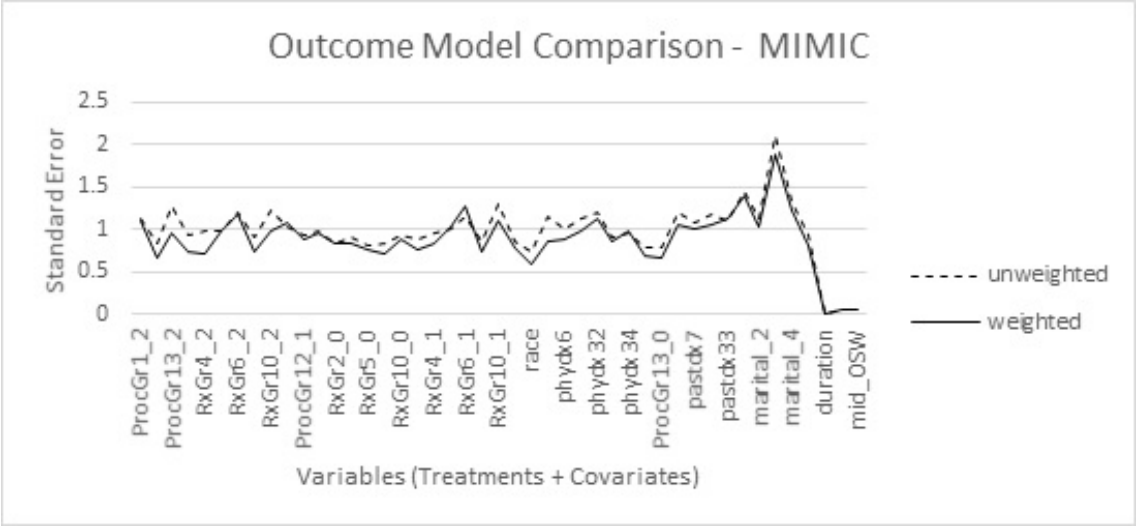



Figure 3.15: Comparison of Standard Errors in Outcome Models - MIMIC

# Chapter 4

# Conclusion and Future Work

Chapter 2, through a framework of simulation studies, demonstrates how to mitigate time varying confounding in observational studies. Under the assumption of unmeasured confounders, a generalized IPTW process is developed and implemented to address time varying confounding with multiple treatments. In case of independent treatments, this approach is straightforward since the joint probabilities of treatments can be decomposed as the product of their marginals, conditioned on the covariates. This is demonstrated using an example with three binary treatments. Correlated treatments are estimated by decomposing them using the chain rule of probability. The conditional joint probabilities can be decomposed using multiple orderings and, thus, without knowledge of the true ordering, these joint probabilities are estimated by averaging across factorizations using all the possible orderings. However, the number of possible factorizations increases with the number of treatments. This is demonstrated using three examples.

The above mentioned factorization approach fails under correlated example 2 because of high correlations between treatments that lead to data separation. This problem is addressed in chapter 3, in which the application of the generalized IPTW process successfully addresses time varying confounding for an interdisciplinary adaptive pain management program. This paper discusses alternatives to the factorization approach to estimate the joint probability for IPTW. These approaches are more catered towards solving real world problems and take into account the component of uncertainty. The methods, namely GreedyMin, GreedyMax and MIMIC, not only show promising results, but are

also computationally efficient, especially compared to the factorization approach. The emphasis on the correct ordering of the chain rule for the joint probability of treatment is demonstrated using a simulation study.

Although the generalized IPTW process, in particular, using MIMIC, shows promising results, additional work is currently underway to develop state transition and outcome models . A correlation plot of the covariates showed small to mild correlation between some of the variables for the pain management case study. With higher correlation, this could cause severe mulitcollinearity issues that might cause data separation. Moreover, an increase in the numbers of variables and stages generates a very large feature space, especially if higher order interactions are included. This motivates a desire for a parsimonious model. However, parsimony can inadvertently induce unmeasured confounder effects, and, thus, more work is needed to develop these models.

In some cases the treatment variables are highly imbalanced. The logistic regression models can produce biased results towards the majority class. Possible approaches to mitigate this include various sampling methods, such as up-sampling, down-sampling or the synthetic minority oversampling SMOTE technique. A tree based approach could also be used to handle the large feature space with correlated variables.