LASSO Based State Transition Modeling with Interactions in Adaptive Interdisciplinary Pain Management

By

Amith Viswanatha[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 2022

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

Acknowledgments

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

Acknowledgments

I would like to thank my advisors Dr Victoria Chen and Dr Jay Rosenberger for guiding me through my doctoral journey. They have played a pivotal role in my life and whatever research work I have completed is a direct result of their support, feedback, and encouragement. Their knowledge and experience are an asset for any student, and I am grateful to them for sharing it with me. They opened the door to the research world of COSMOS in the Summer of 2016, when as a master's student I started working with Dr Nilabh Ohol and Dr Gazi Iqbal, both doctoral students at the time, on the pain management data. This experience prompted me to apply for the PhD program and I cannot thank them enough for agreeing to be my advisors and shaping my research career.

I would like to thank Dr Aera LeBoulluec for her support and guidance. I would also like to thank Dr Carl Noe, MD, for agreeing to be on my dissertation committee and providing valuable feedback on the research. I would also like to thank Dr Yuan Zhou for being a part of the dissertation committee and supporting me as a Graduate Research Assistant for the past two years.

I would like to thank my fellow COSMOSians Dr Nilabh Ohol, Dr Gazi Iqbal, Dr Ashkan Farahani, Dr Shirish Rao, Dr Srividya Sekar, Dr Alireza Fallahi, Dr Ukesh Chawal and Dr Nahal Sakhavanand for being my friends and mentors. I would like to thank Richard Zercher and Ann Hoang for providing technical and administrative support from the department.

I would finally like to thank my parents, family, and friends for being there and supporting me through the ups and downs of life. I would like to end my acknowledgement with the quote from the marathon world record holder Eliud Kipchoge, "100% of me is nothing compared to 1% of the whole team". I would like to thank everyone for being a part of my team.

Abstract

LASSO Based State Transition Modeling with Interactions in Adaptive Interdisciplinary Pain Management

Amith Viswanatha

The University of Texas at Arlington, 2022

Supervising Professors: Dr Victoria Chen, Dr Jay Rosenberger

The Eugene McDermott Center for Pain Management at the University of Texas Southwestern Medical Center has an interdisciplinary pain management program for chronic pain. This program treats patients with a holistic view of reducing chronic pain and improving their physical, mental, and social well-being through treatment interventions. The development of an adaptive treatment decision tool is main goal of the research project.

This program is modeled as a two-stage adaptive treatment decision problem, with state transition models representing the transition of patient state, treatment, and outcome variables from stage 1 to stage 2. Interactions between the patient state and treatments play a major role in determining a personalized treatment plan for individual patients. In this research, we address the challenge of modeling state-treatment interactions. We propose a LASSO based approach to develop the state transition models. The proposed approach is studied using a simulated case study structured based on the McDermott Center data. The state transition models built using the proposed method are then formulated within the multi-objective two-stage stochastic programming optimization to obtain an optimal treatment plan.

Table of Contents

1 Introduction 1

2 Literature Review **7**

2.1 Pain Management Program 7

2.2 Multi-Stage Optimization 9

2.3 State Transition Models 13

2.4 Feature and Interaction Selection 16

2.4.1 LASSO 17

2.4.2 Interaction Selection 19

2.5 Contribution 20

3 LASSO Based State Transition Modeling with Interactions in Pain Management Simulation Case Study 24

3.1 Introduction 25

3.1.1 Addressing Time-varying Confounding 26

3.1.2 Optimization Approaches for Adaptive Interdisciplinary Pain Management 28

3.1.3 Contribution in State Transition and Outcome Modeling 28

3.2 Modeling Framework 29

3.2.1 HierNet 29

3.2.2 HierNet-IPTW Modeling Framework 30

3.3 Pain Management Simulation Case Study 32

3.4 Performance Metrics 35

3.4.1 Feature/Variable Selection Metrics 35

3.4.2 Interaction Selection Metrics 36

3.5 Simulation Experiment Results 37

3.5.1 Feature Selection Sensitivity 37

3.5.1.1 Preliminary Analysis 37

3.5.1.2 Analysis of Variance (ANOVA) 40

3.5.2 Feature Selection Specificity 45

3.5.2.1 Preliminary Analysis 45

3.5.2.2 Analysis of Variance (ANOVA) 47

3.5.3 Interaction Selection Sensitivity 52

3.5.3.1 Preliminary Analysis 52

3.5.3.2 Analysis of Variance (ANOVA) 54
3.5.4 Interaction Selection FDR 59
3.5.4.1 Preliminary Analysis 59
3.5.4.2 Analysis of Variance (ANOVA) 61
3.6 Conclusion and Future Work 66
4 Using LASSO based State Transition Model with Interactions in Multiple Objective Two-Stage Stochastic Programming for Adaptive Interdisciplinary Pain Management 68
4.1 Introduction 69
4.2 Stochastic Programming Formulation 73
4.3 Treatment Usage Analysis 75
4.3.1 Stage 1 Treatment Analysis 75
4.3.1.1 Stage 1 2SP IQ-IPTW vs 2SP IQ Treatment Analysis 76
4.3.1.2 Stage 1 2SP IQ-IPTW vs 2SP AL-IPTW Treatment Analysis 77
4.3.1.3 Stage 1 Treatment Analysis between low and high IPTW weight patients 78
4.3.2 Stage 2 Treatment Analysis 81
4.3.1.1 Stage 1 2SP IQ-IPTW vs 2SP IQ Treatment Analysis 81
4.3.1.2 Stage 1 2SP IQ-IPTW vs 2SP AL-IPTW Treatment Analysis 82
4.3.1.3 Stage 1 Treatment Analysis between low and high IPTW weight patients 83
4.4 Conclusion and Future Work 86
5 Discussion and Future Work 88

List of Figures

1 Biopsychosocial model of pain [8] 2

2 Two-stage interdisciplinary pain management program [22] 3

3 Adaptive interdisciplinary pain management research objectives [26] 4

4 Adaptive interdisciplinary pain management framework [26] 20

5 Adaptive interdisciplinary pain management research overview [26] 25

6 Time-varying confounding in one treatment case [26,168] 27

7 HierNet-IPTW modeling flowchart 30

8 Baseline model flowchart 31

9 Pain Management Causal Diagram [168] 32

10 Correlation Design for the simulation case study [168] 33

11 Feature selection sensitivity against the use of IPTW for different correlation structures between treatments 38

12 Feature selection sensitivity against the use of IPTW for different correlation structures between treatments and covariates 39

13 Feature selection sensitivity against the use of IPTW for different correlation structures between treatments and spurious variables 40

14 Tukey line plot for feature selection sensitivity 42

15 Feature selection specificity against the use of IPTW for different correlation structures between treatments 45

16 Feature selection specificity against the use of IPTW for different correlation structures between treatments and covariates 46

17 Feature selection specificity against the use of IPTW for different correlation structures between treatments and spurious variables 47

18 Tukey line plot for feature selection specificity 50

19 Interaction selection sensitivity against the use of IPTW for different correlation structures between treatments 52

20 Interaction selection sensitivity against the use of IPTW for different correlation structures between treatments and covariates 53

21 Interaction selection sensitivity against the use of IPTW for different correlation structures between treatments and spurious variables 54

22 Tukey line plot for interaction selection sensitivity 57

23 Interaction selection FDR against the use of IPTW for different correlation structures between treatments 59

24 Interaction selection FDR against the use of IPTW for different correlation structures between treatments and covariates 60

25 Interaction selection FDR against the use of IPTW for different correlation structures between treatments and spurious variables 61

26 Tukey line plot for interaction selection FDR 64

27 Two-stage interdisciplinary pain management program [22] 69

28 Stage 1 Treatment Usage 2SP IQ-IPTW, 2SP IQ, and observed data 77

29 Stage 1 Treatment Usage 2SP IQ-IPTW, 2SP AL-IPTW, and observed data 78

30 Stage 1 Treatment Usage 2SP IQ for Low and High IPTW weight patients 79

31 Stage 1 Treatment Usage 2SP AL-IPTW for Low and High IPTW weight patients 79

32 Stage 1 Treatment Usage 2SP HierNet-IPTW for Low and High IPTW weight patients 80

33(a) Stage 1 treatment assignment comparison for low IPTW weight patients 81

33(b) Stage 1 treatment assignment comparison for high IPTW weight patients 81

34 Stage 2 Treatment Usage 2SP IQ-IPTW, 2SP IQ, and observed data 82

35 Stage 2 Treatment Usage 2SP IQ-IPTW, 2SP AL-IPTW, and observed data 83

36 Stage 2 Treatment Usage 2SP IQ for Low and High IPTW weight patients 84

37 Stage 2 Treatment Usage 2SP AL-IPTW for Low and High IPTW weight patients 84

38 Stage 2 Treatment Usage 2SP HierNet-IPTW for Low and High IPTW weight patients 85

39(a) Stage 2 treatment assignment comparison for low IPTW weight patients 86

39(b) Stage 2 treatment assignment comparison for high IPTW weight patients 86

40 Residuals versus experiment treatments feature selection sensitivity 90

41 Normal Probability Plot (Q-Q Plot) feature selection sensitivity 91

42 Residuals versus experimental treatments for feature selection specificity 91

43 Normal Probability Plot (Q-Q Plot) feature selection specificity 92

44 Residuals versus experimental treatments for interaction selection sensitivity 92

45 Normal Probability Plot (Q-Q Plot) interaction selection sensitivity 93

46 Residuals versus experimental treatments for interaction selection FDR 93

47 Normal Probability Plot (Q-Q Plot) interaction selection FDR 94

List of Tables

1 Pain management variables 10
2 State and outcome Transition [25] 21
3 Pain Management Research Summary 21
4 Other factors in pain management simulation case study 34
5 Confusion matrix for feature selection [171] 35
6 Confusion matrix for interaction selection 36
7 ANOVA for feature selection sensitivity 41
8 Tukey comparison factor combinations for feature selection sensitivity 41
9 Tukey analysis p-value for feature selection sensitivity 42
10 Tukey analysis 90% CI for feature selection sensitivity 42
11 Comparison of feature selection sensitivity means with the use of IPTW 43
12 Contrast for simultaneous Bonferroni comparison 44
13 Simultaneous Bonferroni comparison results for feature selection sensitivity 44
14 ANOVA for feature selection specificity 48
15 Tukey comparison factor combinations for feature selection specificity 49
16 Tukey analysis p-value for feature selection specificity 49
17 Tukey analysis 90% CI for feature selection specificity 50
18 Comparison of feature selection specificity means with the use of IPTW 51
19 Simultaneous Bonferroni comparison results for feature selection specificity 52
20 ANOVA for interaction selection sensitivity 55
21 Tukey comparison factor combinations for interaction selection sensitivity 56
22 Tukey analysis p-value for interaction selection sensitivity 56
23 Tukey analysis 90% CI for interaction selection sensitivity 57
24 Comparison of interaction selection sensitivity means with the use of IPTW 58
25 Simultaneous Bonferroni comparison results for interaction selection sensitivity 58
26 ANOVA for interaction selection FDR 62
27 Tukey comparison factor combinations for interaction selection FDR 63
28 Tukey analysis p-value for interaction selection FDR 63
29 Tukey analysis 90% CI for interaction selection FDR 64
30 Comparison of interaction selection FDR means with the use of IPTW 65
31 Simultaneous Bonferroni comparison results for interaction selection FDR 65
32 Pain Outcome Measures 69

33 Treatment Decision Variables 70
34 Pain Management Research Summary 71
35 Description of state, treatment, and outcome variables in pain management dataset [27] 96
36 Adverse Treatment Interaction Constraints [27] 99
37 Stage 1 Pain Outcome Transition Models 2 SP AL-IPTW 99
38 Stage 2 Pain Outcome Transition Models 2 SP AL-IPTW 101
39 Stage 1 Pain Outcome Transition Models 2 SP IQ-IPTW 102
40 Stage 2 Pain Outcome Transition Models 2 SP IQ-IPTW 105
41 Stage 1 Pain Outcome Transition Models 2 SP IQ 108
42 Stage 2 Pain Outcome Transition Models 2 SP IQ 111

1 Introduction

Pain is a common problem that most people encounter during their lifetimes. Pain can be broadly classified into two types, acute and chronic. Acute pain usually occurs after an injury/accident and is treated with either analgesic medications or surgery followed by rehabilitation. Chronic pain on the other hand can occur due to multiple reasons like poor lifestyle, age, pre-existing health conditions, surgery, cancer, etc. Chronic pain is defined as pain that lasts more than 3 to 6 months [1]. Chronic pain is one of the main causes for adults seeking medical care [2]. Chronic pain that severely inhibits life or work activities is classified as high-impact chronic pain (HICP) [3, 4]. The Centers for Disease Control and Prevention (CDC) estimated that around 20.4% (50.0 million) of US adults have chronic pain, and 8.0% (19.6 million) of US adults have high-impact chronic pain based on the 2016 National Health Survey Interview [5]. A higher prevalence of chronic pain and HICP was reported among older adults and economically vulnerable adults [5].

Chronic pain is often characterized by unrelenting and debilitating symptoms and leads to unpleasant physical and mental experiences [6]. Pain is very subjective, where several factors, such as the patient's social and financial background, personal relations, and mental conditions, impact the perception and description of their conditions. Chronic pain not only affects the patient's quality of life, but it also affects family and social circles since relatives and caregivers must monitor the patient's medication, side effects, and state of mind. [7]. The earlier models of chronic pain were based on a biomedical model that focused only on the processes within the body and assumed independence between the body and mind. This model favored treating the condition rather than the patient, with medications being the preferred choice of treatment. The development of neuroscience lead to a marked shift in the way chronic pain was viewed, with equal focus being given to body and mind. To understand the impact of chronic pain on a patient's daily activities, mental and physical health, and social relations, the biopsychosocial model of pain (Figure 1) was developed, which provides a holistic framework for understanding how different aspects of pain are related through an assessment of sensorial, cognitive, and interpersonal factors [8, 9].

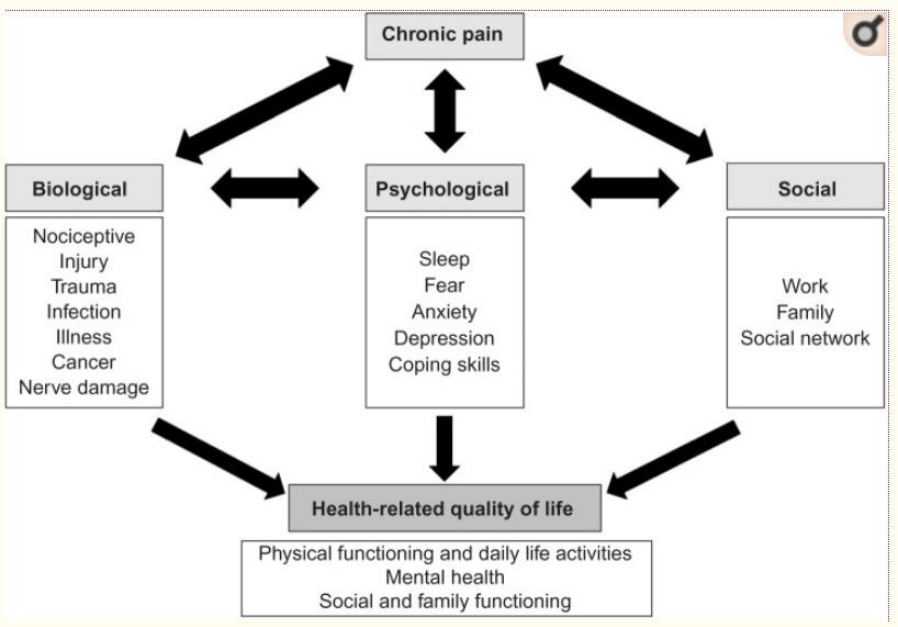


Figure 1: Biopsychosocial model of pain [8]

The typical treatment for chronic pain is analgesics, opioids, and other medications. Short-term medications help relieve severe pain in patients, but long-term usage can be detrimental to their health [10]. Deyo et al. [11] found that opioids were prescribed for over 60% of patients with non-cancer pain, and almost 20% became long-term users. Despite opioid therapy, a majority of the patients have persistently high levels of pain and poor quality of life [12]. Opioid medications also pose a significant risk of misuse and abuse [13].

Surgical interventions are another preferred treatment for chronic pain. Rajaee et al. [14] found that between 1998 and 2008, there was more than a 100% increase in spinal fusion surgeries for low back pain. There have been concerns about surgical interventions leading to high disability rates after the procedures [15].

The utility of the biopsychosocial model of pain along with the limitations of a medication-only treatment plan has led to the evolution of interdisciplinary treatment strategies. Interdisciplinary treatment includes physical therapy, psychotherapy cognitive behavioral therapy (CBT), and other procedural interventions along with medications to treat chronic pain. Interdisciplinary care consists of greater

coordination, communication among the different healthcare professionals, and active patient involvement to ensure effective and comprehensive treatment [16].

The Eugene McDermott Center for Pain Management at the University of Texas (UT) Southwestern Medical Center has one such interdisciplinary pain management program for chronic pain. The subjective nature makes it difficult to measure pain [17], but several outcome measures help quantify the patient's pain experience. The McDermott Center evaluates patients on five different pain outcome measures that evaluate patients' physical, psychological, and overall wellbeing. These pain outcomes are listed below:

- Oswestry Pain Disability Index (OSW) is a measure of functional disability due to pain [18].
- Pain Drawing Analogue (PDA) is an analogue scale of 0-10, with 0 corresponding to no pain and 10 corresponding to worst possible pain [19].
- Beck Depression Inventory (BDI) measures the severity of depression [20].
- Short Form Survey Physical Component Score (SF36pcs) and Short Form Survey Mental Component Score (SF36mcs), are general health status profile surveys designed to measure the physical and mental health status of the patient respectively [21].

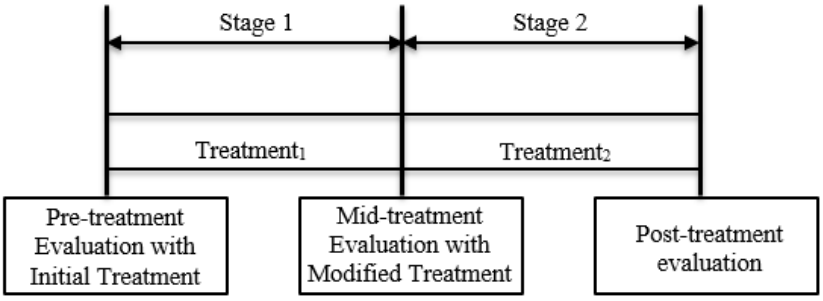


Figure 2: Two-stage interdisciplinary pain management program [22]

Lin et al [22] modeled this program as a two-stage adaptive treatment framework [23]. The patient entering the program undergoes a pre-treatment evaluation at the beginning of Stage 1, as shown in Figure 2. Based on the pre-treatment evaluation, an initial treatment plan is prescribed by the interdisciplinary team of experts. The individual patient's treatment plan depends on data including their demographic

information, past medical and surgical history, past treatments, and past pain outcome scores. These are called as patient *state* variables. Interactions between the state variables and treatments play a major role in determining a personalized treatment plan for individual patients. The initial treatment period (Stage 1) can last from a few weeks to months depending on the individual patient's characteristics and the severity of their pain. At the end of this period the patient is evaluated again, where their pain outcome scores are measured. Depending on this evaluation, the team of experts continues or modifies the treatment regimen, if appropriate. This is called the mid-treatment evaluation, where Stage 1 ends and the patient transitions into Stage 2 of their treatment plan. At the end of Stage 2, the patient undergoes a post-treatment evaluation, which concludes the two-stage program. The patient will go through another evaluation one year after the completion of the program [24].

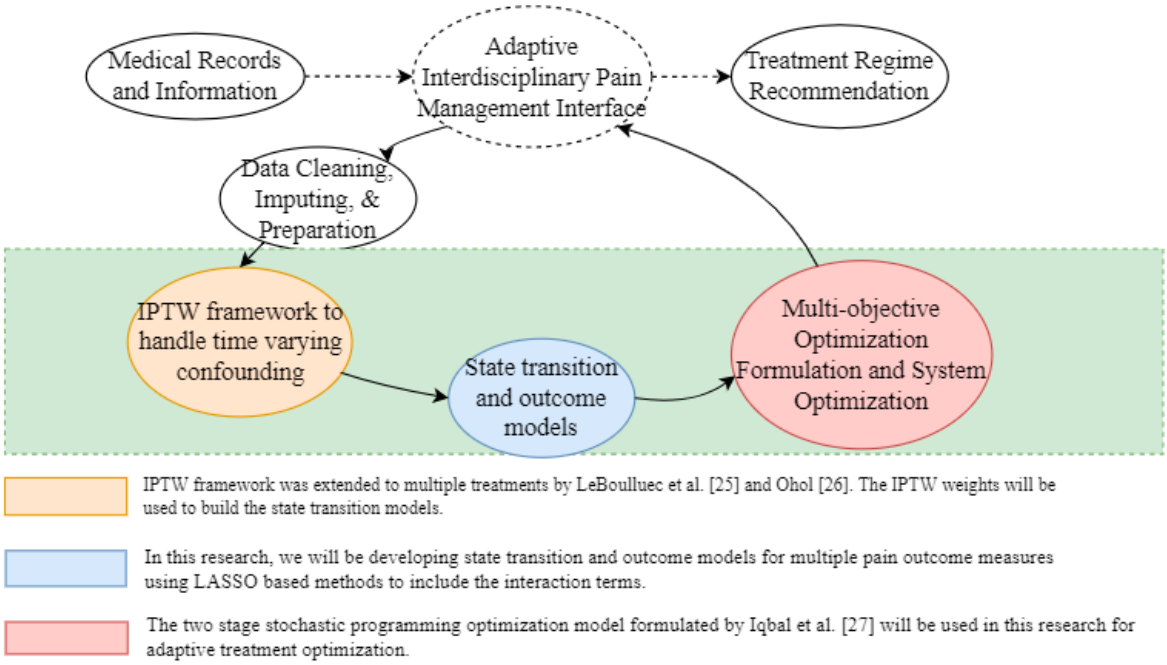


Figure 3: Adaptive interdisciplinary pain management research objectives [26]

The research in this dissertation completes the research objectives of a larger pain management project, shown in Figure 3. The development of an adaptive treatment decision tool is the main goal. Adaptive treatment strategies develop decision rules that depend on the patient's state [23]. In an adaptive treatment

environment, the patient's current state influences the treatments recommended, which in turn affect the patient's future states. When clinical expertise is employed to select the treatment, the observational data from this adaptive treatment process involves time varying confounding, for which it is not possible to separate whether it is the sequence of treatments that produced the patient's outcomes or if it is the evolution of the patient's state characteristics that led to these outcomes. Consequently, time varying confounding leads to biased estimates of the treatment effects on pain outcomes. There are several methods to address time-varying confounding. For the pain management project in Figure 3, LeBoulluec et al. [25] and Ohol [26] addressed time-varying confounding using the Inverse Probability of Treatment Weighting (IPTW) technique.

The adaptive treatment decision optimization problem has been studied in two form. Lin et al. [22] utilized a stochastic dynamic programming approach that conducts decision-making over multiple stages. Because the pain management program in Figure 2 only requires two stages, Wang et al. [28] and Iqbal et al. [27] formulated the decision optimization as a two-stage stochastic programming problem. The optimization in this dissertation is based on the formulation by Iqbal et al. [27] because they considered all five pain outcome measures. For both the optimization approaches, state transition models are used to represent the transition of patient state, treatment, and pain outcome variables from stage 1 to stage 2. While the prior work on the large pain management project did build state transition models, this dissertation specifically addresses the challenge of modeling state-treatment interactions that are critical to enable personalized treatment for individual patients. The methods proposed in this dissertation use a Least Absolute Shrinkage and Selection Operator (LASSO) [29] based technique named HierNet [30]. This state transition modeling approach also addresses time varying confounding by incorporating IPTW techniques developed by Ohol [26]. Finally, the state transition models built using the proposed method are then formulated within the multi-objective two-stage stochastic programming optimization to complete the objectives of the larger pain management project.

The outline of the dissertation is as follows. Chapter 2 provides a literature review on pain management and adaptive treatment strategies, multi-stage optimization and state transition modeling, feature, and interaction selection. Chapter 3 discusses the proposed LASSO based modeling approach to build state transition models with interactions. The performance of the proposed method on feature and interaction selection is evaluated on simulated pain management data. Chapter 4 discusses application of the proposed method to build state transition models with interactions on the pain management dataset. The state transition models are used in the optimization module, and the optimal treatment recommendations are evaluated.

2 Literature Review

2.1 Pain Management Program

Pain management programs are primarily aimed at treating and managing a patient's pain. Over the last few decades, significant advances have been made in our knowledge of basic pain mechanisms. Melzack and Wall's gate control theory, which proposed that pain perception is determined by various psychological factors in addition to sensory input, was a major driving force behind rapid developments in research on chronic pain [31]. The improvements in understanding the underlying causes of pain and acknowledging the influence of social, economic, and psychological factors on an individual's pain experience have led researchers and clinics to focus on holistic approaches that shift the focus from directly treating pain symptoms to improving patient quality of life [32, 33]. The Centers for Disease Control and Prevention (CDC) *Guideline for Prescribing Opioids for Chronic Pain* (2016) and the Department of Health and Human Services' *National Pain Strategy* (2016) have recommended the biopsychosocial approach in the treatment of chronic pain [30, 35]. The CDC guideline recommends the use of cognitive behavioral therapy, physical therapy, and non-opioid medications as the first line of treatment [34]. This has led to the development of interdisciplinary and multidisciplinary pain management programs, shifting the burden of a patient's pain management from primary care physicians to a specialized team of health care providers [36]. The American anesthesiologist John J. Bonica established one of the first multidisciplinary pain centers at the University of Washington in Seattle [36, 37].

The main difference between multidisciplinary and interdisciplinary pain management programs is the level of coordination between the different care providers. In multidisciplinary programs although care is provided by several health care providers, it may not be coordinated, resulting in parallel treatment plans and goals rather than an integrated approach [16, 36]. Interdisciplinary care involves greater coordination of services and frequent communication among health care professionals, where the different treatment plans complement each other [16]. The patient and their caregivers are active participants in this program

[16]. Interdisciplinary pain management programs have a patient-centered approach, with a focus on patient education and cognitive behavioral changes [38]. Short- and long-term treatment goals are discussed and reviewed regularly in line with the expectations of the patient, family members, and clinicians [37].

Several studies have shown the effectiveness of multidisciplinary/interdisciplinary programs in managing pain and improving patient outcomes. Flor et al. [39] conducted a meta-analytic review of sixty-five studies on multidisciplinary treatments for chronic back pain and found that multidisciplinary treatments for chronic pain were superior to other treatment plans. Turk [40] found that pain rehabilitation programs provide a comparable reduction in pain outcomes to alternative treatment modalities, but significantly improve medication use, health care utilization, functional activities, and return to work among patients, and are more cost-effective than surgical interventions. Gagnon et al. [41] analyzed the efficacy of an interdisciplinary pain management program on workers' compensation patients with chronic back pain and found a significant decrease in patients' emotional stress, pain intensity, and an increase in their return-to-work status at program completion. The integration of physiotherapy and clinical psychology in pain management and its effectiveness has been reviewed by Johnson and Morales [42]. The success of interdisciplinary pain management programs has led to their widespread use in many clinics [16, 37].

The interdisciplinary pain management program follows an adaptive treatment strategy. Adaptive treatments can be described as a set of sequential decision rules adapted based on the patient's characteristics and response to treatments over multiple stages [23]. The research on adaptive treatments is divided into two categories: randomized experimentation and Markov decision processes [22]. Randomized experimentation includes the multiphase optimization strategy (MOST) and sequential multiple assignment randomized trials (SMART) [43]. The adaptive treatment decision problem is formulated as a stochastic dynamic program (SDP), which is discussed in the next section. The adaptive treatment decision framework developed for the pain management program is discussed in Chapter 4.

2.2 Multi-Stage Optimization

Multi-stage optimization involves problems where decisions must be made sequentially in stages under conditions of uncertainty. Multi-stage optimization has applications in energy [44, 45], finance [46], supply chain [47, 48, 49], manufacturing [50], healthcare [51] and other fields. For the adaptive interdisciplinary pain management project in Figure 3 from the previous chapter, two types of multi-stage optimization have been employed. Lin et al. [22] used stochastic dynamic programming (SDP), which can handle two or more stages. Because the interdisciplinary pain management program at the McDermott Center can be modeled as a two-stage program (Figure 2), Wang et al. [28] and Iqbal et al. [27] used two-stage stochastic programming. Here, we discuss the adaptive interdisciplinary pain management framework of Lin et al. [22] using SDP since this was the first optimization framework for this problem. Later, in Chapter 4, we build on the work of Wang et al. [28] and Iqbal et al. [27] to construct the two-stage stochastic programming formulation employed for the research in this dissertation.

Dynamic programming is a collection of mathematical tools to analyze and solve sequential decision problems and was first introduced by Bellman [52, 53]. These problems are most commonly modeled in discrete (time) stages, where a decision is made in each stage, then additional information is observed, followed by a subsequent decision in the next stage, and so on [54]. SDP models multi-stage optimization problems under conditions of uncertainty where some of the parameters in the problem are modeled as stochastic variables. There are five main elements in a SDP formulation: stages *T*, are the sequential time stages when decisions are made; state variables $S_t \in R^n$, represents the state of the system at time t before we make a decision; decision variables $u_t \in R^m$, are controlled to optimize the solution; the feasible decision space is denoted by X_t, $u_t \in X_t$, state transition functions $f_t(\cdot)$, model the transition of state variables from stage t to $t+1$; the objective function $C_t(\cdot)$: $R^{n+m+l} \to R^l$, could be a cost/reward/utility function which may depend on the state S_t and the decision u_t [55].

An SDP formulation over T discrete time stages is shown in Equation (1) [56, 57]

$$\min_{u_1,u_2} E\{\sum_{t=1}^{2} C_t(s_t, u_t, \varepsilon_t)\} \quad (1)$$

$$\text{subject to } s_{t+1} = f_t(s_t, u_t, \varepsilon_t) \quad for\ t = 1, \dots\dots, T-1,$$

$$(s_t, u_t) \in \Gamma_t \quad for\ t = 1, \dots\dots, T,$$

$$u_t \in X_t \quad for\ t = 1, \dots\dots, T,$$

where $\varepsilon_t \in R^l$ is the random vector, $\Gamma_t \in R^{n+m}$ is the set of constraints and X_t is the feasible decision space.

The pain management treatment optimization is a two-stage stochastic programming problem, a common type of multi-stage optimization problems. Two-stage stochastic programming problems can be found in healthcare [58, 59], staffing and scheduling [60, 61], energy [62] and other applications. The state, decision, and outcome variables in the pain management research are shown in Table 1. The state and decision variables are a mix of continuous and categorical variables, while the outcome variables are continuous.

Table 1: Pain management variables

State Variables	Patient demographics
	Patient medical history
	Patient surgical history
	Patient treatment history
	Patient past pain outcome measures
Decision Variables	Pharmaceutical treatments
	Procedural treatments
Outcome Variables	Pain outcome measures: OSW, PDA, BDI, SF36-mcs, and SF36-mcs

The two-stage stochastic programming formulation is shown in Equation (2) [63, 64]

$$\min_{x} c^T x + E[Q(x,w)] \tag{2}$$

$$s.t. Ax = b$$

$$x \in R_+^{n_1} \times Z_+^{s_1}$$

$$where\ \ Q(x,w) := \min_{y} q^T y$$

$$s.t. Wy = h - Tx$$

$$y \in R_+^{n_2} \times Z_+^{s_2},$$

where x and y are the stage 1 and stage 2 decision variables, n_1, s_1 are the number of continuous and integer variables in stage 1 decision vector and n_2, s_2 are the number of continuous and integer variables in stage 2 decision vector respectively. The second stage data is uncertain and is represented by $w = (q, T, W, h)$. The objective is to minimize the first stage decision cost and the expected second stage decision cost. c is the first stage decision cost vector and $Q(x,w)$ is the second stage decision cost function. The uncertain data w can be modeled as a probability function to represent a discrete finite number of scenarios $w_1, \dots, w_k$ with probabilities $p_1, \dots, p_k$ respectively. The expected second stage cost function is the summation of the second stage cost function $Q(x,w)$ over k scenarios. The two-stage problem can then be formulated as one mixed integer linear programming (MILP) model, shown in Equation 3 [63, 64]

$$\min_{x, y_1, \dots y_k} c^T x + \sum_{i=1}^{k} p_k q_k^T y_k \tag{3}$$

$$s.t. Ax = b$$

$$W_k y_k = h_k - T_k x \qquad k = 1, \dots, k$$

$$x \in R_+^{n_1} \times Z_+^{s_1}, y \in R_+^{n_2} \times Z_+^{s_2}$$

The second stage scenario generation is an important topic of research since there is a tradeoff between computation time and solution quality. The more scenarios we generate, the better the solutions but the computational time increases. A survey of different sample generation methods can be found in [65, 66].

The equivalent MILP problem will be solved after generating the scenarios, assuming that the cost functions and state transition constraints are linear. The state transition constraints correspond to the transition functions that map the state, decision, and outcome variables from stage 1 to stage 2.

Lin et al. [22] used approximate dynamic program (ADP) to solve the pain management optimization problem and used linear regression to model the state transitions. Wang et al. [28] used the two-stage stochastic programming approach with stepwise regression used to build the state transition models. Since the regression models had interaction terms, the state transition functions were quadratic and non-convex. Linearization techniques using piecewise linear functions was applied to make these constraints linear and formulate the optimization as an MILP problem. The objective of the optimization was to minimize the pain outcome measure, OSW while also penalizing excessive treatment costs under the state transition function and treatment interaction constraints. The MILP solutions were compared with the solutions obtained by solving the original MINLP problem without linearizing the interaction constraints. Iqbal et al.[27] developed a multi-objective two-stage stochastic programming optimization approach using piecewise linear networks (PLN) to build the state transition functions. The objective was to minimize the multiple pain outcome measures along with treatment costs. The optimization problem was solved as an MILP problem.

Markov decision process (MDP) is another technique to formulate sequential decision-making problems in discrete time steps. MDP is characterized by a finite and discrete state space, with the state transitions modeled as probability matrices. MDPs are based on the Markov property that future state transitions depend only on the current state, and decisions are independent of past states and past decisions. MDPs have been used in medical decision problems to determine the optimal timing of interventions [67, 68, 69]. Alagoz et al. [70] model an infinite horizon stationary MDP to determine the optimal timing of liver

transplantation. Shechter et al. [71] use MDP for optimal initiation of HIV treatment. Denton et al. [72] use MDP to optimize the start time of statin therapy in diabetes patients. The main advantages of MDP are that it allows for a simpler representation of the future states and possible transitions that may occur, and it is preferred over decision trees for complex problems [69].

MDP models the state transition at the cohort level, while microsimulation (MSM) models the state transition at an individual level [73, 74]. MSM simulates events and outcomes at the individual level to provide information that can guide policy decisions [75]. The state transitions are modeled as individual probability matrices. MSM models are not limited by the Markovian assumptions since they simulate one individual at a time [76]. MSM has been used in cancer research [77], diabetes research [78], and health policy [79]. These models are computationally intensive and often require simulating millions of individuals to obtain stable outcome values [76]. MDP and MSM are structured around a set of mutually exclusive and exhaustive states [76], which can create a computationally intractable problem.

MDP and MSM are suitable for discrete state space problems, while the pain management state space has a mix of continuous and discrete variables. Discretization techniques can be employed in the continuous state space, but this is not desired to avoid potential information loss. The state transition probability matrices also do not capture the interactions between the state and decision variables.

In this research, optimization framework is based on the multi-objective two stage stochastic programming approach of Iqbal et al.[27] since all the pain outcome measures were considered in that study. An important aspect of two-stage stochastic programs is to model the state transition functions, and this is discussed in the next sections.

2.3 State Transition Models

In any stochastic programming formulation, modeling the state and outcome transitions is an important step. Since the pain management data are from an observational study, we look into the literature on longitudinal data analysis for methods to model transition functions. Longitudinal data are data resulting

from observing subjects repeatedly over time [80]. They allow the researcher to track the changes in the response variable over time. Correlation between variables is common in longitudinal data due to the repeated measurements over time, and multiple treatments being prescribed together depending on the patient's condition. The data structure includes baseline data (e.g., age, gender, race, etc.) and time-varying signals (e.g., treatments, medical conditions, outcome measures, etc.) and includes both categorical and continuous variables [81].

Random effects models have been used in longitudinal data analysis to model individual-specific random effects on the outcome variable in the time-varying setting [82]. The linear mixed effect model (LME) is one such regression-based random effects model for continuous outcome response. It is based on the premise that there is a subject-specific mean response profile over time, with a specific functional form [80]. The general LME model form for outcome $Y_{i(t+1)}$ measured subsequent to the t-th time point for $i = 1, \dots, n$ individuals have the form [80, 82, 83] shown in Equation (4)

$$Y_{i(t+1)} = {X_{it}}^T \beta + {Z_{it}}^T b_i + \varepsilon_{i(t+1)}, \tag{4}$$

where X_{it} and Z_{it} are the known design matrices prior to the t-th time point for the fixed-effects and random-effects coefficients respectively. β is the vector of fixed-effects coefficients, $b_i \sim N(0, D)$ is the vector of random effects coefficients, and $\varepsilon_{i(t+1)} \sim N(0, \sigma^2)$ is the random error. The random effects are assumed to be independent of the error terms $\varepsilon_{i(t+1)}$ and normally distributed, with mean zero and variance-covariance matrix *D*. Maximum Likelihood principles are used in estimating the model parameters.

Generalized estimating equations (GEE) is another regression-based approach for longitudinal data analysis [84]. In this approach, two models are specified. The first is the regression model for mean response and the second is a correlation model for the within-subject correlation. The purpose of the correlation model is to apply the covariance inverse weights to the observations and obtain regression coefficient estimates and the standard error for the estimated coefficients [85]. The drawback is that it requires large sample sizes to obtain unbiased and consistent estimation [86].

The LME and GEE fall under a category of longitudinal study called parametric models. Kvaløy et al [87] and Nordseth et al [88] use a non-parametric regression method to model cardiac arrest data including the state history. Due to the non-parametric nature, this method can handle data under minimal assumptions. A detailed review of non-parametric and semi-parametric modeling methods can be found in Huang [89].

Hidden Markov Models [90, 91] and Gaussian Process State Space Models [92] are other approaches used in modeling observational health data. These methods involve complex model specifications, involve a large number of parameters, and require training [81].

The methods discussed under longitudinal data analysis are designed for studies where the covariates are assumed to be fixed. They do not model the effects of past treatments and past outcomes on current treatments and outcomes [93]. This leads to inferential challenges when one tries to apply this model with time-varying confounding variables [82]. These limitations make them unsuitable for building state transition models on the pain management dataset.

A review of best modeling practices for state transition modeling can be found in the ISPOR-SMDM report [76], which states that while modeling the effectiveness of treatment interventions in observational studies, it is important to account for time-varying confounding. The sequential nature of the pain management program results in the treatment effects in a particular stage being influenced by patient state variables, past treatments, and past pain outcome measures. This is referred to as time-varying confounding. There are several methods to handle time-varying confounding and identify the true treatment effects. The most studied methods are propensity scores [94, 95, 96], marginal structural models [97, 98, 99, 100], g-estimation [101, 102, 103, 104, 98] and IPTW weights [105, 106, 107, 108, 109, 110]. A detailed review of the different methods to handle time-varying confounding can be found in the work of LeBoulluec et al. [25] and Ohol [26]. Robins et al. [106] introduced Inverse Probability of Treatment Weighting (IPTW) to obtain unbiased estimates of treatment effects in a one-treatment setting. LeBoulluec et al.[25] extended the IPTW framework to multiple treatments. Ohol [26] considered the case with multiple correlated treatments and used MIMIC [111] to estimate the joint probability distribution of treatments.

It is important to perform feature selection on the state space to identify the relevant features to be used in the state transition function. A review of the feature selection methods is given in the next section.

2.4 Feature and Interaction Selection

A feature is an individual measurable property of the system being observed [112]. With the development of technology, data storage capacity, and computing systems, the number of features observed, and the data collected has increased exponentially [113]. The analysis of these data to draw meaningful insights about the system processes, and to build models is an important goal for researchers. Data pre-processing is the first step in the model-building process, where the data are processed before being presented to any learning, discovering, or visualizing algorithm [114]. There are three steps in data pre-processing: feature construction, feature extraction, and feature selection. Feature construction is the process where missing information about the relationship between features is discovered and the feature space is augmented by inferring or creating additional features [115, 116]. Feature extraction is a mapping process from the original feature space to a lower dimensional one [117]. Feature selection is the process of selecting a subset of features from the original input feature space and evaluating this subset on the research objective like response prediction/classification or uncovering the cause-effect relationship between features and response [117].

Feature selection methods are further classified into the filter, wrapper, and embedded methods [112]. Filter methods rely on the characteristics of the training data to rank and select features based on certain relevance criteria [118] and are independent of the model-building process. There are different relevance criteria used to rank features [119, 120], based on statistical measures, such as Pearson's correlation [121], Linear Discriminant Analysis, ANOVA, Chi-square [122], Wilcoxon Mann Whitney test [123], and Mutual Information [124, 125]. The faster computation time is the main advantage of filter methods and is preferred when the input feature space is large. Some of the ranking methods do not consider the correlation between the features, which leads to the selection of a redundant subset [120]. It is hard to select a suitable modeling

algorithm with filter methods since they do not consider the performance of the algorithm with the selected feature subset [126]. In feature ranking, important features that are less informative on their own but are informative when combined with others could be discarded [112]. Filter-based methods have been used in medical imaging [127, 128], DNA microarray data [129, 130] and signal processing [131, 132].

Wrapper methods search for an optimal feature subset along with the performance of the predictive algorithm on the feature subset [133]. The predictive algorithm is modeled as a black box and repeatedly runs on the dataset using various feature subsets. The feature subset with the best prediction is selected. There are broadly two search algorithms to search for the feature subset: Sequential Selection Algorithm and Heuristic Search Algorithms. The sequential selection algorithm starts with an empty set (full set) and adds features (removes features) until the optimal objective function is obtained [112]. The heuristic algorithm generates subsets around the search space by generating solutions to the optimization problem [112]. Wrapper methods have better predictive performance than the filter-based methods since the feature selection and prediction model building go hand in hand. The downside of a wrapper method is that it is computationally intensive, and the search space grows exponentially as the number of features increases [133]. Wrapper-based methods have been used in brain MRI studies [134, 135] and medical image processing [136, 137].

In embedded methods, the feature selection is embedded in the modeling algorithm [138]. Embedded methods include decision tree algorithms like CART [139] and Random Forest [140]. LASSO [29] is an embedded method that performs feature selection based on L_1 regularization and is computationally fast.

2.4.1 LASSO

LASSO is a regularization technique for simultaneous estimation and feature selection [29]. The LASSO estimates are shown in Equation (5)

$$\hat{\beta} \in \arg\min_{\beta} \left\| y - \sum_{j=1}^{p} x_j \beta_j \right\|^2 + \lambda \sum_{j=1}^{p} |\beta_j|, \tag{5}$$

where y is the response vector, β_j is the coefficient estimate of input feature x_j, p is the number of input features, and λ is the regularization parameter. The second term is called the "l_1 penalty", which shrinks the coefficients to zero as λ increases. The LASSO estimates could be inconsistent since they apply the same shrinkage for all the variables estimates. They achieve consistent variable selection and optimal estimation, which is referred to as the oracle property, only under certain necessary conditions [141, 142, 143].

Adaptive LASSO was developed by using an adaptively weighted l_1 penalty [143] and shown to satisfy the oracle property more likely than LASSO. The Adaptive LASSO estimates are calculated using Equation (6).

$$\hat{\beta} \in \arg\min_{\beta} \left\| y - \sum_{j=1}^{p} x_j \beta_j \right\|^2 + \lambda \sum_{j=1}^{p} w_j |\beta_j|, \tag{6}$$

where $w_j = \frac{1}{|\hat{\beta}|^{\gamma}}$, $\gamma > 0$ and $\hat{\beta}$ can be Ordinary Least Squares estimates.

The LASSO algorithm shown in Equation (3) has been modified over the years leading to the development of several algorithms. The most prominent among these are Elastic Net [144], Group LASSO [145], Adaptive Group LASSO [146], and overlapped Group LASSO [147]. Elastic Net was developed to perform consistent feature selection under multicollinearity. Group LASSO was an extension of LASSO to select groups of variables together, for example in the multifactor analysis of variance models. Group LASSO was shown to have the same variable selection inconsistency as LASSO, and Adaptive Group LASSO was proposed, where different groups of variables were weighted differently, similar to Adaptive LASSO. The overlapped Group LASSO considers groups of features, allowing overlap between the groups.

LASSO has been widely used in high-dimensional data modeling where parsimony is desired. Zhang et al. [148] and Wang et al. [149] use group lasso regularization to perform feature selection for neural networks. They are used in the estimation of sparse graphical models and neighborhood selection, which find applications in molecular biology, gene expression, and network analysis [150, 151]. Farahani et al.

[152] demonstrated the application of LASSO based method for causal variable selection on the pain management data while addressing time-varying confounding.

2.4.2 Interaction Selection

The feature selection methods discussed so far focus mostly on the main effects, while in practice it is important to consider the influence of interactions between the input variables on the response variable. Interaction models are useful in social, political, economic, epidemiology, and genetic studies and provide better insight into the association between variables [153, 154]. Gunter et al. [155] discuss the importance of selecting interaction variables in optimal decision-making problems.

Interaction selection/screening is a major subject of study with different methods developed to discover and model interactions. In statistical modeling of interactions, it is a common practice to allow interaction in the model only if the corresponding main effects are present in the model [156, 157, 158]. This is referred to as heredity, marginality, or strong hierarchy. Weak hierarchy is when an interaction is allowed in the model when at least one of the corresponding main effects is present in the model.

Strong Hierarchy: $\widehat{\Theta}_{ij} \neq 0, \Rightarrow \hat{\beta}_i \neq 0 \; and \; \hat{\beta}_j \neq 0,$

Weak Hierarchy: $\widehat{\Theta}_{ij} \neq 0, \Rightarrow \hat{\beta}_i \neq 0 \; or \; \hat{\beta}_j \neq 0,$

where $\widehat{\Theta}_{ij}$ is the interaction estimate and $\widehat{\beta}_i, \hat{\beta}_j$ are main effect estimates for variables i and j respectively.

Several methods aim at building models while satisfying this hierarchy constraint. The multi-step approach is one such iterative process built on the stepwise framework, where variables are added or removed iteratively. The stepwise framework screens the main effects first and then searches for interactions between the selected main effects, thus enforcing a strong hierarchy [159, 160]. There are several optimization-based approaches, which formulate the hierarchy constraint as convex and non-convex. Zhao et al. [161] introduced composite absolute penalties (CAP), a broad class of penalties that can achieve group and hierarchical sparsity. Choi et al. [162] formulated a non-convex optimization for

sparse hierarchical interaction models. Bien et al. [30] developed a method called HierNet, where they introduced a set of convex constraints to LASSO to satisfy the hierarchy condition and build sparse interaction models. Lim and Hastie [163] developed a method called Glinternet, where they formulated a constrained overlapped group LASSO to enforce hierarchy and solve it using an equivalent group LASSO formulation. Chipman [164] used the Bayesian viewpoint to build hierarchical interaction models adapting the stochastic search variable selection approach (SSVS) of George and McCulloch [165]. The multi-step procedures have computational advantages when the data dimension is large, but there have been questions raised over their theoretical validity [153]. Some of the optimization-based approaches have been shown to produce consistent estimates under the strong hierarchy condition [162, 166], but the major limitation of these methods is the computational cost and memory requirement with large datasets [153]. The interaction screening methods have been studied on genome-wide association studies, UCI spam base data, gene expression dataset, and Boston Housing dataset [153, 163, 167].

2.5 Contribution

The main objective of pain management research is to develop an adaptive decision framework to identify an optimal treatment regime. The different components that make up this framework are shown in Figure 4. The highlighted section is the focus of this research work.

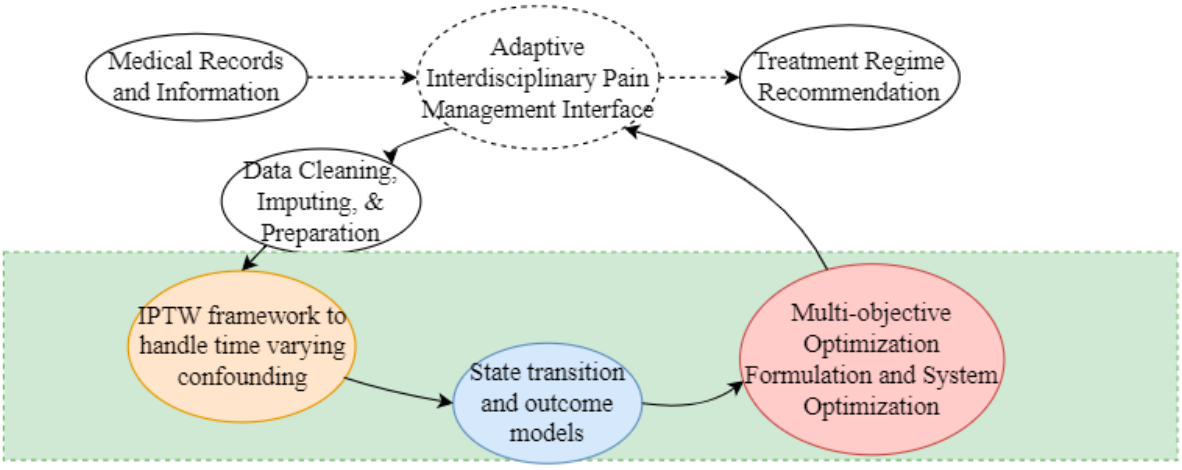
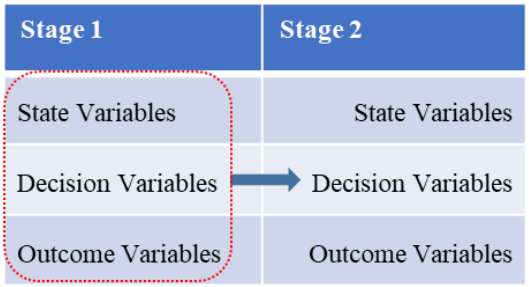


Figure 4: Adaptive interdisciplinary pain management framework [26]

The pain management decision problem is formulated as a two-stage stochastic programming problem with an adaptive treatment framework. The data has state, decision, and outcome variables for the two stages. The state and outcome transition functions model the transition of these variables between the stages, as shown in Table 2. The interaction between the state and decision variables has a significant impact on the outcome variable in an adaptive treatment setting since the treatments are prescribed depending on the patient's state [155]. It is therefore important to model these interactions in the transition models.

Table 2: State and outcome Transition [25]



The summary of the research work done on the pain management program is shown in Table 3. The proposed research work is highlighted in the table.

Table 3: Pain Management Research Summary

	State Transition Modeling				
Paper	**Model type**	**Interactions**	**Feature selection**	**IPTW**	**Optimization**
Lin et al. [22]	Linear model	State treatment interactions	Stepwise least squares	No	SDP
LeBoulluec et al. [25]	Linear model	State treatment interactions	Stepwise least squares	Derived weights	No
Ohol [26]	Linear model	No	No	Derived weights	No
Farahani et al. [152]	Linear model	No	LASSO regularization	Used Ohol weights	No
Wang et al. [28]	Linear model	State treatment interactions	Stepwise least squares	Used LeBoulluec's weights	Two-stage SP
Iqbal et al. [27]	Piecewise Linear Network	Network interactions	Piecewise Linear Network	Used LeBoulluec's weights	Two-stage SP
Viswanatha	Linear model	State treatment interactions	LASSO regularization	Derived weights	Two-stage SP

Lin et al. [22] modeled an optimization framework based on approximate dynamic programming (ADP) with linear regression to model the state transitions. The recommended treatment regime minimized adverse patient pain outcomes along with treatment costs. The adaptive and sequential nature of the pain management program introduces time-varying confounding, where the treatment effects are confounded by past treatments and patient state variables. LeBoulluec et al. [25] extended the IPTW framework to address this time-varying confounding in a multiple treatment setting. Ohol [26] further extended the IPTW method to consider correlated treatments. Farahani et al. [152] used LASSO-based regularization along with IPTW to perform causal feature selection.

The stochastic dynamic programming approach of Lin et al. [22] optimizes treatments over multiple stages. Since the pain management program has two-stages, Wang et al. [28] and Iqbal et.al [27] formulated the optimization as a two-stage stochastic programming problem. Wang et al. [28] used weighted least squares method to develop the state transition models using the IPTW weights from LeBoulluec et al. [25]. The state transition constraints in the optimization were non-convex as they modeled the state treatment interaction terms. A linearization technique using piecewise linear function was proposed to approximate the non-convex constraints and formulate the optimization problem as an approximated mixed integer linear problem (MILP). The objective of the optimization was to minimize the pain outcome measure, OSW while also penalizing excessive treatment costs under the state transition function and treatment interaction constraints. The optimization results from the approximate MILP were compared with the solutions obtained from the original mixed integer nonlinear problem (MINLP) formulation without linearizing the interaction constraints.

Iqbal et al.[27] developed a multi-objective two-stage stochastic programming optimization approach using piecewise linear networks (PLN) to build the state transition functions. The objective is to minimize all the pain outcome measures considered in this study along with treatment costs. A survey among caregivers was conducted to identify the relation between the different pain outcome measures and a convex

quadratic programming approach was used to obtain weights to penalize the different pain measures. An equivalent MILP model was used to solve the optimization problem.

Iqbal et al. [27] was the only work that considered multi-objective optimization and included IPTW weights while building the state transition models. The state transition models did not include the state treatment interaction effects but had network interactions. In this research work, our primary research goal is to model the state treatment interactions while building the state transition models. The second research goal is to use these state transition models in the multi-objective optimization framework based on Iqbal et.al [27] and study the treatment recommendation patterns from the proposed approach.

With this background, it is important to select a state transition modeling approach that identifies the right features from the state space. The work of Farahani et al. [152] showed that LASSO-based techniques can be used to build models on pain management data while achieving consistent feature selection. We propose to use HierNet [30], a LASSO-based interaction modeling approach while incorporating the IPTW-based weighting technique developed by Ohol [26] to effectively model the state transition functions with state and treatment interaction effects.

We evaluate the performance of the proposed HierNet-IPTW method on feature and interaction selection metrics in a case study designed to simulate the pain management data. This is discussed in Chapter 3. In Chapter 4, we use the proposed HierNet-IPTW modeling method to develop state transition models with state treatment interactions on pain management data. We build on the multi-objective stochastic programming optimization developed by Iqbal et al. [27], using the newly developed state transition models. The state transition functions are non-convex as they include interaction terms. Instead of linearizing the interaction terms and using an approximate MILP model, we use a Mixed Integer Quadratically Constrained Program (MIQCP) to formulate and solve the optimization problem. Optimal treatment recommendations are then compared with solutions obtained from state transition models without interactions and models without IPTW. This will help us understand the impact of including interaction effects and IPTW on optimal treatment recommendations.

Chapter 3

LASSO Based State Transition Modeling with Interactions in Pain Management Simulation Case Study

Abstract

The Eugene McDermott Center for Pain Management at the University of Texas Southwestern Medical Center has an interdisciplinary pain management program for chronic pain. This program is modeled as a two-stage adaptive treatment decision problem, with state transition models representing the transition of patient state, treatment, and outcome variables from stage 1 to stage 2.

In an adaptive treatment environment, the patient's current state influences the treatments recommended, which in turn affect the patient's future states leading to time-varying confounding and biased treatment estimates. Inverse Probability of Treatment Weighting (IPTW) [106] is one technique to address time-varying confounding. Interactions between the patient state and treatments also play a major role in determining a personalized treatment plan for individual patients. It is important for the modeling method to identify the causal features and interactions. The advantage of LASSO based methods is in building parsimonious models and handling correlation between variables, which is not the case with least squares-based methods.

In this paper, LASSO based method named HierNet [30] is combined with IPTW from Ohol [26] to build state transition and outcome models that enable feature selection and modeling of interaction effects in the presence of time-varying confounding. The proposed approach is studied using a simulated case study structured based on the McDermott Center data. The proposed approach is compared on the feature and interaction selection metrics against the baseline method that does not use IPTW.

3.1 Introduction

The objective of adaptive interdisciplinary pain management is to provide personalized pain treatment regime that considers non-pharmaceutical procedures as alternatives to medications. Given the wide range of potential treatment regimes, an adaptive treatment framework is needed to identify the optimal treatment regime. Of particular importance in personalizing treatment is the modeling of interactions between possible treatments and a patient's pain and health characteristics [155]. Interactions capture how the different treatments affect the pain outcomes for patients with differing characteristics. In this paper, we build on past research to complete the pain management adaptive treatment framework illustrated in Figure 5. The highlighted section in the figure shows the three modules that form the scope of the presented research.

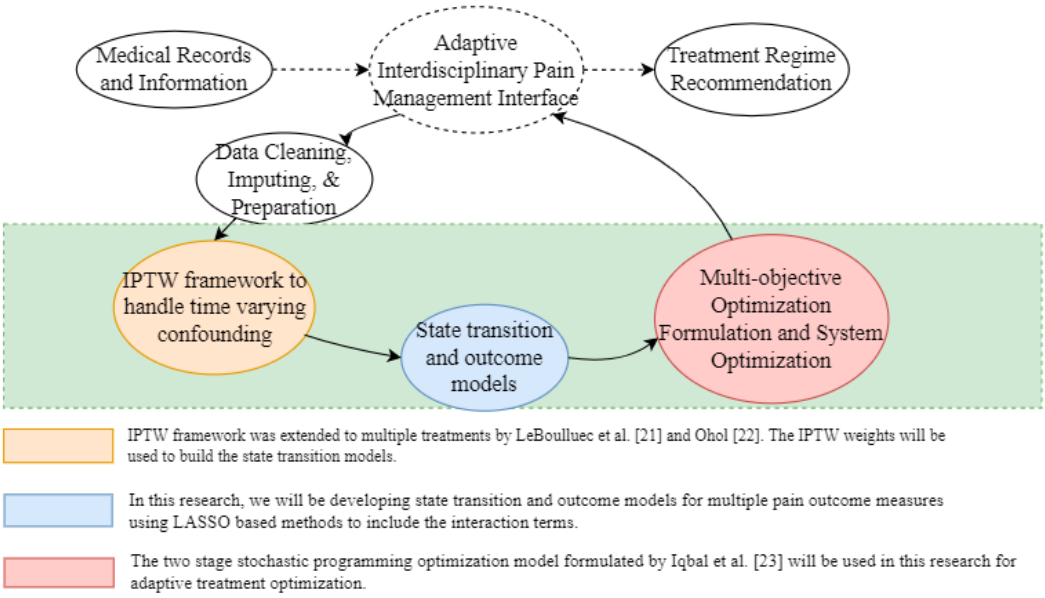


Figure 5: Adaptive interdisciplinary pain management research objectives [26]

Medical records and patient information for our study were provided by the Eugene McDermott Center for Pain Management at the University of Texas Southwestern Medical Center at Dallas. The raw dataset included patient information, past medication information, past medical records, treatments, and patient pain outcome measures collected over the course of the program. Data cleaning and imputation via

regression models was done to preserve as much of the data set as possible and to prepare the data in a format suitable for mathematical analysis. Please refer Lin et al. [22] for details on data preparation. Prior work cited in Figure 5 provides background specifying the state variables that define a patient's pain and health history, decision variables that define the categories of treatments, and pain outcome measures that are optimized via the adaptive treatment framework. State variables provide information about the patient and include past medical and surgical history, physical condition, age, gender, past medications etc. Decision variables are the different pharmaceutical and procedural treatments, such as analgesics, anti-depressants, physical therapy, and cognitive behavioral therapy. Outcome measures are OSW [18], PDA [19], BDI [20], sf36pcs [21] and sf36mcs [21]. The pain management program is a two-stage decision problem, with treatment decisions made in stages 1 and 2 respectively. Patient pain outcome measures are also collected at stage 1 and stage 2. The final cleaned dataset consisted of 294 patient observations, 62 state variables, 14 stage 1 decision variables, 13 stage 2 decision variables, and 5 stage 1 and stage 2 outcome variables respectively.

3.1.1 Addressing Time-varying Confounding

A primary challenge addressed in the work of LeBoulluec et al. [25] and Ohol [26] is conducting unbiased estimation of treatment effects using observational data. In order to optimize a treatment regime, it is essential to appropriately represent treatment effects within the decision optimization method. Ideally, a randomized controlled trial would be implemented to provide this information, but the availability of observed clinical data raises the opportunity to leverage such data. Since certain treatments are prescribed together and certain medical conditions are treated by specific treatments, we further observe correlation between some of the state variables, and between some of the state and decision variables.

The pain management program is adaptive and sequential, where treatment effects in a particular stage are influenced by past patient state variables, past treatments, and past pain outcome measures. This is referred to as time-varying confounding or endogeneity [106]. This leads to biased estimation of the causal effects of the treatments on the pain outcome measures. Consider this one treatment example shown in

Figure 6. The treatment T_1 is prescribed in stage 1 and patient outcome measure Y_1 is measured at the end of stage 1. The treatment T_2 is prescribed in stage 2 and patient outcome measure Y_2 is measured at the end of stage 2. The red causal path shows the time-varying confounding effect of stage 1 patient outcome Y_1 and stage 1 treatment T_1 on the prescribed treatment in stage 2, T_2. The effect of treatment T_2 on patient outcome Y_2 is confounded by the past patient outcome and past treatments. The orange causal path shows the effect of patient state variables like age, gender, race, past medical history, etc. on treatments and patient outcomes. These are considered confounders since they are associated both with the treatments and outcomes. To find the impact of treatment on outcome measures, the effect of these confounders needs to be accounted for. In this study, the confounders are time-invariant since patient state variables are assumed to be constant throughout the treatment plan. The past treatments and past outcomes are time-varying confounders since they vary from one stage to another.

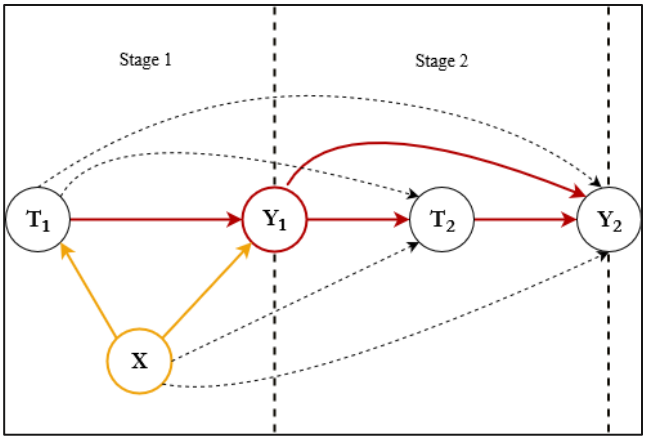


Figure 6: Time-varying confounding in one treatment case [26, 168]

There are several methods to address time-varying confounding like G-estimation [102, 103, 104], marginal structure models (MSM) [97, 98, 99, 100], G-computation [98, 101], inverse probability of treatment weighting (IPTW) [105, 106, 107, 108, 109, 110]. The IPTW approach is the preferred method for handling time varying confounding in an adaptive treatment setting. IPTW generates weights that can be conceptually thought of as creating replicates of the rare observations in the study with the goal of mimicking data that would be expected from clinical trials. The important assumption while calculating the

weights using this method is that there are no unmeasured confounders, which implies all confounders are accounted for in the model.

LeBoulluec et al. [25] extended the Inverse Probability of Treatment Weighting (IPTW) method to address time-varying confounding in the multiple treatment setting assuming treatment independence. Ohol [26] extended the IPTW method further by considering multiple correlated treatments and used the MIMIC [111] algorithm to estimate the joint probability distribution of treatments. Given the correlated nature of treatments in interdisciplinary treatment, the method of Ohol [26] is employed in the current work.

3.1.2 Optimization Approaches for Adaptive Interdisciplinary Pain Management

Referring back to three modules highlighted Figure 1, IPTW weights are first generated from the available observational data, then used to build unbiased state transition and outcome models which are finally incorporated into the optimization of the adaptive treatment framework. In addition to the IPTW work of LeBoulluec et al.[25] and Ohol [26], the current research builds on prior work for optimization. Lin et al. [22] created an optimization framework based on approximate dynamic programming with linear regression to model the state transitions. Wang et al. [28] developed IPTW based weighted least square state transition models. The interactions in the state transition models were linearized using a piecewise linear approximation method and modeled as constraints in a two-stage stochastic programming (2SP) framework. Iqbal et al. [27] used piecewise linear networks (PLN) to model the state transitions in a multiple objective two-stage stochastic programming framework, which included all the five pain outcome measures.

3.1.3 Contribution in State Transition and Outcome Modeling

Both LeBoulluec et al. [25] and Ohol [26] constructed state transition and outcome models to study their IPTW methods, but neither considered the importance of state-treatment interactions or addressed feature selection. Lin et al [22] did study interaction terms and feature selection via stepwise regression but did not address time-varying confounding. To address causal feature selection in the presence of time-varying confounding, Farahani et al. [152] developed a LASSO based outcome adaptive Elastic Net algorithm for

main effect (no interaction) models. The advantage of LASSO based methods is in building parsimonious models and handling correlation between variables, which is not the case with least squares-based methods. In this paper, LASSO based methods are combined with IPTW from Ohol [26] to build state transition and outcome models that enable feature selection and modeling of interaction effects in the presence of time-varying confounding. This approach is studied using a simulated case study structured based on the McDermott Center data with time varying confounding and various correlation structures between the variables. The benefit of a simulated case study is the ability to control the truth and compare the modeling results directly to the truth. The performance of the proposed modeling approach is compared against a baseline model that also attempts to model interactions but does not address time varying confounding.

3.2 Modeling Framework

3.2.1 HierNet

HierNet is a LASSO-based method developed by Bien et al. [30] for finding interactions, which produces sparse estimates of the main and interaction effects while satisfying the strong or weak hierarchy constraint. The optimization problem shown in Equation (7) is solved in HierNet

$$\text{argmin}_{\mu,\beta,\theta} \frac{1}{2}\sum_{i=1}^{n}(y_i - \mu - x_i^T\beta - \frac{1}{2}x_i^T\Theta x_i)^2 + \lambda 1^T(\beta^+ + \beta^-) + \frac{\lambda}{2}\|\Theta\|_1, \quad (7)$$

$$\text{subject to } \Theta = \Theta^T, \left\|\Theta_j\right\|_1 \leq \left(\beta_j^+ + \beta_j^-\right),\ \beta_j^+ \geq 0,\ \beta_j^- \geq 0,$$

where y_i is the response variable, x_i is the input vector, β is the main effect vector, λ is the regularization parameter, μ is the fixed intercept, β_j is the main effect of variable j, β_j^+ and β_j^+ are such that $|\beta_j| = (\beta^+ + \beta^-)$, Θ is the matrix of interaction effects, Θ_j is the jth row of Θ.

The constraint $\left\|\Theta_j\right\|_1 \leq \left(\beta_j^+ + \beta_j^-\right)$, is called the symmetry constraint and enforces strong hierarchy among the solutions. Weak hierarchy can be obtained by relaxing this constraint. In this study, we consider strong hierarchy since we want the main effects corresponding to the interactions to be in the model.

For this study, two important modifications were made to HierNet. They are summarized here and detailed in Appendix I. First, the optimization selects the optimal treatments with the objective of reducing patients' pain outcomes and treatment costs. In order to enable representation of all treatment variables, the HierNet algorithm was modified to maintain all treatment variables. Second, to address time varying confounding via IPTW, the HierNet loss function was modified to utilize weighted loss function. The complete iterative state transition and outcome modeling framework using IPTW and HierNet is explained in the next section.

3.2.2 HierNet-IPTW Modeling Framework

Because different weights on the observations will affect the results from feature selection, and different sets of features corresponding will affect the IPTW weights, an iterative process is presented in Figure 7 that alternates between generating IPTW weights and conducting feature selection with HierNet.

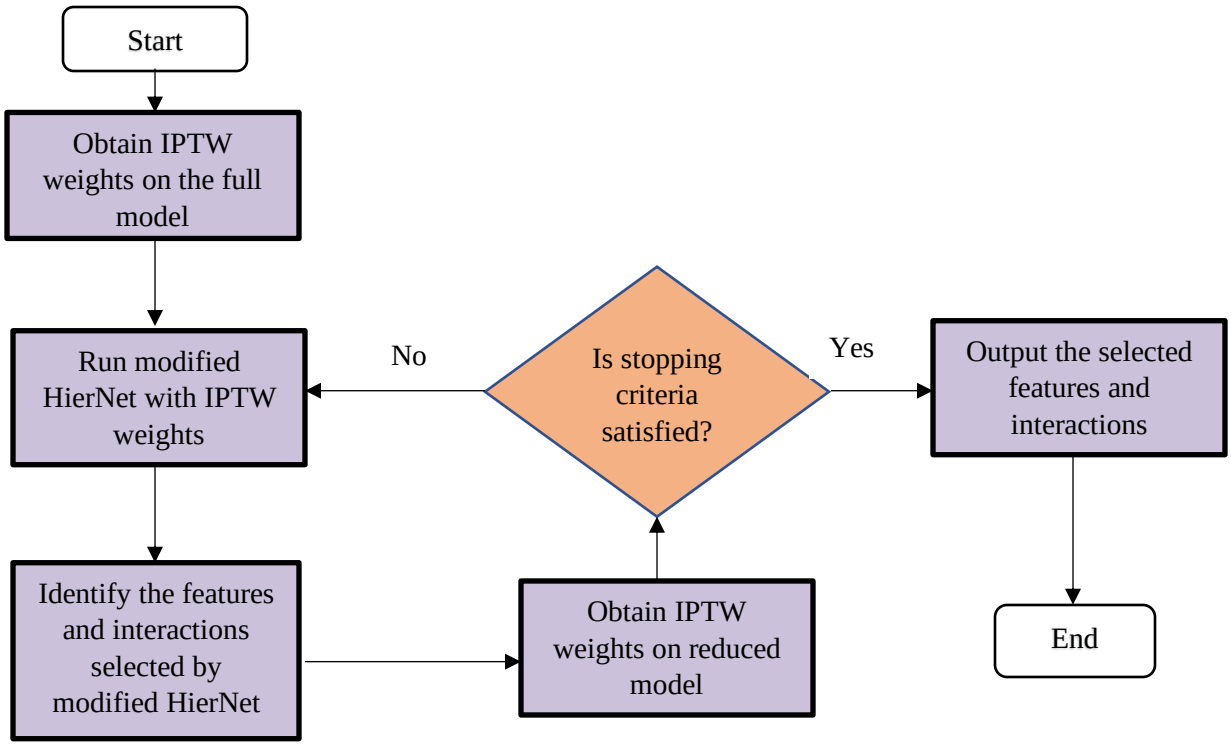


Figure 7: HierNet-IPTW modeling flowchart

The basic steps of the algorithm are as follows:

Step 1: In the first step, the IPTW weights are obtained on the full model in which all the state and decision variables are included.

Step 2: The modified HierNet algorithm is run using the IPTW weights obtained from the full model.

Step 3: The IPTW weights are obtained on the features identified by the HierNet algorithm in step 2. These weights are obtained on the reduced model using only the state and treatment variables from step 2 and are used to rebuild the models in step 2 in an iterative process till the stopping criteria are met.

Step 4: The iterative process of calculating the IPTW weights and rebuilding models using HierNet is conducted until the stopping criteria are satisfied. In this study, the stopping criteria are the percentage change in weights from the previous iteration. If the maximum percentage change in weights is less than 10%, the iterative process stops, and the selected features and interactions are output. The stopping criteria are shown in Equation (8).

$$\max \frac{(IPTW_n^i - IPTW_n^{i-1}) \times 100}{IPTW_n^{i-1}} < 10, \forall n \in N \quad (8)$$

where N is the total observations, $IPTW_n^i$ is the IPTW weight for observation n in iteration i, and $IPTW_n^{i-1}$ is the IPTW weight for observation n in iteration $i-1$.

The performance of the proposed HierNet-IPTW approach is compared against a baseline model, shown in Figure 8, that also attempts to model interactions but does not address time varying confounding.

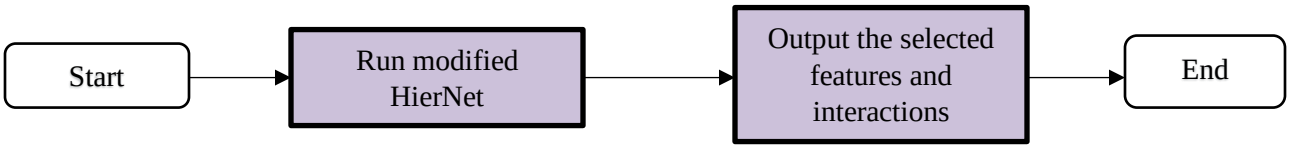


Figure 8: HierNet model flowchart

3.3 Pain Management Simulation Case Study

The primary goal of this simulation case study is to evaluate the performance of the proposed HierNet-IPTW modeling framework for identifying the true features and interactions using data with characteristics based on the McDermott Center data. Simulated data were created under various correlation conditions and time varying confounding to represent a variety of possible patterns in interdisciplinary pain management data. The causal diagram of the interdisciplinary pain management program is shown in Figure 9.

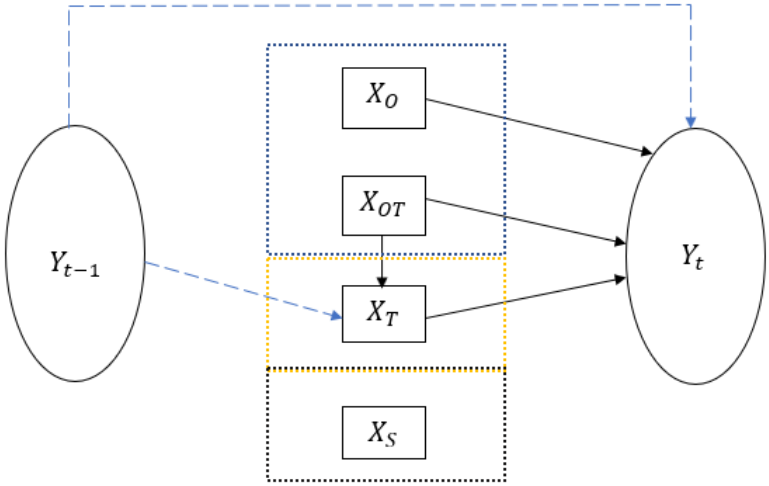


Figure 9: Pain Management Causal Diagram [168]

The causal diagram shows the second stage of the two-stage program, where Y_{t-1} is the pain outcome measured at the end of stage 1 of the program. The outcome measure Y_{t-1} becomes a state variable in stage 2, that influences the stage 2 pain outcome measure Y_t and the stage 2 treatment variables X_T . The spurious variables are X_S. The covariates are divided into two groups, outcome covariates X_O, that influence the stage 2 pain outcome measure Y_t, and confounding variables X_{OT}, that influence Y_t and the stage 2 treatment variables X_T. The confounding variables X_{OT} also include the treatment variables prescribed in stage 1 and the stage 1 pain outcome measure Y_{t-1}. All the stage 2 treatment variables are assumed to be causal variables influencing Y_t. This assumption is important from the optimization perspective since to optimize the treatment allocation, all the treatment variables need to be in the final state transition model. The

treatments are binary variables that follow the probability distribution as a function of Y_{t-1} and X_{OT}, shown in Equation (9).

$$P_{trt_t} = \frac{exp(\beta^T X_T + \beta_{trt}^{OT} X_{OT} + \gamma_{t-1} Y_{t-1})}{1 + exp(\beta^T X_T + \beta^{OT} X_{OT} + \gamma_{t-1} Y_{t-1})} \quad (9)$$

We create a covariate group, X_{CO} that includes X_O, X_{OT}, and Y_{t-1}. The pain outcome measure Y_t is a function of X_{CO}, X_T and interaction terms between them as shown in Equation (10)

$$Y_t = \beta^T X_T + \beta^{CO} X_{CO} + \gamma_{CO}^T X_T X_{CO} + \gamma_{CO}^{CO} X_{CO} X_{CO} + \gamma_T^T X_T X_T + \varepsilon_i \quad (10)$$

where β^T is the main effect of treatment X_T, β^{CO} is the main effect of covariates X_{CO}, γ_{OT}^T is the interaction effect between treatments and covariates, γ_{CO}^{CO} is the interaction effect between covariates and γ_T^T is the interaction effect between treatment variables. The error term ε_i is selected such that the proportion of variance explained by the true model is 0.9. The pain management data also has a few rare outlier observations. To mimic these, the noise factor was increased for 10% of the observations so that the proportion of variance explained by the true model is 0.75.

The second area of interest in this study is to understand the impact of different correlation structures between the treatments, covariates, and spurious variables on feature selection and interaction selection. The correlation structure used in the simulation study is shown in Figure 10, and is based on Farahani [168]

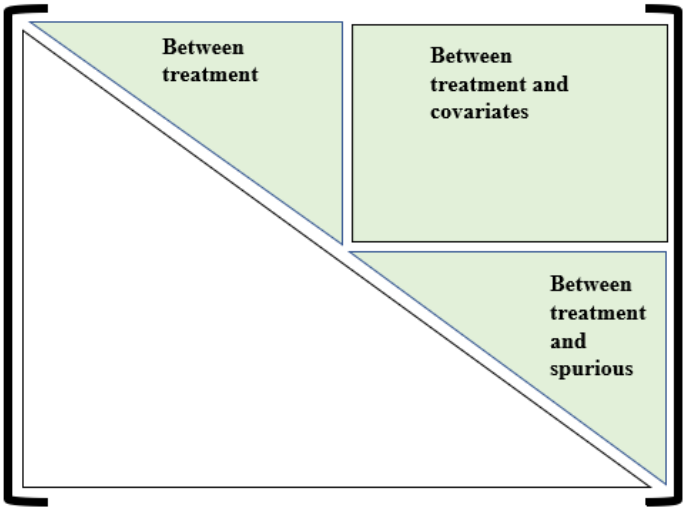


Figure 10: Correlation Design for the simulation case study [168]

The following are the factors considered in this simulation case study:

1. **Correlation between treatments**: {Low, High} The correlation between treatments were randomly generated to be between (0, 0.4) for low correlation and [0.5, 0.9] for high correlation.
2. **Correlation between treatments and covariates**: {Low, High} The correlation between treatments and covariates were randomly generated to be between (0, 0.4) for low correlation and [0.5, 0.9] for high correlation.
3. **Correlation between treatments and spurious variables**: {Low, High} The correlation between treatments and spurious variables were randomly generated to be between (0, 0.4) for low correlation and [0.5, 0.9] for high correlation.
4. **Use of IPTW**: {Baseline, IPTW-MIMIC} The baseline algorithm following Figure 8 does not use IPTW, so it does not address time-varying confounding. The IPTW-MIMIC algorithm following Figure 7 with IPTW using the MIMIC method from Ohol [26] to handle correlated treatments. The weights obtained from MIMIC are incorporated in the HierNet-IPTW model, which does the feature selection and model building as in Figure 7.

A 2^4 full factorial experiment with 100 replications and 100 observations per replication was conducted. Oher factors which were fixed in the study are shown in Table 4:

Table 4: Other factors in simulation case study

Number of treatments (X_T)	4
Number of confounding variables (X_{OT})	9
Number of outcome covariates (X_O)	4
Number of covariates (X_{CO})	13 (This is a union of X_{OT} and X_O)
Number of spurious variables	8
Number of randomly simulated interaction effects	10

3.4 Performance Metrics

The performance metrics considered in this case study are classified into two groups, namely feature selection metrics and interaction selection metrics.

3.4.1 Feature/Variable Selection Metrics

The confusion matrix is primarily used to evaluate the performance of classification models and is based on the count of observations in the test dataset correctly and incorrectly predicted by the model [169, 170]. The confusion matrix concept extended to feature/variable selection is shown in Table 5.

Table 5: Confusion matrix for feature selection [171]

		Predicted feature classes	
		Causal feature	Spurious feature
True feature classes	Causal features	a: The number of causal features classified correctly	b: The number of causal features classified incorrectly
	Spurious features	c: The number of spurious features classified incorrectly	d: The number of spurious features classified correctly

The confusion matrix for feature/variable selection is used to evaluate the feature selection performance of the model based on the features correctly and incorrectly identified by the model [171]. The proportion of correctly classified causal features among all causal features is called sensitivity and the proportion of correctly classified spurious features among all spurious features is called specificity. Sensitivity and specificity provide specific information on feature selection with regards to causal and spurious features respectively.

$$\text{Sensitivity} = \frac{a}{a+b} \tag{11}$$

$$\text{Specificity} = \frac{d}{c+d} \tag{12}$$

Sensitivity and Specificity values range between 0 and 1, with desired values closer to 1. Feature selection sensitivity and specificity are the two feature selection metrics used in this study.

3.4.2 Interaction Selection Metrics

The confusion matrix concept is extended to evaluate how the interactions are classified in the models and are shown in Table 6.

Table 6: Confusion matrix for interaction selection

		Predicted interaction classes	
		Causal interactions	Spurious interactions
True interaction classes	Causal interactions	p: The number of causal interactions classified correctly	q: The number of causal interactions classified incorrectly
	Spurious interactions	r: The number of spurious interactions classified incorrectly	s: The number of spurious interactions classified correctly

The proportion of correctly classified interactions among all causal interactions is called sensitivity and the proportion of correctly classified spurious interactions among all spurious interactions is called specificity. The sensitivity and specificity formulas shown in Equations (9) and (10) respectively are extended to the confusion matrix in Table 5. The False Discovery Rate (FDR) is the number of false interaction terms among all the predicted interaction terms in the model and was used by Lim and Hastie [163] in their study to compare the performance of Glinternet and HierNet.

$$\text{FDR} = \frac{r}{p+r} \tag{13}$$

Specificity and FDR both look at the number of spurious interactions in the model. The specificity values in this study would be very high since there are 25 state variables and 300 possible two-factor interaction effects. Both interaction sensitivity and FDR are used to compare the modeling methods in this study.

3.5 Simulation Experiment Results

The results of the 2^4 full factorial experiment conducted using pain management simulation data is analyzed in this section. The feature and interaction selection metrics are the outcomes of interest in this study. The factors controlled in the experiment are the correlation between treatments, the correlation between treatments and covariates, the correlation between treatments and spurious variables, and the use of IPTW.

3.5.1 Feature Selection Sensitivity

3.5.1.1 Preliminary Analysis

The main purpose of the experiment is to analyze the difference in feature selection performance between the proposed modeling method that uses IPTW-MIMIC and the baseline method with no IPTW, under different correlation structures between treatments, covariates, and spurious variables.

The boxplot in Figure 11 compares the use of IPTW under correlation between treatments. We can observe that with both the baseline and IPTW-MIMIC framework, the average sensitivity is higher with both the model frameworks when the correlation between treatments is low. We can infer that low correlation between treatments results in more true features being correctly identified by the model. We also observe a clear difference between the baseline and IPTW-MIMIC, with IPTW-MIMIC having higher average sensitivity than the baseline.

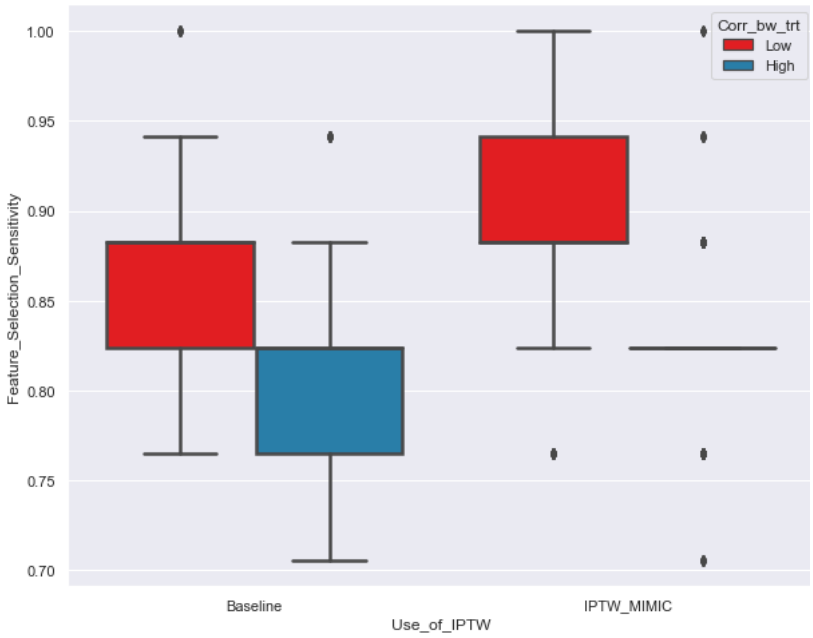


Figure 11: Feature selection sensitivity against the use of IPTW for different correlation structures between treatments

The boxplot in Figure 12 compares the use of IPTW under correlation between treatments and covariates. Similar to the previous plot, the average sensitivity is higher with both the model frameworks when the correlation between treatments and covariates is higher. There is a greater variation with IPTW-MIMIC at low correlation. A higher correlation between treatments and covariates results in more true features being picked by the model. The IPTW-MIMIC framework has higher average sensitivity than the baseline for a given correlation structure.

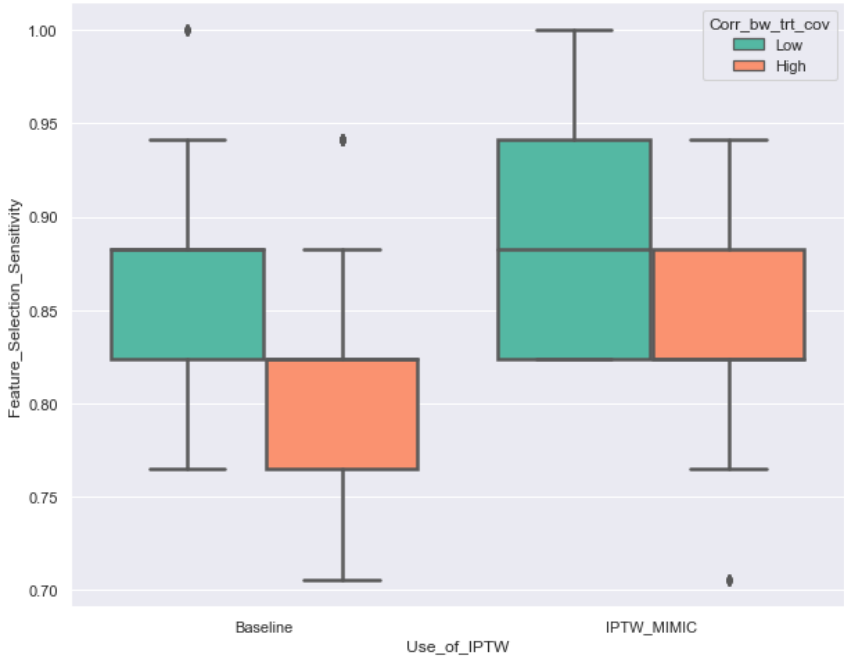


Figure 12: Feature selection sensitivity against the use of IPTW for different correlation structures between treatments and covariates

The boxplot in Figure 13 compares the use of IPTW under different correlations between treatments and spurious variables. We do not see any clear difference between the average sensitivity at high and low correlation levels. With the baseline model under high correlations, there is a greater variation in specificity than under low correlations. The performance of IPTW-MIMIC vs. the baseline are quite similar. A possible explanation for this could be that since we are maintaining all the treatment variables in the model, the correlation between treatments and spurious variables has a lower impact on feature selection.

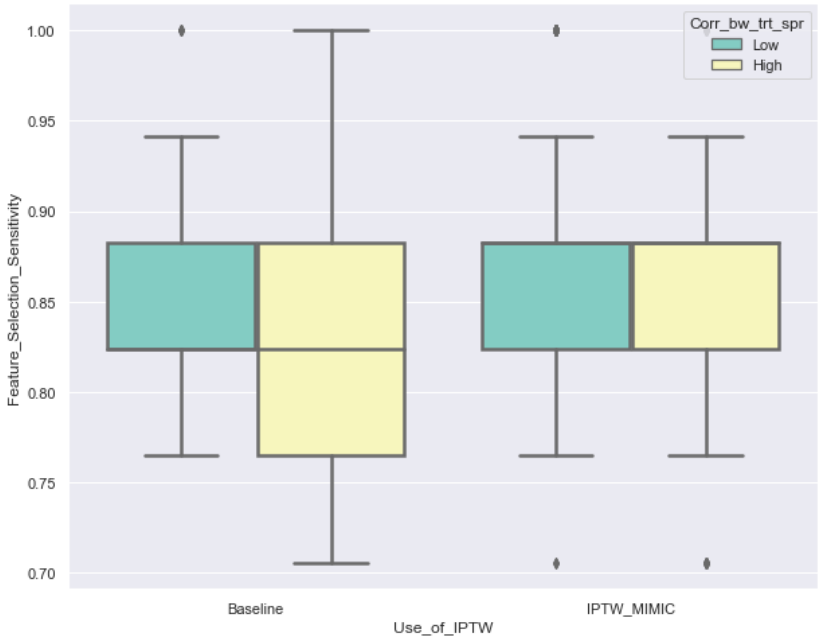


Figure 13: Feature selection sensitivity against the use of IPTW for different correlation structures between treatments and spurious variables

3.5.1.2 Analysis of Variance (ANOVA)

The correlation between treatments, the correlation between treatments and covariates, the correlation between treatments and spurious variables and the use of IPTW are the four factors studied in this ANOVA. The results of a full factorial ANOVA using SAS are shown in Table 7, with the highlighted boxes indicating the significant effects. The main effects of all the factors were significant. The two-factor interaction between the use of IPTW and the correlation between treatments and covariates was significant. The three-factor interaction between the use of IPTW, the correlation between treatments, and the correlation between treatments and covariates was significant at a significance level of $\alpha = 0.1$

From the results of the ANOVA, we can see that the feature selection sensitivity is affected by the use of IPTW and the different correlation structures. A Tukey multiple comparison considering the use of IPTW, the correlation between treatments, and the correlation between treatments and covariates is conducted since the interaction between these factors was found to be significant.

Table 7: ANOVA for feature selection sensitivity

Source	DF	Anova SS	Mean Square	F Value	Pr > F
Corr_Trt	1	1.60320285	1.60320285	909.37	<.0001
Corr_Trt_Cov	1	0.950625	0.950625	539.21	<.0001
Corr_Trt_Spr	1	0.04801254	0.04801254	27.23	<.0001
Use_of_IPTW	1	0.26643815	0.26643815	151.13	<.0001
Corr_Trt*Corr_Trt_Cov	1	0.00048659	0.00048659	0.28	0.5994
Corr_Trt*Corr_Trt_Spr	1	0.000625	0.000625	0.35	0.5517
Corr_Trt*Use_of_IPTW	1	0.00048659	0.00048659	0.28	0.5994
Corr_Trt_Cov*Corr_Trt_Spr	1	0.00264922	0.00264922	1.5	0.2204
Corr_Trt_Cov*Use_of_IPTW	1	0.03488106	0.03488106	19.79	<.0001
Corr_Trt_Spr*Use_of_IPTW	1	0.00017517	0.00017517	0.1	0.7526
Corr_Trt*Corr_Trt_Cov*Corr_Trt_Spr	1	0.00048659	0.00048659	0.28	0.5994
Corr_Trt*Corr_Trt_Cov*Use_of_IPTW	1	0.00654196	0.00654196	3.71	0.0542
Corr_Trt*Corr_Trt_Spr*Use_of_IPTW	1	0.00157656	0.00157656	0.89	0.3445
Corr_Trt_Cov*Corr_Trt_Spr*Use_of_IPTW	1	0.00437933	0.00437933	2.48	0.1152
Corr_Trt*Corr_Trt_Cov*Corr_Trt_Spr*Use_of_IPTW	1	0.0023551	0.0023551	1.34	0.2479

The experimental combinations are shown in Table 8, the p-value results of the Tukey analysis are shown in Table 9, and the 90% CI in Table 10 respectively. The Tukey line plot is shown in Figure 14. At significance level $\alpha = 0.1$, the experimental combinations 3 and 5 highlighted in the tables and figure are not statistically different. With the baseline model framework, the feature selection sensitivity is not statistically different when either one of the correlations between treatments and the correlation between treatment covariates is high and the other low. All other experimental combinations are statistically different.

Table 8: Tukey comparison factor combinations for feature selection sensitivity

Corr_Trt	Corr_Trt_Cov	Use_of_IPTW	Feature_Selection_Sensitivity LSMEAN	LSMEAN Number
High	High	Baseline	0.78176471	1
High	High	IPTW_MIMIC	0.81176471	2
High	Low	Baseline	0.83470588	3
High	Low	IPTW_MIMIC	0.85411765	4
Low	High	Baseline	0.83882353	5
Low	High	IPTW_MIMIC	0.87911765	6
Low	Low	Baseline	0.90205882	7
Low	Low	IPTW_MIMIC	0.91558824	8

Table 9: Tukey analysis p-value for feature selection sensitivity

Least Squares Means for effect Corr_T*Corr_T*Use_of
Pr > |t| for H0: LSMean(i)=LSMean(j)

Dependent Variable: Feature_Selection_Sensitivity

i/j	1	2	3	4	5	6	7	8
1		<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
2	<.0001		<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
3	<.0001	<.0001		0.0001	0.9772	<.0001	<.0001	<.0001
4	<.0001	<.0001	0.0001		0.0068	<.0001	<.0001	<.0001
5	<.0001	<.0001	0.9772	0.0068		<.0001	<.0001	<.0001
6	<.0001	<.0001	<.0001	<.0001	<.0001		<.0001	<.0001
7	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001		0.0283
8	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	0.0283	

Table 10: Tukey analysis 90% CI for feature selection sensitivity

Corr_Trt	Corr_Trt_Cov	Use_of_IPTW	Feature_Selection_Sensitivity LSMEAN	90% Confidence Limits	
High	High	Baseline	0.781765	0.776878	0.786651
High	High	IPTW_MIMIC	0.811765	0.806878	0.816651
High	Low	Baseline	0.834706	0.829819	0.839592
High	Low	IPTW_MIMIC	0.854118	0.849231	0.859004
Low	High	Baseline	0.838824	0.833937	0.843710
Low	High	IPTW_MIMIC	0.879118	0.874231	0.884004
Low	Low	Baseline	0.902059	0.897172	0.906945
Low	Low	IPTW_MIMIC	0.915588	0.910702	0.920475

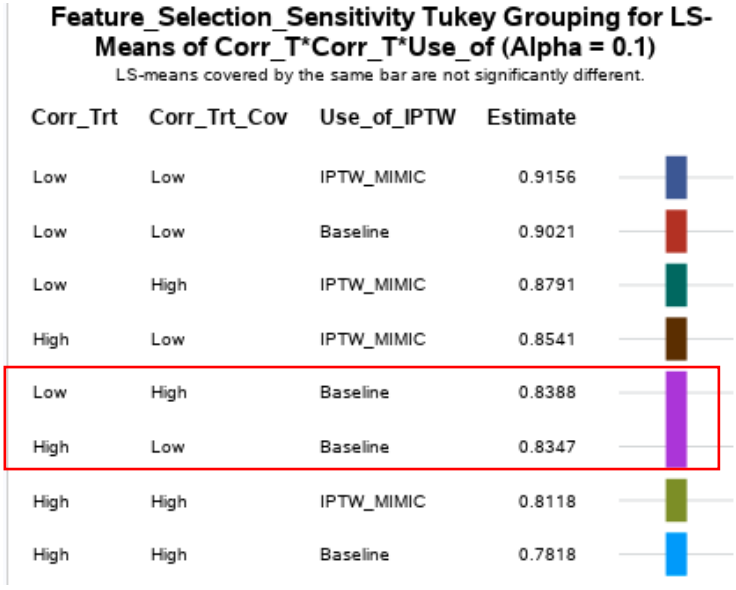


Figure 14: Tukey line plot for feature selection sensitivity

We perform a means analysis between the IPTW-MIMIC and the baseline across all the correlation combinations. This would give us an insight into the difference in performance with the use of IPTW. The results of this analysis are shown in Table 11. We can see that the difference in feature selection due to the use of IPTW is statistically different with the IPTW-MIMIC performing better than the baseline with higher average sensitivity. We can conclude that the IPTW-MIMIC enables true features to be identified by the modeling method better than the baseline without IPTW.

Table 11: Comparison of feature selection sensitivity means with the use of IPTW

Use_of_IPTW	Feature_Selection_Sensitivity LSMEAN	H0:LSMean1=LSMean2 Pr > |t|
Baseline	0.83933824	<.0001
IPTW_MIMIC	0.86514706	

Use_of_IPTW	Feature_Selection_Sensitivity LSMEAN	90% Confidence Limits	
Baseline	0.839338	0.836895	0.841781
IPTW_MIMIC	0.865147	0.862704	0.867590

Least Squares Means for Effect Use_of_IPTW				
i	j	Difference Between Means	Simultaneous 90% Confidence Limits for LSMean(i)-LSMean(j)	
1	2	-0.025809	-0.029264	-0.022354

The next analysis was to simultaneously compare the contrasts shown in Table 12, where each contrast is highlighted with the same color. We are simultaneously comparing the 8 contrasts using the Bonferroni method. These contrasts are defined to help us compare the use of IPTW for different correlation combinations between treatments, covariates, and spurious variables. The result of the simultaneous Bonferroni comparison for these 8 contrasts is shown in Table 13. The results show that the difference in performance between IPTW-MIMIC and the baseline is statistically significant for all the correlation combinations except the case where the correlation between treatments, the correlation between treatments and covariates is low, and the correlation between treatments and spurious variables is high. We can conclude from this analysis that the IPTW-MIMIC performs as well or better than the baseline without IPTW under the different correlation combinations.

Table 12: Contrast for simultaneous Bonferroni comparison

Corr_Trt	Corr_Trt_Cov	Corr_Trt_Spr	Use_of_IPTW
High	High	High	Baseline
High	High	High	IPTW-MIMIC
High	High	Low	Baseline
High	High	Low	IPTW-MIMIC
High	Low	High	Baseline
High	Low	High	IPTW-MIMIC
High	Low	Low	Baseline
High	Low	Low	IPTW-MIMIC
Low	High	High	Baseline
Low	High	High	IPTW-MIMIC
Low	High	Low	Baseline
Low	High	Low	IPTW-MIMIC
Low	Low	High	Baseline
Low	Low	High	IPTW-MIMIC
Low	Low	Low	Baseline
Low	Low	Low	IPTW-MIMIC

Table 13: Simultaneous Bonferroni comparison results for feature selection sensitivity

Corr_Trt	Corr_Trt_Cov	Corr_Trt_Spr	DF	Sum of Squares	Mean Square	F Value	Pr > F
High	High	High	1	0.056211	0.056211	31.88	<.0001
High	High	Low	1	0.035035	0.035035	19.87	<.0001
High	Low	High	1	0.022422	0.022422	12.72	0.0004
High	Low	Low	1	0.015571	0.015571	8.83	0.0030
Low	High	High	1	0.099931	0.099931	56.68	<.0001
Low	High	Low	1	0.064377	0.064377	36.52	<.0001
Low	Low	High	1	0.002093	0.002093	1.19	0.2760
Low	Low	Low	1	0.021194	0.021194	12.02	0.0005

The residual analysis is conducted to check the assumptions that the error terms are normally distributed and have constant variance. This is discussed in Appendix II.

3.5.2 Feature Selection Specificity

3.5.2.1 Preliminary Analysis

The boxplot in Figure 15 compares the use of IPTW under different correlation between treatments. The average specificity is similar at high and low correlation levels, and with both the baseline and IPTW-MIMIC. There is a greater variation in specificity with the baseline at high correlation between treatments. Specificity measures how many spurious variables are correctly identified as spurious by the model, the correlation between treatments does not directly impact specificity as all the treatment variables in the case study are causal variables and not spurious. That is why specificity is similar with both IPTW-MIMIC and baseline under different correlations between treatments.

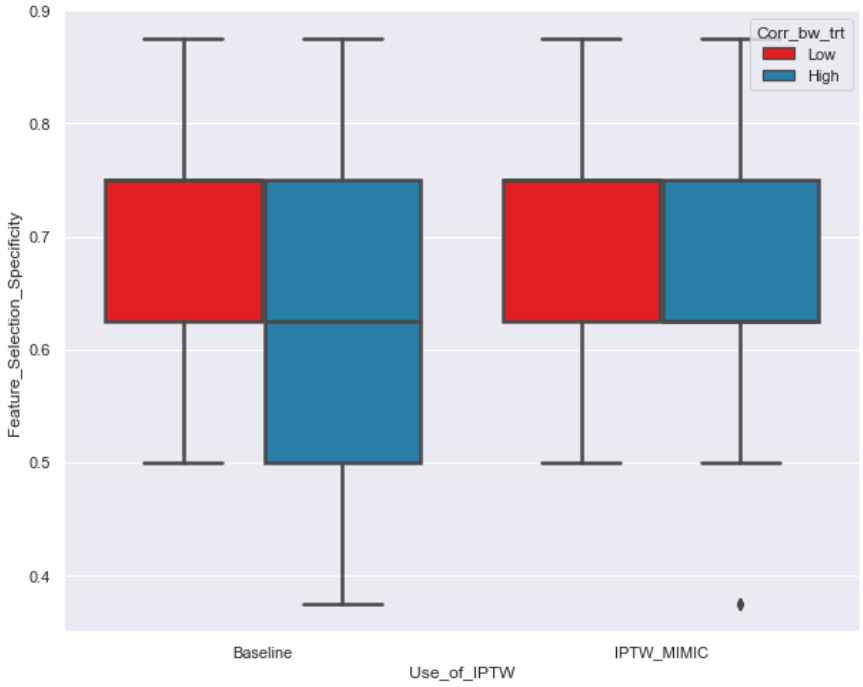


Figure 15: Feature selection specificity against the use of IPTW for different correlation structures between treatments

The boxplot in Figure 16 compares the use of IPTW under different correlation between treatments and covariates. The interpretation of these plots is similar to the previous figure, where we do not see major difference under the different correlation levels and between the use of IPTW. The effect of correlation between treatments and covariates on average specificity is similar between IPTW-MIMIC and the baseline. The variation in specificity is high with the baseline at high correlation. We can infer that the correlation between treatments and covariates along with the use of IPTW does not directly impact the selection of spurious variables in the model.

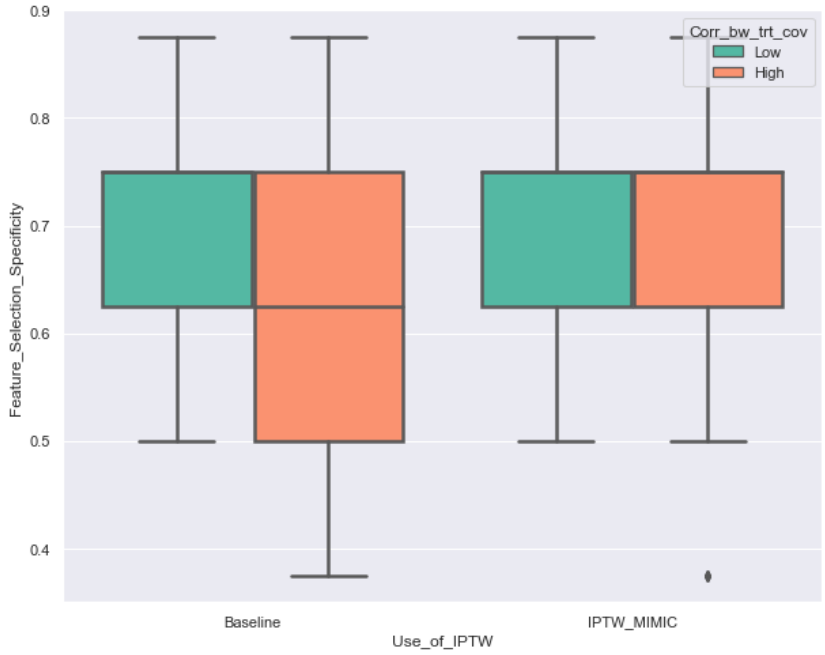


Figure 16: Feature selection specificity against the use of IPTW for different correlation structures between treatments and covariates

The boxplot in Figure 17 shows a clear difference in average specificity under different correlation between treatments and spurious variables. The average specificity is less at high correlation than at low correlation with both the baseline and IPTW-MIMIC. A high correlation between treatments and spurious variables results in more spurious variables being included in the model, and hence lower average specificity. The average specificity is greater with IPTW-MIMIC than the baseline indicating that it

performs better when there is correlation between treatments and spurious variables. The variation in specificity is similar across the two methods.

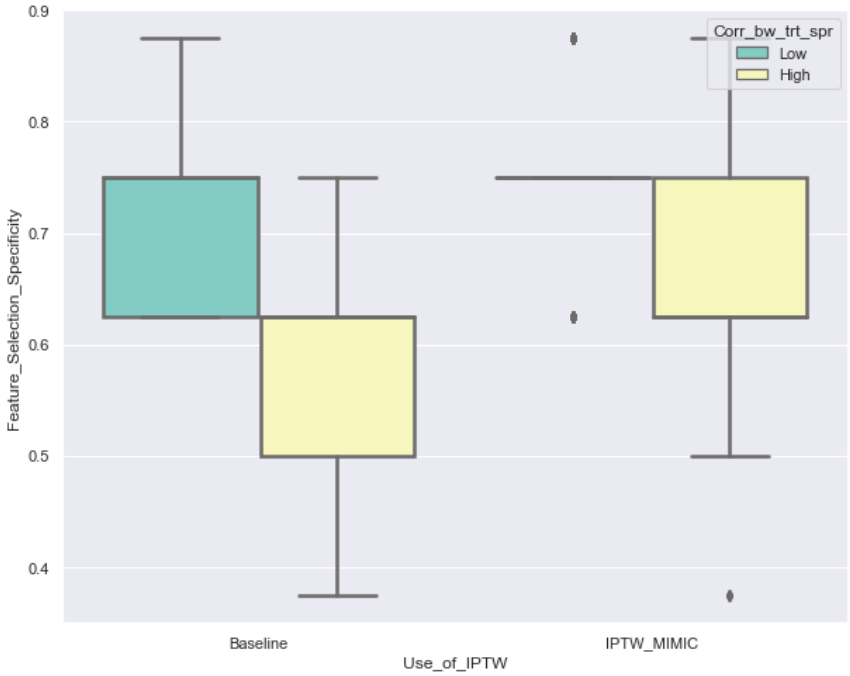


Figure 17: Feature selection specificity against the use of IPTW for different correlation structures between treatments and spurious variables

3.5.2.2 Analysis of Variance (ANOVA)

The correlation between treatments, the correlation between treatments and covariates, the correlation between treatments and spurious variables and the use of IPTW are the four factors studied in this ANOVA. The results of a full factorial ANOVA using SAS are shown in Table 14, with the highlighted boxes indicating the significant effects. The significance level $\alpha = 0.1$ was used in this analysis. The main effects of all the factors were significant. All the two-factor interaction effects except for the interaction between the use of IPTW with the correlation between treatments and the use of IPTW with the correlation between treatments and covariates were significant, which follows the inferences from the preliminary boxplot analysis. The three-factor interaction between the correlation between treatments, the correlation between treatments and covariates, and the correlation between treatments and spurious variables was significant.

Table 14: ANOVA for feature selection specificity

Source	DF	Anova SS	Mean Square	F Value	Pr > F
Corr_Trt	1	1.27266602	1.27266602	244.65	<.0001
Corr_Trt_Cov	1	0.47696289	0.47696289	91.69	<.0001
Corr_Trt_Spr	1	5.56665039	5.56665039	1070.11	<.0001
Use_of_IPTW	1	0.31290039	0.31290039	60.15	<.0001
Corr_Trt*Corr_Trt_Cov	1	0.21102539	0.21102539	40.57	<.0001
Corr_Trt*Corr_Trt_Spr	1	0.54852539	0.54852539	105.45	<.0001
Corr_Trt*Use_of_IPTW	1	0.00821289	0.00821289	1.58	0.2091
Corr_Trt_Cov*Corr_Trt_Spr	1	0.19415039	0.19415039	37.32	<.0001
Corr_Trt_Cov*Use_of_IPTW	1	0.00610352	0.00610352	1.17	0.2789
Corr_Trt_Spr*Use_of_IPTW	1	0.02954102	0.02954102	5.68	0.0173
Corr_Trt*Corr_Trt_Cov*Corr_Trt_Spr	1	0.13829102	0.13829102	26.58	<.0001
Corr_Trt*Corr_Trt_Cov*Use_of_IPTW	1	0.00165039	0.00165039	0.32	0.5733
Corr_Trt*Corr_Trt_Spr*Use_of_IPTW	1	0.00352539	0.00352539	0.68	0.4105
Corr_Trt_Cov*Corr_Trt_Spr*Use_of_IPTW	1	0.00821289	0.00821289	1.58	0.2091
Corr_Trt*Corr_Trt_Cov*Corr_Trt_Spr*Use_of_IPTW	1	0.00165039	0.00165039	0.32	0.5733

A Tukey multiple comparison considering all the factor combinations is conducted since the different significant interaction effects included all the factors. The factor combinations are shown in Table 15, the p-value results of the Tukey analysis are shown in Table 16, and the 90% CI in Table 17 respectively. At significance level $\alpha = 0.1$, the highlighted cells in Table 15 indicate the factor combinations whose p-value is greater than 0.1, and hence are not statistically different. In the Tukey line plot, shown in Figure 18, the factor combinations covered by the same-colored bar are not statistically different. We can observe that several factor combinations are not statistically different under low correlation between treatments and spurious variables, while at the high correlation between treatments and spurious variables, there are more statistically different combinations. The factor combination with the baseline and IPTW-MIMIC with the corresponding correlation factors high are the only two that are statistically different from the rest of the factor combinations.

Table 15: Tukey comparison factor combinations for feature selection specificity

Least Squares Means
Adjustment for Multiple Comparisons: Tukey

Corr_Trt	Corr_Trt_Cov	Corr_Trt_Spr	Use_of_IPTW	Feature_Selection_Specificity LSMEAN	LSMEAN Number
High	High	High	Baseline	0.50125000	1
High	High	High	IPTW_MIMIC	0.54375000	2
High	High	Low	Baseline	0.70500000	3
High	High	Low	IPTW_MIMIC	0.73125000	4
High	Low	High	Baseline	0.60375000	5
High	Low	High	IPTW_MIMIC	0.63750000	6
High	Low	Low	Baseline	0.72125000	7
High	Low	Low	IPTW_MIMIC	0.74875000	8
Low	High	High	Baseline	0.63375000	9
Low	High	High	IPTW_MIMIC	0.68125000	10
Low	High	Low	Baseline	0.73625000	11
Low	High	Low	IPTW_MIMIC	0.74750000	12
Low	Low	High	Baseline	0.66125000	13
Low	Low	High	IPTW_MIMIC	0.68375000	14
Low	Low	Low	Baseline	0.74375000	15
Low	Low	Low	IPTW_MIMIC	0.75625000	16

Table 16: Tukey analysis p-value for feature selection specificity

Least Squares Means for effect Corr*Corr*Corr*IPTW_
Pr > |t| for H0: LSMean(i)=LSMean(j)

Dependent Variable: Feature_Selection_Specificity

i/j	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1		0.0034	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
2	0.0034		<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
3	<.0001	<.0001		0.4180	<.0001	<.0001	0.9697	0.0020	<.0001	0.6024	0.1445	0.0034	0.0020	0.7758	0.0142	<.0001
4	<.0001	<.0001	0.4180		<.0001	<.0001	0.9998	0.9428	<.0001	0.0001	1.0000	0.9697	<.0001	0.0004	0.9979	0.5090
5	<.0001	<.0001	<.0001	<.0001		0.0729	<.0001	<.0001	0.1961	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
6	<.0001	<.0001	<.0001	<.0001	0.0729		<.0001	<.0001	1.0000	0.0020	<.0001	<.0001	0.6024	0.0007	<.0001	<.0001
7	<.0001	<.0001	0.9697	0.9998	<.0001	<.0001		0.3338	<.0001	0.0090	0.9856	0.4180	<.0001	0.0221	0.6930	0.0500
8	<.0001	<.0001	0.0020	0.9428	<.0001	<.0001	0.3338		<.0001	<.0001	0.9979	1.0000	<.0001	<.0001	1.0000	1.0000
9	<.0001	<.0001	<.0001	<.0001	0.1961	1.0000	<.0001	<.0001		0.0004	<.0001	<.0001	0.3338	0.0001	<.0001	<.0001
10	<.0001	<.0001	0.6024	0.0001	<.0001	0.0020	0.0090	<.0001	0.0004		<.0001	<.0001	0.8464	1.0000	<.0001	<.0001
11	<.0001	<.0001	0.1445	1.0000	<.0001	<.0001	0.9856	0.9979	<.0001	<.0001		0.9994	<.0001	<.0001	1.0000	0.8464
12	<.0001	<.0001	0.0034	0.9697	<.0001	<.0001	0.4180	1.0000	<.0001	<.0001	0.9994		<.0001	<.0001	1.0000	1.0000
13	<.0001	<.0001	0.0020	<.0001	<.0001	0.6024	<.0001	<.0001	0.3338	0.8464	<.0001	<.0001		0.6930	<.0001	<.0001
14	<.0001	<.0001	0.7758	0.0004	<.0001	0.0007	0.0221	<.0001	0.0001	1.0000	<.0001	<.0001	0.6930		<.0001	<.0001
15	<.0001	<.0001	0.0142	0.9979	<.0001	<.0001	0.6930	1.0000	<.0001	<.0001	1.0000	1.0000	<.0001	<.0001		0.9979
16	<.0001	<.0001	<.0001	0.5090	<.0001	<.0001	0.0500	1.0000	<.0001	<.0001	0.8464	1.0000	<.0001	<.0001	0.9979	

Table 17: Tukey analysis 90% CI for feature selection specificity

Corr_Trt	Corr_Trt_Cov	Corr_Trt_Spr	Use_of_IPTW	Feature_Selection_Specificity LSMEAN	90% Confidence Limits	
High	High	High	Baseline	0.501250	0.489380	0.513120
High	High	High	IPTW_MIMIC	0.543750	0.531880	0.555620
High	High	Low	Baseline	0.705000	0.693130	0.716870
High	High	Low	IPTW_MIMIC	0.731250	0.719380	0.743120
High	Low	High	Baseline	0.603750	0.591880	0.615620
High	Low	High	IPTW_MIMIC	0.637500	0.625630	0.649370
High	Low	Low	Baseline	0.721250	0.709380	0.733120
High	Low	Low	IPTW_MIMIC	0.748750	0.736880	0.760620
Low	High	High	Baseline	0.633750	0.621880	0.645620
Low	High	High	IPTW_MIMIC	0.681250	0.669380	0.693120
Low	High	Low	Baseline	0.736250	0.724380	0.748120
Low	High	Low	IPTW_MIMIC	0.747500	0.735630	0.759370
Low	Low	High	Baseline	0.661250	0.649380	0.673120
Low	Low	High	IPTW_MIMIC	0.683750	0.671880	0.695620
Low	Low	Low	Baseline	0.743750	0.731880	0.755620
Low	Low	Low	IPTW_MIMIC	0.756250	0.744380	0.768120

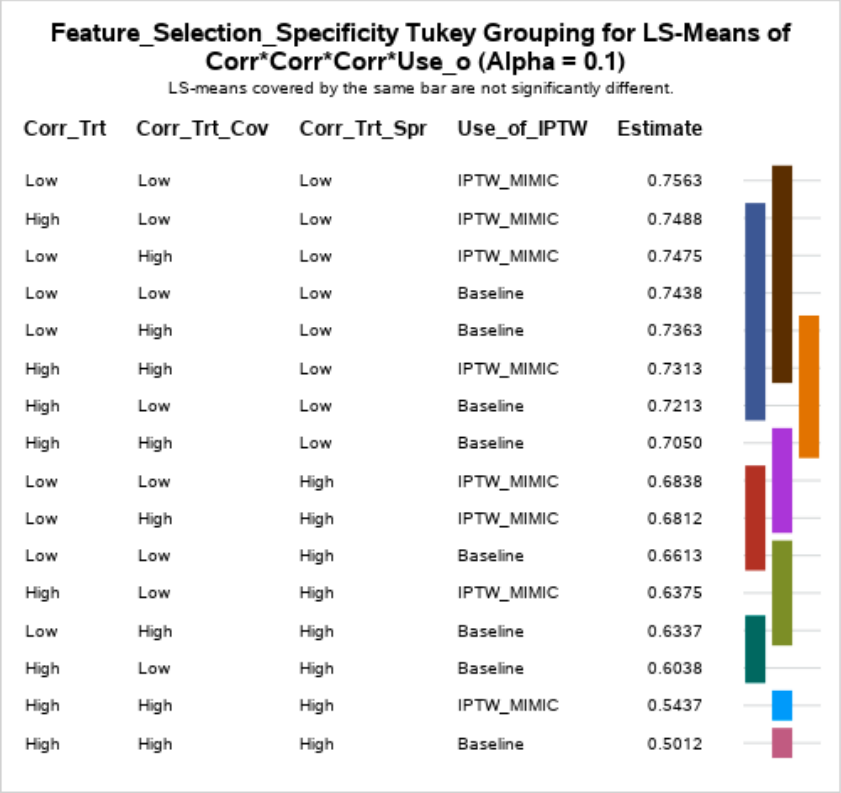


Figure 18: Tukey line plot for feature selection specificity

A means analysis of the use of IPTW across all the correlation combinations was performed. The results of this analysis are shown in Table 18. We can see that the difference in average feature selection specificity due to the use of IPTW is statistically significant with IPTW-MIMIC having higher average specificity than the baseline.

Table 18: Comparison of feature selection specificity means with IPTW model form

Use_of_IPTW	Feature_Selection_Specificity LSMEAN	H0:LSMean1=LSMean2 Pr > |t|
Baseline	0.66328125	<.0001
IPTW_MIMIC	0.69125000	

Use_of_IPTW	Feature_Selection_Specificity LSMEAN	90% Confidence Limits	
Baseline	0.663281	0.659084	0.667478
IPTW_MIMIC	0.691250	0.687053	0.695447

Least Squares Means for Effect Use_of_IPTW				
i	j	Difference Between Means	Simultaneous 90% Confidence Limits for LSMean(i)-LSMean(j)	
1	2	-0.027969	-0.033904	-0.022034

A simultaneous Bonferroni comparison of contrasts is performed to identify the performance of the use of IPTW across the different correlation combinations. The set of contrasts for comparison is shown in Table 12, where each contrast is highlighted with the same color. The result of the simultaneous Bonferroni comparison is shown in Table 19. The results show that the difference in average specificity between IPTW-MIMIC and the baseline approach is statistically significant for all the correlation combinations except the two highlighted. When the correlation between treatments, and the correlation between treatments and spurious variables is low, the difference between IPTW-MIMIC and the baseline is not statistically significant. We can conclude from this analysis that the IPTW-MIMIC performs as well or better than the baseline without IPTW under the different correlation combinations, and the correlation between treatments and spurious variables is an important factor in determining model performance.

Table 18: Simultaneous Bonferroni comparison results for feature selection specificity

Corr_Trt	Corr_Trt_Cov	Corr_Trt_Spr	DF	Sum of Squares	Mean Square	F Value	Pr > F
High	High	High	1	0.090313	0.090313	17.36	<.0001
High	High	Low	1	0.034453	0.034453	6.62	0.0102
High	Low	High	1	0.056953	0.056953	10.95	0.0010
High	Low	Low	1	0.037812	0.037812	7.27	0.0071
Low	High	High	1	0.112813	0.112813	21.69	<.0001
Low	High	Low	1	0.006328	0.006328	1.22	0.2702
Low	Low	High	1	0.025313	0.025313	4.87	0.0275
Low	Low	Low	1	0.007812	0.007812	1.50	0.2206

3.5.3 Interaction Selection Sensitivity

3.5.3.1 Preliminary Analysis

The boxplot in Figure 19 shows the impact of the use of IPTW along with the correlation between treatments on interaction selection sensitivity. From the plot, we can observe that the average and variation in interaction sensitivity are similar with IPTW-MIMIC and the baseline under both low and high correlation.

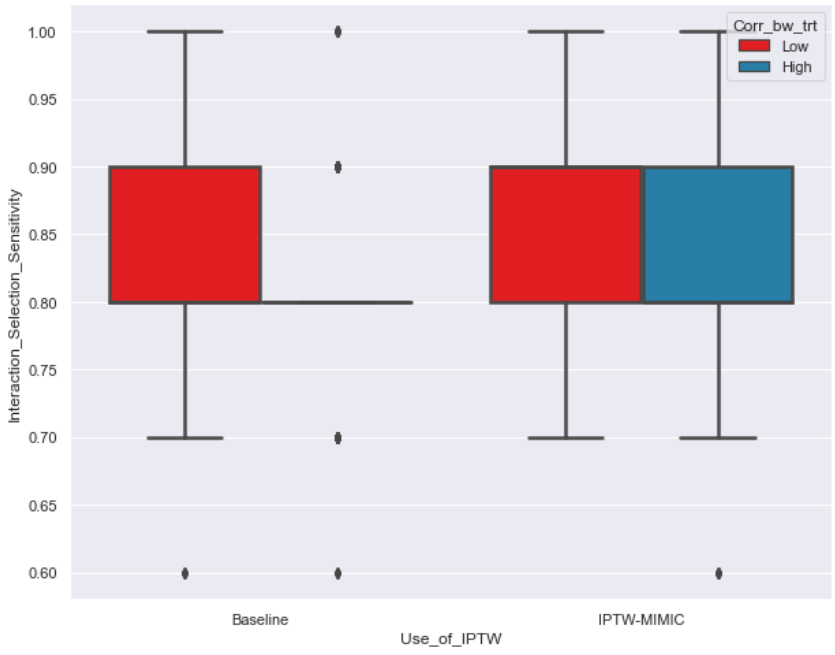


Figure 19: Interaction selection sensitivity against the use of IPTW for different correlation structures between treatments

The boxplot in Figure 20 shows the use of IPTW against the correlation between treatments and covariates. The average sensitivity values are similar across the use of IPTW and correlation combinations. The variation in sensitivity is also similar. This is similar to the boxplot in Figure 19. The use of IPTW along with the correlation between treatments or correlation between treatments and covariates does not show any significant difference in interaction selection.

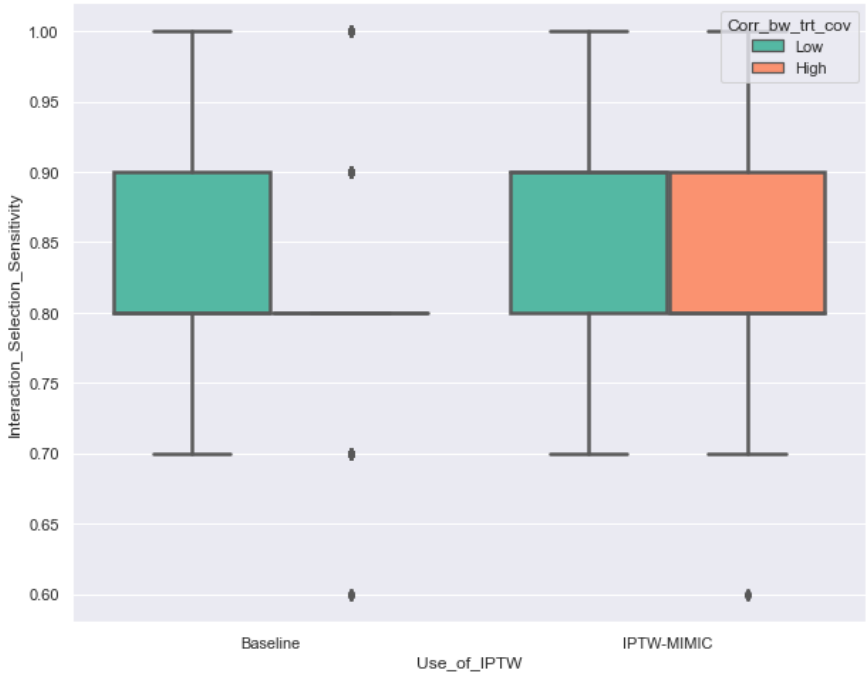


Figure 20: Interaction selection sensitivity against the use of IPTW for different correlation structures between treatments and covariates

The boxplot in Figure 21 shows interaction sensitivity plotted against the use of IPTW for a given correlation structure between treatments and spurious variables. With IPTW-MIMIC, the average sensitivity is similar at low and high correlation, while with the baseline, the average sensitivity is lower at high correlation. The variation is similar across the boxplots. We can infer that at high correlation between treatments and spurious variables, the baseline approach results in less causal interaction terms being identified than with the IPTW-MIMIC approach.

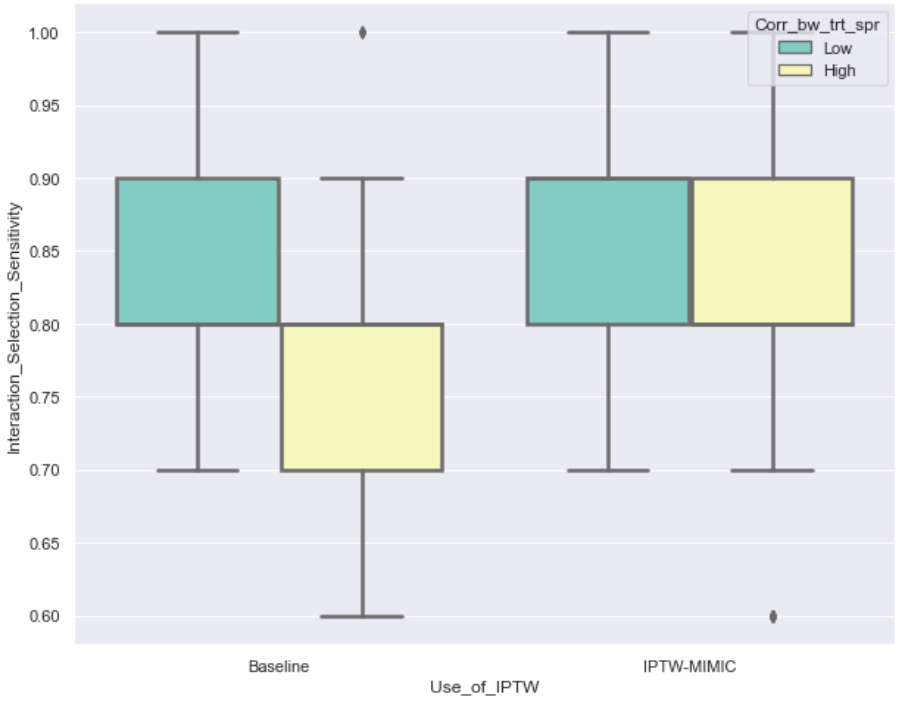


Figure 21: Interaction selection sensitivity against the use of IPTW framework for different correlation structures between treatments and spurious variables

3.5.3.2 Analysis of Variance (ANOVA)

The correlation between treatments, the correlation between treatments and covariates, the correlation between treatments and spurious variables and the use of IPTW are the four factors studied in this ANOVA. The results of a full factorial ANOVA using SAS are shown in Table 20, with the highlighted boxes indicating the significant effects. The significance level $\alpha = 0.1$ was used in this analysis. The main effects of all the factors were significant. The two-factor interaction effects that were statistically significant involved the correlation between treatments and spurious variables along with the correlation between treatments and the correlation between treatments and covariates respectively. A three-factor interaction effect between the use of IPTW, the correlation between treatments, and the correlation between treatments and spurious variables is significant.

A Tukey multiple comparison considering all the factor combinations is conducted since the different significant interaction effects included all the factors in the study.

Table 20: ANOVA for interaction selection sensitivity

Source	DF	Anova SS	Mean Square	F Value	Pr > F
Corr_Trt	1	0.85100625	0.85100625	193	<.0001
Corr_Trt_Cov	1	0.58905625	0.58905625	133.59	<.0001
Corr_Trt_Spr	1	1.53140625	1.53140625	347.3	<.0001
Use_of_IPTW	1	0.40005625	0.40005625	90.73	<.0001
Corr_Trt*Corr_Trt_Cov	1	0.00075625	0.00075625	0.17	0.6788
Corr_Trt*Corr_Trt_Spr	1	0.02975625	0.02975625	6.75	0.0095
Corr_Trt*Use_of_IPTW	1	0.00030625	0.00030625	0.07	0.7922
Corr_Trt_Cov*Corr_Trt_Spr	1	0.07700625	0.07700625	17.46	<.0001
Corr_Trt_Cov*Use_of_IPTW	1	0.00050625	0.00050625	0.11	0.7348
Corr_Trt_Spr*Use_of_IPTW	1	0.00525625	0.00525625	1.19	0.2751
Corr_Trt*Corr_Trt_Cov*Corr_Trt_Spr	1	0.00390625	0.00390625	0.89	0.3467
Corr_Trt*Corr_Trt_Cov*Use_of_IPTW	1	0.00140625	0.00140625	0.32	0.5723
Corr_Trt*Corr_Trt_Spr*Use_of_IPTW	1	0.01500625	0.01500625	3.4	0.0653
Corr_Trt_Cov*Corr_Trt_Spr*Use_of_IPTW	1	0.00030625	0.00030625	0.07	0.7922
Corr_Trt*Corr_Trt_Cov*Corr_Trt_Spr*Use_of_IPTW	1	0.00950625	0.00950625	2.16	0.1422

The factor combinations for the Tukey multiple comparison are shown in Table 21, the p-value results of the Tukey analysis are shown in Table 22, and the 90% CI in Table 23 respectively. At significance level $\alpha = 0.1$, the highlighted cells in Table 22 indicate the factor combinations whose p-value is greater than 0.1, and hence are not statistically different. In the Tukey line plot, shown in Figure 22, the factor combinations covered by the same-colored bar are not statistically different.

From the Tukey Line plot, we can see different groups of factor combinations are not statistically different. The baseline approach with high correlation across the other factors is the only statistically different combination not overlapping with other groups. The two-factor and three-factor interaction effects between the factors could be one possible reason we have different groups of factor combinations not statistically different from each other. To better understand the difference between the performance of IPTW-MIMIC and the baseline across the different correlation factors, a means analysis is performed.

Table 21: Tukey comparison factor combinations for interaction selection sensitivity

Adjustment for Multiple Comparisons: Tukey

Corr_Trt	Corr_Trt_Cov	Corr_Trt_Spr	Use_of_IPTW	Interaction_Selectn_Sensitivity LSMEAN	LSMEAN Number
High	High	High	Baseline	0.73500000	1
High	High	High	IPTW-MIMIC	0.77200000	2
High	High	Low	Baseline	0.82400000	3
High	High	Low	IPTW-MIMIC	0.85800000	4
High	Low	High	Baseline	0.79600000	5
High	Low	High	IPTW-MIMIC	0.81900000	6
High	Low	Low	Baseline	0.84300000	7
High	Low	Low	IPTW-MIMIC	0.87900000	8
Low	High	High	Baseline	0.79300000	9
Low	High	High	IPTW-MIMIC	0.82700000	10
Low	High	Low	Baseline	0.86100000	11
Low	High	Low	IPTW-MIMIC	0.88700000	12
Low	Low	High	Baseline	0.83700000	13
Low	Low	High	IPTW-MIMIC	0.88400000	14
Low	Low	Low	Baseline	0.89500000	15
Low	Low	Low	IPTW-MIMIC	0.91100000	16

Table 22: Tukey analysis p-value for interaction selection sensitivity

Least Squares Means for effect Corr*Corr*Corr*IPTW_
Pr > |t| for H0: LSMean(i)=LSMean(j)

Dependent Variable: Interaction_Effects_Sensitivity

i/j	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1		0.0084	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
2	0.0084		<.0001	<.0001	0.4310	<.0001	<.0001	<.0001	0.6711	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
3	<.0001	<.0001		0.0268	0.1778	1.0000	0.8122	<.0001	0.0746	1.0000	0.0084	<.0001	0.9921	<.0001	<.0001	<.0001
4	<.0001	<.0001	0.0268		<.0001	0.0036	0.9689	0.6711	<.0001	0.0746	1.0000	0.1356	0.6711	0.2884	0.0084	<.0001
5	<.0001	0.4310	0.1778	<.0001		0.5104	<.0001	<.0001	1.0000	0.0746	<.0001	<.0001	0.0015	<.0001	<.0001	<.0001
6	<.0001	<.0001	1.0000	0.0036	0.5104		0.4310	<.0001	0.2884	1.0000	0.0009	<.0001	0.8682	<.0001	<.0001	<.0001
7	<.0001	<.0001	0.8122	0.9689	<.0001	0.4310		0.0125	<.0001	0.9460	0.8682	0.0003	1.0000	0.0015	<.0001	<.0001
8	<.0001	<.0001	<.0001	0.6711	<.0001	<.0001	0.0125		<.0001	<.0001	0.8682	1.0000	0.0009	1.0000	0.9460	0.0539
9	<.0001	0.6711	0.0746	<.0001	1.0000	0.2884	<.0001	<.0001		0.0268	<.0001	<.0001	0.0003	<.0001	<.0001	<.0001
10	<.0001	<.0001	1.0000	0.0746	0.0746	1.0000	0.9460	<.0001	0.0268		0.0268	<.0001	0.9996	<.0001	<.0001	<.0001
11	<.0001	<.0001	0.0084	1.0000	<.0001	0.0009	0.8682	0.8682	<.0001	0.0268		0.2884	0.4310	0.5104	0.0268	<.0001
12	<.0001	<.0001	<.0001	0.1356	<.0001	<.0001	0.0003	1.0000	<.0001	<.0001	0.2884		<.0001	1.0000	1.0000	0.4310
13	<.0001	<.0001	0.9921	0.6711	0.0015	0.8682	1.0000	0.0009	0.0003	0.9996	0.4310	<.0001		<.0001	<.0001	<.0001
14	<.0001	<.0001	<.0001	0.2884	<.0001	<.0001	0.0015	1.0000	<.0001	<.0001	0.5104	1.0000	<.0001		0.9987	0.2287
15	<.0001	<.0001	<.0001	0.0084	<.0001	<.0001	<.0001	0.9460	<.0001	<.0001	0.0268	1.0000	<.0001	0.9987		0.9460
16	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	0.0539	<.0001	<.0001	<.0001	0.4310	<.0001	0.2287	0.9460	

Table 23: Tukey analysis 90% CI for interaction selection sensitivity

Corr_Trt	Corr_Trt_Cov	Corr_Trt_Spr	Use_of_IPTW	Interaction_Selectn_Sensitivity LSMEAN	90% Confidence Limits	
High	High	High	Baseline	0.735000	0.724071	0.745929
High	High	High	IPTW-MIMIC	0.772000	0.761071	0.782929
High	High	Low	Baseline	0.824000	0.813071	0.834929
High	High	Low	IPTW-MIMIC	0.858000	0.847071	0.868929
High	Low	High	Baseline	0.796000	0.785071	0.806929
High	Low	High	IPTW-MIMIC	0.819000	0.808071	0.829929
High	Low	Low	Baseline	0.843000	0.832071	0.853929
High	Low	Low	IPTW-MIMIC	0.879000	0.868071	0.889929
Low	High	High	Baseline	0.793000	0.782071	0.803929
Low	High	High	IPTW-MIMIC	0.827000	0.816071	0.837929
Low	High	Low	Baseline	0.861000	0.850071	0.871929
Low	High	Low	IPTW-MIMIC	0.887000	0.876071	0.897929
Low	Low	High	Baseline	0.837000	0.826071	0.847929
Low	Low	High	IPTW-MIMIC	0.884000	0.873071	0.894929
Low	Low	Low	Baseline	0.895000	0.884071	0.905929
Low	Low	Low	IPTW-MIMIC	0.911000	0.900071	0.921929

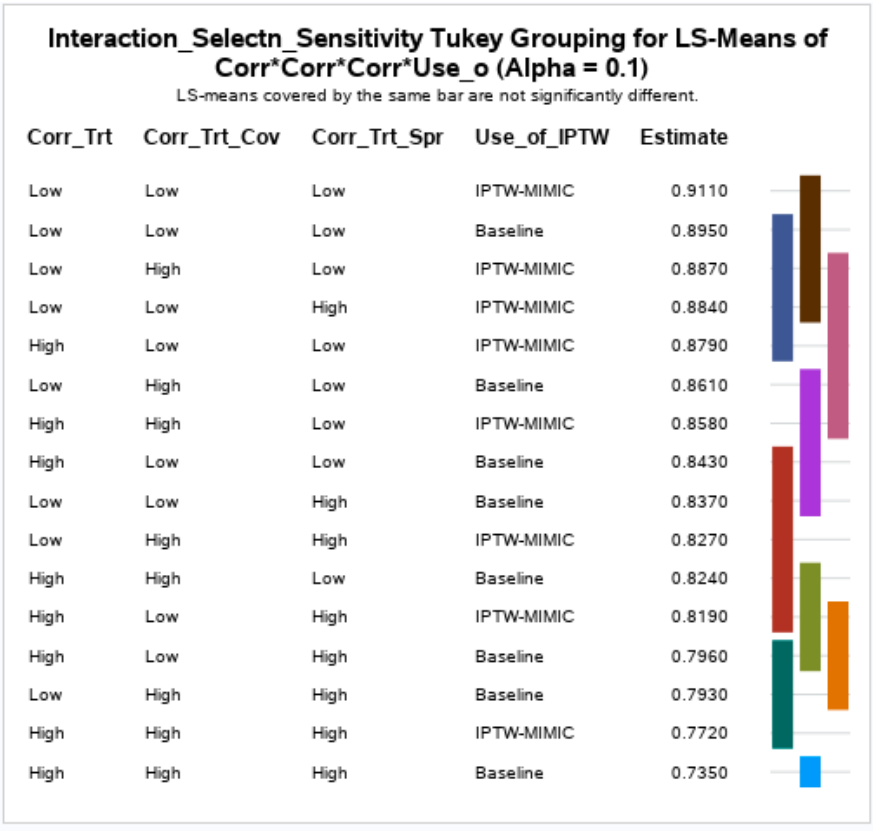


Figure 22: Tukey line plot for interaction selection sensitivity

The result of the means analysis across all the correlation combinations is shown in Table 24. We can see that the baseline and IPTW-MIMIC are statistically different with IPTW-MIMIC having higher average sensitivity than the baseline.

Table 24: Comparison of interaction selection sensitivity means with IPTW model form

Use_of_IPTW	Interaction_Selectn_Sensitivity LSMEAN	H0:LSMean1=LSMean2 Pr > |t|
Baseline	0.82300000	<.0001
IPTW-MIMIC	0.85462500	

Use_of_IPTW	Interaction_Selectn_Sensitivity LSMEAN	90% Confidence Limits	
Baseline	0.823000	0.819136	0.826864
IPTW-MIMIC	0.854625	0.850761	0.858489

Least Squares Means for Effect Use_of_IPTW				
i	j	Difference Between Means	Simultaneous 90% Confidence Limits for LSMean(i)-LSMean(j)	
1	2	-0.031625	-0.037089	-0.026161

A simultaneous Bonferroni comparison of contrasts is performed to identify the difference in the performance of the use of IPTW across the different correlation combinations. The set of contrasts for comparison is shown in Table 12. The results in Table 25 show that all the contrasts are statistically significant at a significance level of $\alpha = 0.1$, indicating that IPTW-MIMIC performs better than the baseline model in identifying the true interaction terms under the different correlation combinations.

Table 25: Simultaneous Bonferroni comparison results for interaction selection sensitivity

Corr_Trt	Corr_Trt_Cov	Corr_Trt_Spr	DF	Sum of Squares	Mean Square	F Value	Pr > F
High	High	High	1	0.068450	0.068450	15.52	<.0001
High	High	Low	1	0.057800	0.057800	13.11	0.0003
High	Low	High	1	0.026450	0.026450	6.00	0.0144
High	Low	Low	1	0.064800	0.064800	14.70	0.0001
Low	High	High	1	0.057800	0.057800	13.11	0.0003
Low	High	Low	1	0.033800	0.033800	7.67	0.0057
Low	Low	High	1	0.110450	0.110450	25.05	<.0001
Low	Low	Low	1	0.012800	0.012800	2.90	0.0886

3.5.4 Interaction Selection FDR

3.5.4.1 Preliminary Analysis

The boxplot for interaction selection FDR with the use of IPTW and correlation between treatments is shown in Figure 23. A lower value of FDR is preferred since it implies that there are few spurious interaction terms in the predicted model. We can observe that with both the baseline and IPTW-MIMIC, the average FDR is high when the correlation between treatments is high. The average FDR value is less with IPTW-MIMIC than the baseline. The variation in FDR values is almost similar with high and low correlation for the baseline model, but with IPTW-MIMIC the FDR values have higher variation at high correlation. These plots suggest that at high treatment correlation, more spurious interaction terms are picked up by both the modeling approaches, with IPTW-MIMIC performing better than the baseline.

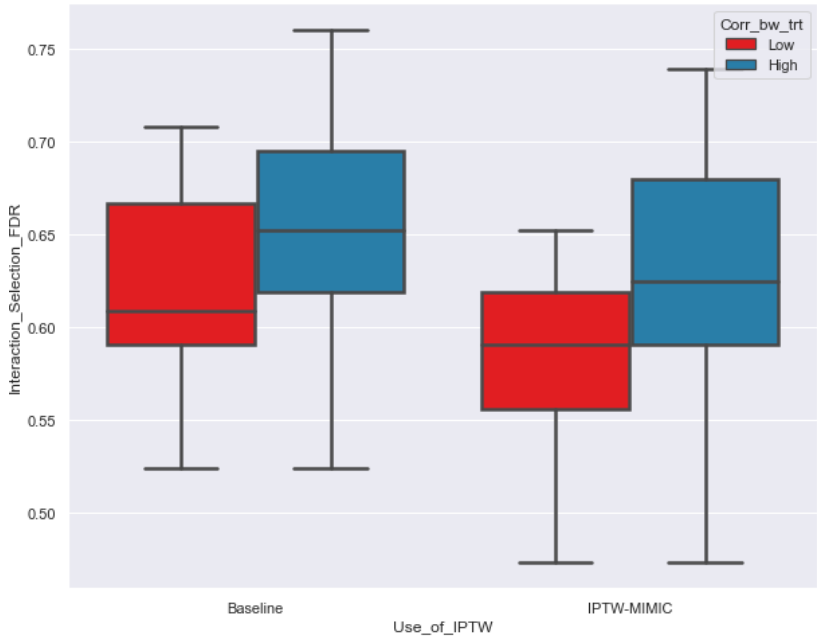


Figure 23: Interaction selection FDR against the use of IPTW for different correlation structures between treatments

The boxplot in Figure 24 shows the effect of the use of IPTW along with the correlation between treatments and covariates on interaction FDR. The average FDR and the variation in FDR are higher with

the baseline than the IPTW-MIMIC. The average FDR values are high at high correlation between treatments and covariates, which implies more spurious interaction terms are in the prediction model when there is high correlation between treatments and covariates. The IPTW-MIMIC performs better than the baseline.

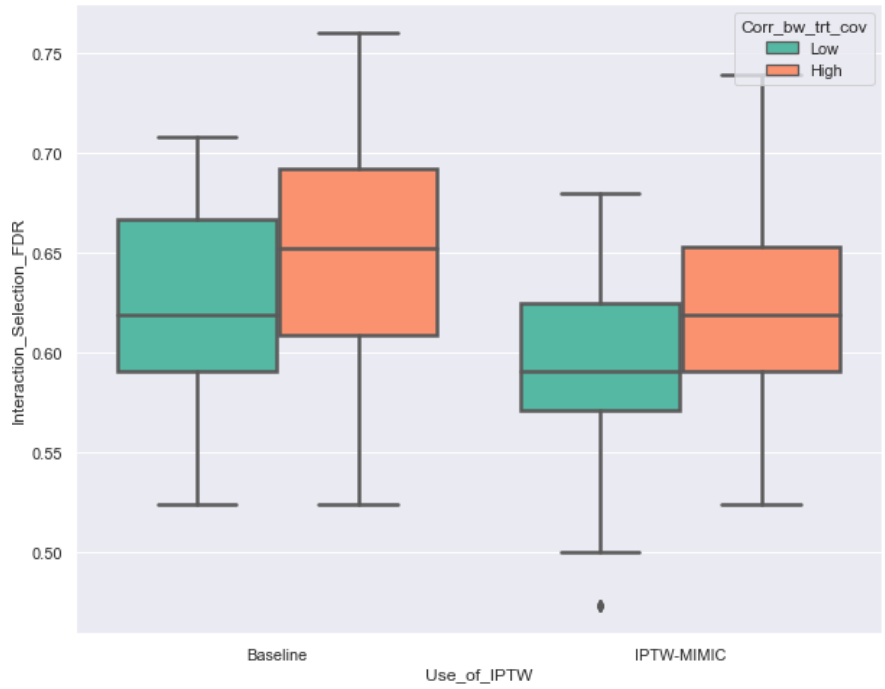


Figure 24: Interaction selection FDR against the use of IPTW for different correlation structures between treatments and covariates

The boxplot in Figure 25 shows the FDR with the use of IPTW under different correlation structures between treatments and spurious variables. The average FDR is higher at high correlation than at low correlation, and the variation in FDR is also higher at high correlation. The IPTW-MIMIC has lower average FDR values for a given correlation structure than the baseline. The performance of IPTW-MIMIC is better than the baseline at both low and high correlation. The impact of high correlation between treatments and spurious variables is more pronounced than the correlation between treatments and correlation between treatments and covariates, with the difference in average FDR values being higher than in Figures 23 and 24.

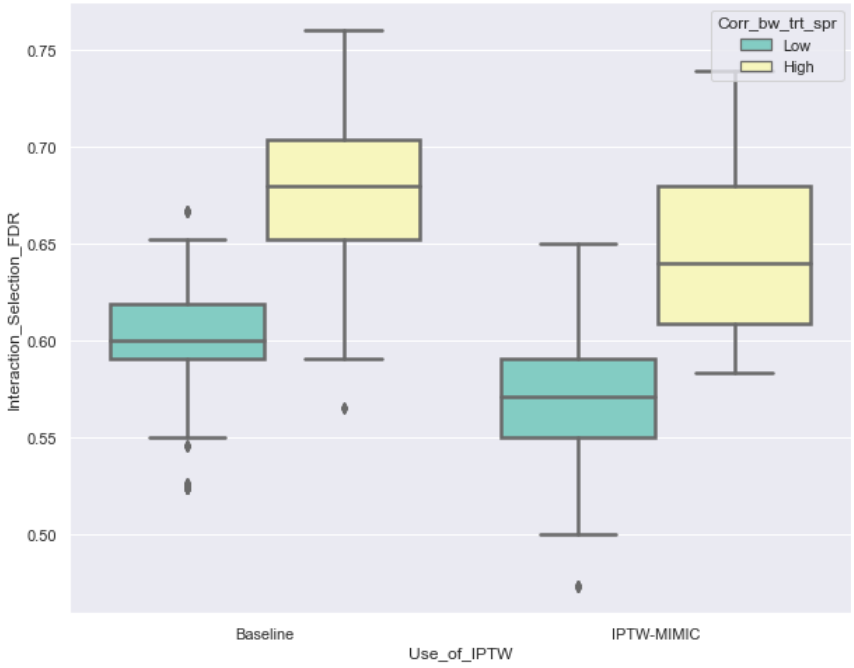


Figure 25: Interaction selection FDR against the use of IPTW for different correlation structures between treatments and spurious variables

3.5.4.2 Analysis of Variance (ANOVA)

The correlation between treatments, the correlation between treatments and covariates, the correlation between treatments and spurious variables and the use of IPTW are the four factors studied in this ANOVA. The results of a full factorial ANOVA using SAS are shown in Table 26, with the highlighted boxes indicating the significant effects at a significance level $\alpha = 0.1$. The main effects of all the factors were significant. All the two-factor interactions were significant except the interaction between the use of IPTW and the correlation between treatments and spurious variables. We can also observe this from Figure 21, where the difference in average FDR between high and low correlation is almost similar between the baseline and IPTW-MIMIC modeling approaches. All three-factor and four-factor interactions are also significant.

A Tukey multiple comparison considering all the factor combinations is conducted since the four-factor interaction effect was significant.

Table 26: ANOVA for interaction selection FDR

Source	DF	Anova SS	Mean Square	F Value	Pr > F
Corr_Trt	1	0.75191704	0.75191704	1372.28	<.0001
Corr_Trt_Cov	1	0.33748692	0.33748692	615.93	<.0001
Corr_Trt_Spr	1	2.32712437	2.32712437	4247.11	<.0001
Use_of_IPTW	1	0.33427392	0.33427392	610.07	<.0001
Corr_Trt*Corr_Trt_Cov	1	0.05980143	0.05980143	109.14	<.0001
Corr_Trt*Corr_Trt_Spr	1	0.04238167	0.04238167	77.35	<.0001
Corr_Trt*Use_of_IPTW	1	0.00225512	0.00225512	4.12	0.0427
Corr_Trt_Cov*Corr_Trt_Spr	1	0.00739022	0.00739022	13.49	0.0002
Corr_Trt_Cov*Use_of_IPTW	1	0.00771801	0.00771801	14.09	0.0002
Corr_Trt_Spr*Use_of_IPTW	1	0.00000469	0.00000469	0.01	0.9263
Corr_Trt*Corr_Trt_Cov*Corr_Trt_Spr	1	0.00715449	0.00715449	13.06	0.0003
Corr_Trt*Corr_Trt_Cov*Use_of_IPTW	1	0.00243515	0.00243515	4.44	0.0352
Corr_Trt*Corr_Trt_Spr*Use_of_IPTW	1	0.002944	0.002944	5.37	0.0206
Corr_Trt_Cov*Corr_Trt_Spr*Use_of_IPTW	1	0.00420714	0.00420714	7.68	0.0057
Corr_Trt*Corr_Trt_Cov*Corr_Trt_Spr*Use_of_IPTW	1	0.00235722	0.00235722	4.3	0.0382

The factor combinations for the Tukey multiple comparison are shown in Table 27, the p-value results of the Tukey analysis are shown in Table 28, and the 90% CI in Table 29 respectively. The highlighted cells in Table 28 indicate the factor combinations whose p-value is greater than the significance level $\alpha = 0.1$, and hence are not statistically different. In the Tukey line plot, shown in Figure 26, the factor combinations covered by the same-colored bar are not statistically different.

From the Tukey Line plot, we observe that the combination with high correlations across the groups with the baseline and IPTW-MIMIC are statistically different from all other combinations. The factor combination with IPTW-MIMIC and low correlation across the groups is also statistically different from all other combinations. There are other factor combinations that are not statistically different. Two such factor combinations of interest are the IPTW-MIMIC combination with high correlation between treatments and spurious variables and low correlation with the other two factors and the IPTW-MIMIC combination with low correlation between treatments and spurious variables and high correlation with the other two

factors. This shows that high correlation between treatments and spurious variables has a higher impact on FDR than the high correlation between treatments or between treatments and covariates.

Table 27: Tukey comparison factor combinations for interaction selection FDR

Adjustment for Multiple Comparisons: Tukey

Corr_Trt	Corr_Trt_Cov	Corr_Trt_Spr	Use_of_IPTW	Interaction_Selection_FDR LSMEAN	LSMEAN Number
High	High	High	Baseline	0.72332120	1
High	High	High	IPTW-MIMIC	0.70543650	2
High	High	Low	Baseline	0.63001482	3
High	High	Low	IPTW-MIMIC	0.60855185	4
High	Low	High	Baseline	0.67956410	5
High	Low	High	IPTW-MIMIC	0.64959058	6
High	Low	Low	Baseline	0.60494391	7
High	Low	Low	IPTW-MIMIC	0.56812987	8
Low	High	High	Baseline	0.66319710	9
Low	High	High	IPTW-MIMIC	0.62534788	10
Low	High	Low	Baseline	0.58865498	11
Low	High	Low	IPTW-MIMIC	0.56778945	12
Low	Low	High	Baseline	0.64256291	13
Low	Low	High	IPTW-MIMIC	0.61220457	14
Low	Low	Low	Baseline	0.57950037	15
Low	Low	Low	IPTW-MIMIC	0.54344298	16

Table 28: Tukey analysis p-value for interaction selection FDR

Least Squares Means for effect Corr*Corr*Corr*IPTW_
Pr > |t| for H0: LSMean(i)=LSMean(j)

Dependent Variable: Interaction_Effects_FDR

i/j	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1		<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
2	<.0001		<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
3	<.0001	<.0001		<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	0.9905	<.0001	<.0001	0.0147	<.0001	<.0001	<.0001
4	<.0001	<.0001	<.0001		<.0001	<.0001	0.9994	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	0.9994	<.0001	<.0001
5	<.0001	<.0001	<.0001	<.0001		<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
6	<.0001	<.0001	<.0001	<.0001	<.0001		<.0001	<.0001	0.0043	<.0001	<.0001	<.0001	0.7502	<.0001	<.0001	<.0001
7	<.0001	<.0001	<.0001	0.9994	<.0001	<.0001		<.0001	<.0001	<.0001	0.0001	<.0001	<.0001	0.7020	<.0001	<.0001
8	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001		<.0001	<.0001	<.0001	1.0000	<.0001	<.0001	0.0495	<.0001
9	<.0001	<.0001	<.0001	<.0001	<.0001	0.0043	<.0001	<.0001		<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
10	<.0001	<.0001	0.9905	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001		<.0001	<.0001	<.0001	0.0074	<.0001	<.0001
11	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	0.0001	<.0001	<.0001	<.0001		<.0001	<.0001	<.0001	0.2903	<.0001
12	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	1.0000	<.0001	<.0001	<.0001		<.0001	<.0001	0.0355	<.0001
13	<.0001	<.0001	0.0147	<.0001	<.0001	0.7502	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001		<.0001	<.0001	<.0001
14	<.0001	<.0001	<.0001	0.9994	<.0001	<.0001	0.7020	<.0001	<.0001	0.0074	<.0001	<.0001	<.0001		<.0001	<.0001
15	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	0.0495	<.0001	<.0001	0.2903	0.0355	<.0001	<.0001		<.0001
16	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	

Table 29: Tukey analysis 90% CI for interaction selection sensitivity

Corr_Trt	Corr_Trt_Cov	Corr_Trt_Spr	Use_of_IPTW	Interaction_Selection_FDR LSMEAN	90% Confidence Limits	
High	High	High	Baseline	0.723321	0.719469	0.727174
High	High	High	IPTW-MIMIC	0.705436	0.701584	0.709289
High	High	Low	Baseline	0.630015	0.626162	0.633867
High	High	Low	IPTW-MIMIC	0.608552	0.604699	0.612404
High	Low	High	Baseline	0.679564	0.675712	0.683417
High	Low	High	IPTW-MIMIC	0.649591	0.645738	0.653443
High	Low	Low	Baseline	0.604944	0.601091	0.608796
High	Low	Low	IPTW-MIMIC	0.568130	0.564277	0.571982
Low	High	High	Baseline	0.663197	0.659345	0.667050
Low	High	High	IPTW-MIMIC	0.625348	0.621495	0.629200
Low	High	Low	Baseline	0.588655	0.584802	0.592507
Low	High	Low	IPTW-MIMIC	0.567789	0.563937	0.571642
Low	Low	High	Baseline	0.642563	0.638710	0.646415
Low	Low	High	IPTW-MIMIC	0.612205	0.608352	0.616057
Low	Low	Low	Baseline	0.579500	0.575648	0.583353
Low	Low	Low	IPTW-MIMIC	0.543443	0.539590	0.547295

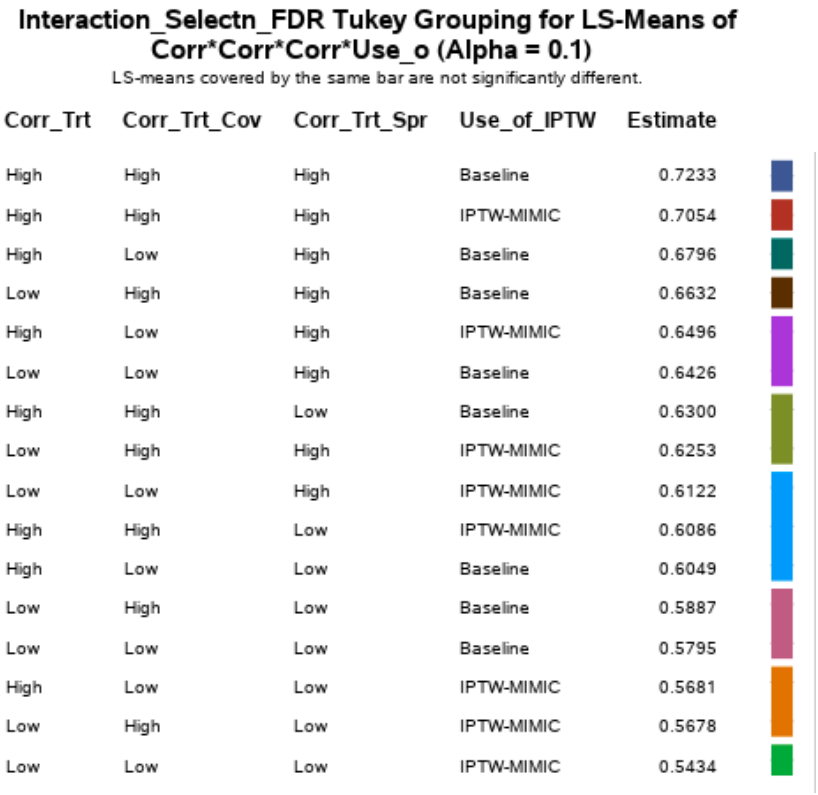


Figure 26: Tukey line plot for interaction selection FDR

A means analysis is performed to compare the IPTW-MIMIC and baseline. The result of the means analysis across all the correlation combinations is shown in Table 30. We see that IPTW-MIMIC and the baseline are statistically different with IPTW-MIMIC having a lower average FDR than baseline.

Table 30: Comparison of feature selection FDR means with IPTW model form

Use_of_IPTW	Interaction_Selection_FDR LSMEAN	H0:LSMean1=LSMean2 Pr > |t|
Baseline	0.63896992	<.0001
IPTW-MIMIC	0.61006171	

Use_of_IPTW	Interaction_Selection_FDR LSMEAN	90% Confidence Limits	
Baseline	0.638970	0.637608	0.640332
IPTW-MIMIC	0.610062	0.608700	0.611424

Least Squares Means for Effect Use_of_IPTW				
i	j	Difference Between Means	Simultaneous 90% Confidence Limits for LSMean(i)-LSMean(j)	
1	2	0.028908	0.026982	0.030834

A simultaneous Bonferroni comparison of contrasts is performed to identify the performance of the use of IPTW across the different correlation combinations. The set of contrasts for comparison is shown in Table 12, where each contrast is highlighted with the same color. The results in Table 31 show that all the contrasts are statistically significant at a significance level of $\alpha = 0.1$, indicating that IPTW-MIMIC performs better than the baseline by identifying fewer false interaction terms in the predicted model across the different correlation combinations.

Table 31: Simultaneous Bonferroni comparison results for feature selection FDR

Corr_Trt	Corr_Trt_Cov	Corr_Trt_Spr	DF	Sum of Squares	Mean Square	F Value	Pr > F
High	High	High	1	0.015993	0.015993	29.19	<.0001
High	High	Low	1	0.023033	0.023033	42.04	<.0001
High	Low	High	1	0.044921	0.044921	81.98	<.0001
High	Low	Low	1	0.067764	0.067764	123.67	<.0001
Low	High	High	1	0.071628	0.071628	130.72	<.0001
Low	High	Low	1	0.021769	0.021769	39.73	<.0001
Low	Low	High	1	0.046081	0.046081	84.10	<.0001
Low	Low	Low	1	0.065007	0.065007	118.64	<.0001

3.6 Conclusion and Future Work

The pain management simulation case study was conducted with the primary goal of evaluating the proposed HierNet-IPTW modeling framework that uses HierNet for model building and IPTW with MIMIC algorithm to address the time-varying confounding. The proposed method was compared against the baseline that uses HierNet for model building and does not address time-varying confounding. The evaluation was done on the feature and interaction selection metrics. The different correlation structures between the treatments, covariates, and spurious variables were simulated to mimic the complexities found in the actual pain management dataset. A full factorial 2^4 experiment was conducted.

The ANOVA was conducted with feature selection sensitivity, feature selection specificity, interaction selection sensitivity, and interaction selection FDR as the outcome measures. From the analysis, all the main effect factors were found to be significant for all the outcome measures. With feature selection sensitivity, interaction terms involving the use of IPTW, the correlation between treatments and correlation between treatments and covariates were found to be significant. With feature selection specificity, interaction selection sensitivity, and interaction selection FDR, 2-factor and 3-factor interaction terms were found to be significant that covered all the factors. Tukey analysis was performed to compare all the experimental combinations and the combinations that were not statistically different were highlighted.

A Bonferroni simultaneous comparison of contrasts was conducted with the contrasts defined to compare the use of IPTW-MIMIC against the baseline with all the possible combinations between the correlation factors. This analysis showed that the IPTW-MIMIC performed as well or better than the baseline in all the cases.

This experimental analysis of the proposed HierNet-IPTW modeling method on the pain management simulation data shows that it performs better than the baseline in correctly identifying the true and spurious features and interactions. The next step would be to implement this modeling method on the pain management dataset to build state transition and outcome models. These state transition models will be

incorporated into the optimization module. For future work, different stopping criteria can be considered in the model framework.

Chapter 4

Using LASSO based State Transition Model with Interactions in Multiple Objective Two-Stage Stochastic Programming for Adaptive Interdisciplinary Pain Management

Abstract

The Eugene McDermott Center for Pain Management at the University of Texas Southwestern Medical Center has a two-stage adaptive interdisciplinary pain management program for chronic pain. The program considers different pain outcome measures to quantify patient's pain, physical and mental health status.

In an adaptive treatment environment, the patient's current state influences the treatments recommended, which in turn affect the patient's future states leading to time-varying confounding and biased treatment estimates. Inverse Probability of Treatment Weighting (IPTW) [106] is one technique to address time-varying confounding. Interactions between the patient state and treatments play a major role in determining a personalized treatment plan for individual patients. In this research, we study the effect of including state-treatment interactions on the treatment optimization problem.

The treatment decision optimization is formulated as a multi-objective two-stage stochastic programming problem based on Iqbal et.al [27]. The objective is to minimize all the pain outcome measures and the treatment costs. The state transition models representing the transition of patient state, treatment, and outcome variables from stage 1 to stage 2 are modeled using the HierNet-IPTW approach. The state transition functions are non-convex and quadratic since they model the interaction effects. The optimization is formulated and solved as a Mixed Integer Quadratically Constrained Optimization (MIQCP) problem. The treatment recommendations from the proposed approach are compared against other approaches that do not model state-treatment interactions and address time varying confounding using IPTW.

4.1 Introduction

The Eugene McDermott Center for Pain Management at UT Southwestern Medical Center has an interdisciplinary pain management program for chronic pain. Lin et al [22] modeled this program as a two-stage adaptive treatment framework as shown in Figure 27.

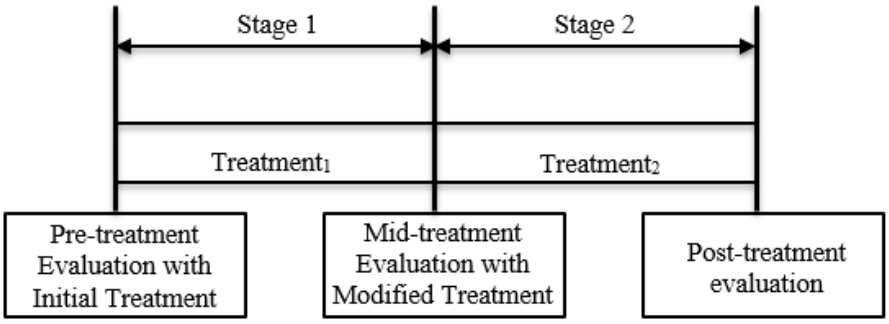


Figure 27: Two-stage interdisciplinary pain management program [22]

The observational dataset has 294 patient observations, with 62 state variables, 14 stage 1 and 13 stage 2 treatment decision variables, and 5 pain outcome measures. A detailed description of these variables is provided in Appendix III. The pain outcome measures, and treatment variables used in the study are shown in Table 32 and Table 33, respectively. The multiple pain outcome measures help caregivers evaluate the patients on several different parameters. These pain outcome measures provide an insight into the physical, mental, and general health profile of the patients. The treatment decision variables include a mix of pharmaceutical and procedural interventions.

Table 32: Pain Outcome Measures

Pain Outcome Measure	**Description**
Oswestry Pain Disability Index (OSW)	A measure of functional disability due to pain [18]
Pain Drawing Analogue (PDA)	An analog scale of 0-10, with 0 corresponding to no pain and 10 corresponding to the worst possible pain [19]
Beck Depression Inventory (BDI)	Used to measure the severity of depression [20]
Short Form Survey Physical Component Score (SF36pcs)	General health status profile surveys designed to measure the physical health status of the patient [21]
Short Form Survey Mental Component Score (SF36mcs)	General health status profile surveys designed to measure the mental health status of the patient [21]

Table 33: Treatment Decision Variables

Treatments	Treatment Type	Description
RxGr1	Pharmaceutical	Tramadol
RxGr2	Pharmaceutical	NSAID
RxGr3	Pharmaceutical	Narcotic
RxGr4	Pharmaceutical	Muscle Relaxant
RxGr5	Pharmaceutical	Anti-depressant
RxGr6	Pharmaceutical	Tranquilizer
RxGr7	Pharmaceutical	Sleeping Pill
RxGr8	Pharmaceutical	Other
ProcGr1	Procedural	Injection
ProcGr2	Procedural	Block Procedure
ProcGr4	Procedural	Stimulation Procedure
ProcGr9	Procedural	Cognitive Behavioral Therapy
ProcGr10	Procedural	Physical Therapy
ProcGr11	Procedural	Additional Procedures

The main objective of the pain management research is to use this observational data in an adaptive decision framework to identify an optimal treatment regime [22]. A summary of the research work done on the pain management program is provided in Table 34. The proposed research work is highlighted in the table.

Lin et al. [22] created an optimization framework based on approximate dynamic programming (ADP) with linear regression to model the state transitions. The state transition models are constructed empirically with the future value function approximated using state space discretization based on Latin hypercube design [22]. The recommended treatment regime minimized adverse patient pain outcomes along with treatment costs. The adaptive and sequential nature of the pain management program introduces time-varying confounding, where the treatment effects are confounded by past treatments and patient state variables. LeBoulluec et al. [25] extended the IPTW framework to address this time-varying confounding in a multiple treatment setting. Ohol [26] further extended the IPTW method to consider correlated

treatments. Farahani et al. [152] used LASSO-based regularization along with IPTW to perform feature selection.

Table 34: Pain Management Research Summary

	State Transition Modeling				
Paper	**Model type**	**Interactions**	**Feature selection**	**IPTW**	**Optimization**
Lin et al. [22]	Linear model	State treatment interactions	Stepwise least squares	No	SDP
LeBoulluec et al. [25]	Linear model	State treatment interactions	Stepwise least squares	Derived weights	No
Ohol [26]	Linear model	No	No	Derived weights	No
Farahani et al. [152]	Linear model	No	LASSO regularization	Used Ohol weights	No
Wang et al. [28]	Linear model	State treatment interactions	Stepwise least squares	Used LeBoulluec's weights	Two-stage SP
Iqbal et al. [27]	Piecewise Linear Network	Network interactions	Piecewise Linear Network	Used LeBoulluec's weights	Two-stage SP
Viswanatha	Linear model	State treatment interactions	LASSO regularization	Derived weights	Two-stage SP

Wang et al. [28] and Iqbal et.al [27] formulated the optimization as a two-stage stochastic programming problem. Wang et al. [28] used weighted least squares method to develop the state transition models using the IPTW weights from LeBoulluec et al. [25]. The state transition constraints in the optimization were non-convex as they modeled the state treatment interaction terms. A linearization method using a piecewise linear function was proposed to approximate the non-convex constraints and formulate the optimization problem as an approximated mixed integer linear problem (MILP). The objective of the optimization was to minimize the pain outcome measure, OSW while also penalizing excessive treatment costs under the state transition function and treatment interaction constraints. The optimization results from the approximate MILP were compared with the solutions obtained from the original mixed integer nonlinear problem (MINLP) formulation without linearizing the interaction constraints.

Iqbal et al.[27] formulated a multi-objective two-stage stochastic programming optimization problem using piecewise linear networks (PLN) to build the state transition functions. The objective is to minimize the multiple pain outcome measures along with treatment costs. A survey among caregivers was conducted to identify the relation between the different pain outcome measures and a convex quadratic programming approach was used to obtain weights to penalize the different pain measures. An equivalent MILP model was used to solve the optimization problem.

Iqbal et al. [27] was the only work that considered multi-objective optimization and included IPTW weights while building the state transition models. The state transition models did not include the state treatment interaction effects but had network interactions.

In this paper, we propose to use the LASSO-based HierNet-IPTW modeling method to develop state transition models with state treatment interactions. We use these state transition models in the multi-objective optimization framework based on Iqbal et.al [27] and study the treatment recommendation patterns from the proposed approach. The state transition functions are non-convex as they include interaction terms. Instead of linearizing the interaction terms and using an approximate MILP model, we use a Mixed Integer Quadratically Constrained Program (MIQCP) to formulate and solve the optimization problem.

The main objective of this research is to show the advantages of modeling interaction terms in an adaptive treatment setting, where the goal is optimal treatment decision-making rather than optimal prediction [155]. We compare the treatment decisions obtained by the proposed method against the treatments recommended using a linear state transition model without interactions. We also show the advantage of using IPTW to address time-varying confounding while building models by comparing the treatments recommended by the proposed method against the treatments obtained using state transition models with interactions modeled without using IPTW.

4.2 Stochastic Programming Formulation

The objective of the research is to find the optimal treatment strategy in stage 1 and stage 2 that will minimize the patient's multiple pain outcome measures and treatment cost. The objective function has two components [27]: the penalty function of pain outcomes and a cost function associated with treatment usage. The objective function ensures that the patient pain outcomes are reduced while avoiding overmedication. The optimization is subject to treatment interaction and state transition constraints [28].

The general two-stage stochastic problem formulation from Iqbal et al. [27] is shown below:

$$min \sum_{i \in N} \sum_{j \in N_i} E(\, P_{u_{ij}}(\bar{Y}_{i2}(\varepsilon_1, \varepsilon_{i2}))) + \, \rho(\sum C(x_1) + \sum E(C(x_2(\varepsilon_1)))) \tag{14.1}$$

$$subject\ to\colon Y_{i1}(\varepsilon_{i1}) = \, h_{i1}(s_1, x_1, \varepsilon_{i1}) \quad \forall i \in N \tag{14.2}$$

$$Y_{i2}(\varepsilon_1, \varepsilon_{i2}) = \, h_{i2}(s_2(\varepsilon_1), x_2(\varepsilon_1), \varepsilon_{i2}) \quad \forall i \in N \tag{14.3}$$

$$\begin{pmatrix} \bar{Y}_{i1}(\varepsilon_{i1}) = \max\,(0, Y_{i1}(\varepsilon_{i1})) \\ \bar{Y}_{i2}(\varepsilon_{1,}\varepsilon_{i2}) = \max\,(0, Y_{i2}(\varepsilon_1, \varepsilon_{i2})) \end{pmatrix} \quad \forall i \in N \tag{14.4}$$

$$\begin{pmatrix} x_1^p \times \, x_1^q = 0 \\ x_2^p(\varepsilon_1) \times \, x_2^q(\varepsilon_1) = 0 \end{pmatrix} \quad \forall x^p, x^q \in \Lambda \tag{14.5}$$

$$s_2(\varepsilon_1) = \, [s_1, x_1, Y_1(\varepsilon_1)] \tag{14.6}$$

$$x_1 \in \, \Gamma_1, x_2(\varepsilon_1) \in \, \Gamma_2 \tag{14.7}$$

where u_{ij} is the penalty weights of pain outcome $i \in N$ for level $j \in N_i$. $P_{u_{ij}}$ is the penalty function on different levels of the multiple pain outcome measures. $C()$ is the treatment cost function with ρ as the treatment cost coefficient. The penalty weights u_{ij} are from Iqbal et al. [27] and the treatment cost function is from Wang et al. [28]. The treatment cost coefficient ρ is set to 0.05 [27] and is selected so that the cost function does not dominate the pain outcome.

The variables $Y_{i1}(\varepsilon_{i1})$ and $Y_{i2}(\varepsilon_{i1}, \varepsilon_{i2})$ are the pain outcome measures i in stage 1 and stage 2 with uncertainties ε_{i1} and ε_{i2}. $Y_1(\varepsilon_1)$ and $\varepsilon_1$are the vectors associated with $Y_{i1}(\varepsilon_{i1})$ and $\varepsilon_{i1}, \forall i \in N$ respectively. x_t^p are the treatment decision variables p in stage t with the feasible decision space Γ_t, and s_t are the state variables in stage t. The set Λ includes treatment interaction restrictions.

The state transition functions for the pain outcome measures are modeled as the constraint set shown in Equations (14.2) and (14.3). h_{i1} and h_{i2} are the state transition models for pain outcome measure i at stage 1 and stage 2 respectively. Equation (14.4) uses the truncating variables $\bar{Y}_{i1}(\varepsilon_{i1})$ and $\bar{Y}_{i2}(\varepsilon_{i1}, \varepsilon_{i2})$ to ensure that the pain outcome measures are non-negative. Equation (14.5) ensures that treatments with adverse interaction effects are not prescribed to patients together. The list of treatments with adverse interaction effects is included in Appendix IV. The state variables in stage 2 include the state variables, treatment variables and pain outcome measures from stage 1, along with the stage 1 uncertainties ε_1 . This is shown in Equation (14.6). Equation (14.7) ensures that the treatment decision variables are in the feasible region.

Iqbal et al.[27] used Piecewise Linear Network (PLN) to model the state transition functions and solved an MILP to optimize treatment. Wang et al. [28] used stepwise regression to model the state transition function with interactions. The interaction terms were linearized using a piecewise linear function and an approximate MILP was solved. We propose to use the HierNet-IPTW approach to build the state transition function with interactions.

The state transition functions h_{i1} and h_{i2}, developed using the proposed HierNet-IPTW approach are quadratic and non-convex since they model the state treatment interaction effects. Instead of linearizing these quadratic non-convex constraints and solving an approximate MILP, we formulate the optimization as a Mixed Integer Quadratically Constrained program (MIQCP). However, we also note that when the state transition models have no interactions, either because interaction terms are ignored in model development or the model finds no statistically significant ones, then the optimization formulation is an MILP.

In two-stage stochastic programming, uncertainty is represented using discrete sampled scenarios. Iqbal et al.[27] ran experiments to compare different sample sizes and selected a sample size of 25 for each stage. We use a sample size of 25 per stage in this research as well.

The optimal stage 1 and stage 2 treatments recommended from the following three optimization approaches are compared:

1. **2SP IQ-IPTW**: The proposed MIQCP optimization approach with HierNet-based state transition models with interactions and with IPTW.
2. **2SP IQ**: The MIQCP optimization approach with HierNet-based state transition models with interactions and without IPTW.
3. **2SP AL-IPTW**: The MILP optimization approach with LASSO-IPTW-based linear state transition models without interactions and with IPTW.

The purpose of the analysis is to compare the difference in treatment usage between models with interactions using IPTW and without IPTW, and between models using IPTW with interactions and without interactions.

The stage 1 and stage 2 state transition models for the 5 pain outcome measures PDA, OSW, BDI, SF-36pcs and SF36-mcs are developed using the above three modeling methods. The MIQCP and MILP formulation is done in AMPL and the Gurobi solver is used to solve the optimization problem. These models are included in Appendix III.

4.3 Treatment Usage Analysis

In this section, we analyze the stage 1 and stage 2 treatments selected by the optimization approaches.

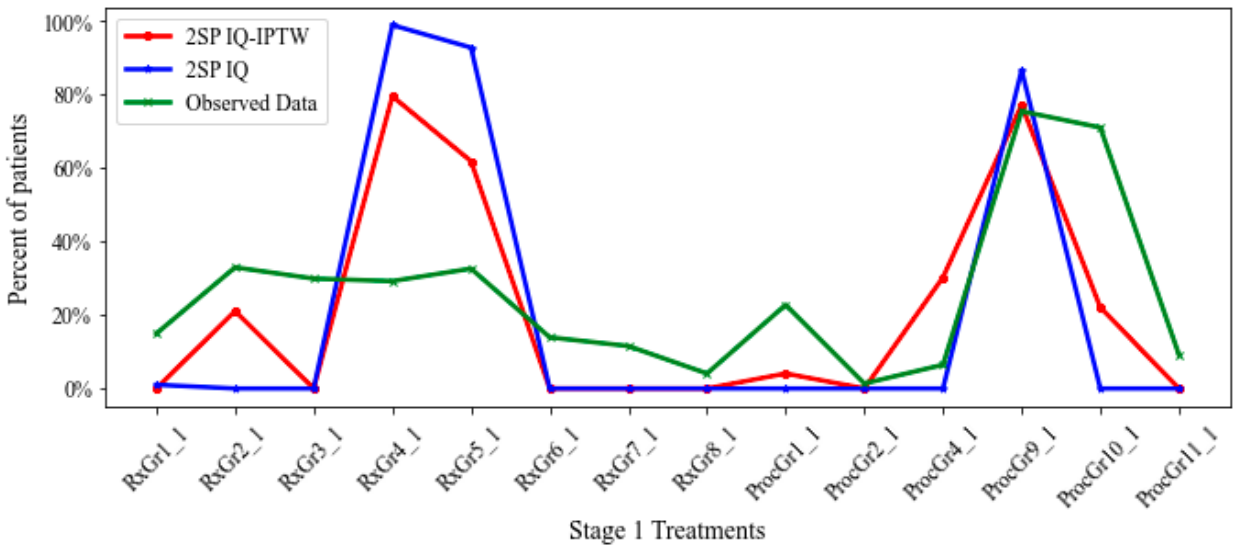
4.3.1 Stage 1 Treatment Analysis

The stage 1 treatments selected by the optimization approaches are compared against each other and with the observed dataset.

4.3.1.1 Stage 1 2SP IQ-IPTW vs 2SP IQ Treatment Analysis

The stage 1 treatment usages from 2SP IQ-IPTW, 2SP IQ, and observed data are shown in Figure 28. The most used treatment in the observed data is ProcGr9_1 (Cognitive Behavioral Therapy), recommended to 76% of the patients. ProcGr9_1 is suggested to 77% of the patients in 2SP IQ-IPTW and 87% of the patients in 2SP IQ. The second most used treatment in the observed data, ProcGr10_1 (Physical Therapy), prescribed to 71% of the patients is not selected by 2SP IQ but is recommended to 22% of the patients in 2SP IQ-IPTW. The pharmaceutical treatments RxGr4 (Muscle Relaxant) and RxGr5 (Anti-depressant) are the most used treatments in 2SP IQ, recommended to 99% and 93% of the patients respectively. RxGr4 is the most used treatment in 2SP IQ-IPTW, recommended to 80% of the patients. In the observed dataset, pharmaceutical treatments RxGr2 (NSAID), RxGr3 (Narcotic), RxGr4, and RxGr5 are recommended to around 30% of the patients, while in 2SP IQ-IPTW, RxGr2 and RxGr5 are recommended to 21%, and 62% of the patients respectively and RxGr3 is not recommended. The procedural treatments ProcGr1 (Injections) and ProcGr4 (Stimulation Procedure) are recommended to 23% and 6% of the patients in the observed dataset, while they are recommended to 4% and 30% of patients in 2SP IQ-IPTW but are not used in 2SP IQ.

The main difference between 2SP IQ-IPTW and 2SP IQ is the number of different treatments recommended by the model. The majority of patients are recommended the treatments RxGr4, RxGr5, and ProcGr9 by 2SP IQ, while 2SP IQ-IPTW recommends a wide range of treatments that includes more procedures like ProcGr1, ProcGr4, ProcGr9, and ProcGr10.

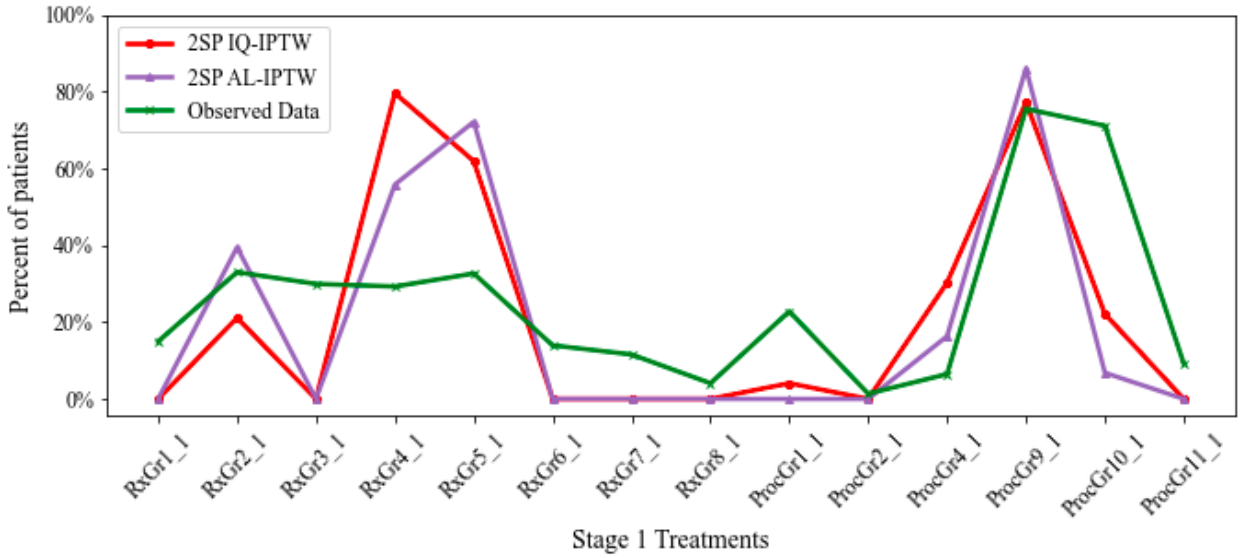


Treatments-Stage 1	RxGr1	RxGr2	RxGr3	RxGr4	RxGr5	RxGr6	RxGr7	RxGr8	ProcGr1	ProcGr2	ProcGr4	ProcGr9	ProcGr10	ProcGr11
2SP IQ-IPTW	0%	21%	0%	80%	62%	0%	0%	0%	4%	0%	30%	77%	22%	0%
2SP IQ	1%	0%	0%	99%	93%	0%	0%	0%	0%	0%	0%	87%	0%	0%
Observed Data	15%	33%	30%	29%	33%	14%	12%	4%	23%	1%	6%	76%	71%	9%

Figure 28: Stage 1 Treatment Usage 2SP IQ-IPTW, 2SP IQ, and observed data

4.3.1.2 Stage 1 2SP IQ-IPTW vs 2SP AL-IPTW Treatment Analysis

The stage 1 treatment usage from 2SP IQ-IPTW, 2SP AL-IPTW, and observed data are shown in Figure 29. The treatments recommended by the 2SP IQ-IPTW and 2SP AL-IPTW are similar but with differences in the percentage of patients recommended. 2SP AL-IPTW recommends RxGr2 (NSAID), RxGr5 (Anti-Depressant), and ProcGr9 (Cognitive Behavioral Therapy) to more patients than 2SP IQ-IPTW while the treatments RxGr4 (Muscle Relaxant), ProcGr4 (Stimulation Procedure) and ProcGr10 (Physical Therapy) are recommended to more patients in 2SP IQ-IPTW. The treatment ProcGr1 (Injection) is recommended to 4% of the patients in 2SP IQ-IPTW while it is not used in 2SP AL-IPTW. From this analysis, we observe that more procedural treatments are recommended by the 2SP IQ-IPTW than 2SP AL-IPTW.



Treatments-Stage 1	RxGr1	RxGr2	RxGr3	RxGr4	RxGr5	RxGr6	RxGr7	RxGr8	ProcGr1	ProcGr2	ProcGr4	ProcGr9	ProcGr10	ProcGr11
2SP IQ-IPTW	0%	21%	0%	80%	62%	0%	0%	0%	4%	0%	30%	77%	22%	0%
2SP AL-IPTW	0%	39%	0%	56%	72%	0%	0%	0%	0%	0%	16%	86%	7%	0%
Observed Data	15%	33%	30%	29%	33%	14%	12%	4%	23%	1%	6%	76%	71%	9%

Figure 29: Stage 1 Treatment Usage 2SP IQ-IPTW, 2SP AL-IPTW, and observed data

4.3.1.3 Stage 1 Treatment Analysis between low and high IPTW weight patients

The objective of incorporating IPTW in the model framework is to identify the rare patient instances in the dataset, which will have higher IPTW weight. The patient dataset is divided into two groups, high IPTW weights, and low IPTW weights. We compare the treatment usage between the two groups to better understand the difference between the three approaches. The dataset has 294 patient observations, 55 of them have IPTW weights greater than 1 and are grouped as high IPTW weight patients while the remaining 239 patients are grouped as low IPTW weight patients.

The treatment recommendation pattern with 2SP IQ is shown in Figure 30. The pharmaceutical treatment usage pattern is similar between low and high IPTW patients while the procedural treatment ProcGr9 (Cognitive Behavioral Therapy) is recommended to 89% of low weight and 76% of high weight patients respectively. The lack of difference between the pharmaceutical treatment usage patterns shows

that the HierNet model without IPTW does not identify the rare patient instances and follows the same treatment recommendation patterns for all the patients.

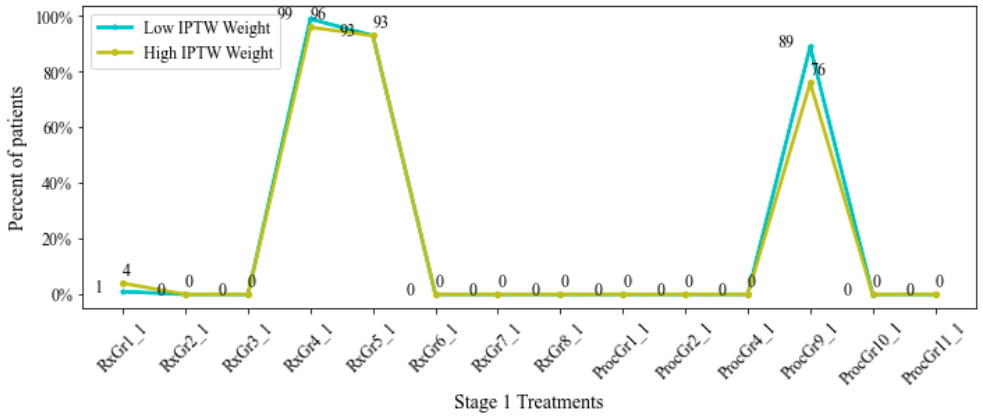


Figure 30: Stage 1 Treatment Usage 2SP IQ for Low and High IPTW weight patients

The treatment usage by 2SP AL-IPTW in Figure 31 does not show major differences in treatment assignment between low and high IPTW patients. The treatments RxGr2 (NSAID), RxGr4 (Muscle Relaxant), RxGr5 (Anti-depressant), and ProcGr10 (Physical Therapy) are prescribed slightly more often to high weight patients while the procedural treatments ProcGr4 (Stimulation procedure) and ProcGr9 (Cognitive Behavioral Therapy) are prescribed more to low weight patients.

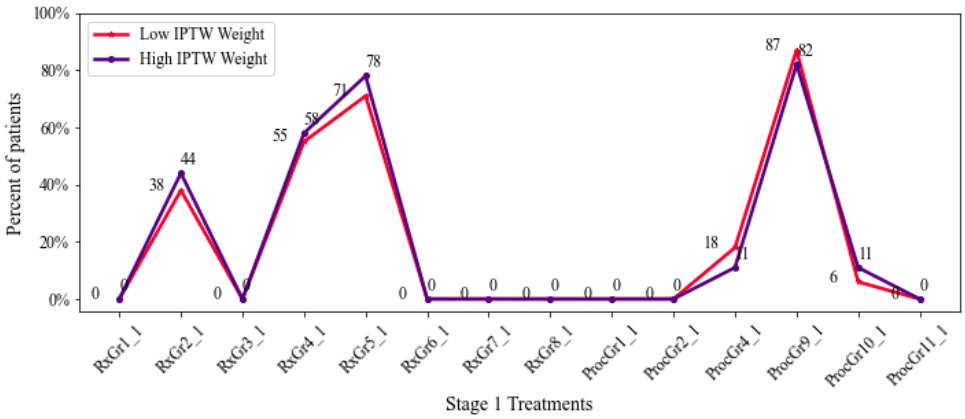


Figure 31: Stage 1 Treatment Usage 2SP AL-IPTW for Low and High IPTW weight patients

The 2SP IQ-IPTW treatment recommendation pattern in Figure 32 shows that certain treatments are prescribed to a greater percentage of high weight patients than to low weight patients. The procedural treatment ProcGr10 (Physical Therapy) is recommended to 53% of high IPTW weight patients while it is suggested to only 15% of low weight patients. The procedural treatments ProcGr1 (Injection) and ProcGr4 (Stimulation Procedure) are also recommended more often to high weight patients. The treatments RxGr4 (Muscle Relaxant) and ProcGr9 (Cognitive Behavioral Therapy) are recommended similarly to around 80% of both low and high weight patients. The pharmaceutical treatments RxGr2 (NSAID) and RxGr5 (Anti-depressant) are recommended to 36% and 69% of high weight patients, while it is recommended to 18% and 60% of low weight patients, respectively.

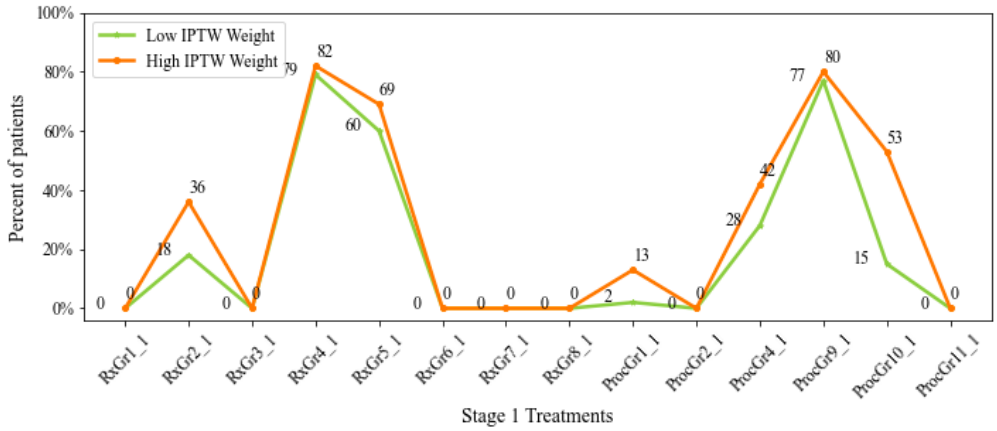


Figure 32: Stage 1 Treatment Usage 2SP HierNet-IPTW for Low and High IPTW weight patients

The comparison of treatment assignment to low and high IPTW patients between 2SP IQ-IPTW, 2SP AL-IPTW, and 2SP IQ is shown in Figures 33(a) and 33(b) respectively. The 2SP IQ-IPTW approach recommends more treatments to high IPTW weight patients than the 2SP AL-IPTW and 2SP IQ approaches. The treatment RxGr2 (NSAID) is recommended by 2SP IQ-IPTW to a smaller percentage of low weight patients than by 2 SP AL-IPTW, but its recommendation to high weight patients is increased by 2SP IQ-IPTW. The treatments ProcGr1 (Injection), ProcGr4 (Stimulation Procedure), and ProcGr10 (Physical Therapy) are recommended to similar percentages of low weight patients by 2SP AL-IPTW and 2SP IQ-IPTW, while we see a significant increase in their recommendation to high weight patients by 2SP IQ-

IPTW. This analysis shows that accounting for both time-varying confounding using IPTW and the interaction effects while building the state transition models results in the optimal treatment allocation being more adaptive to rare patients in stage 1.

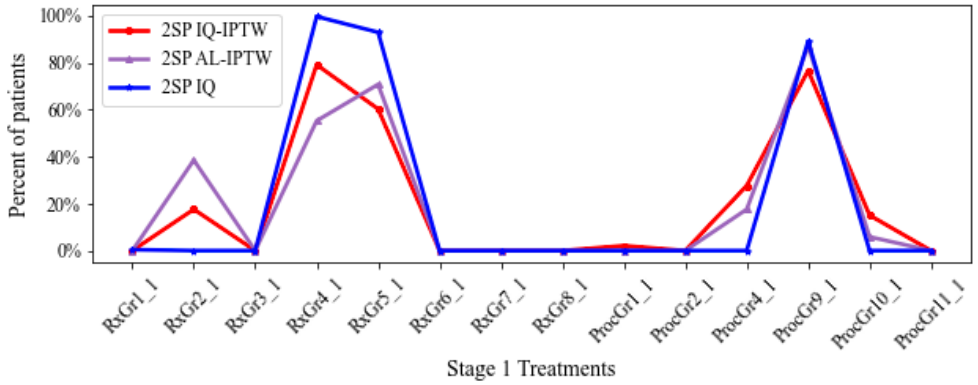


Figure 33(a): Stage 1 treatment assignment comparison for low IPTW weight patients

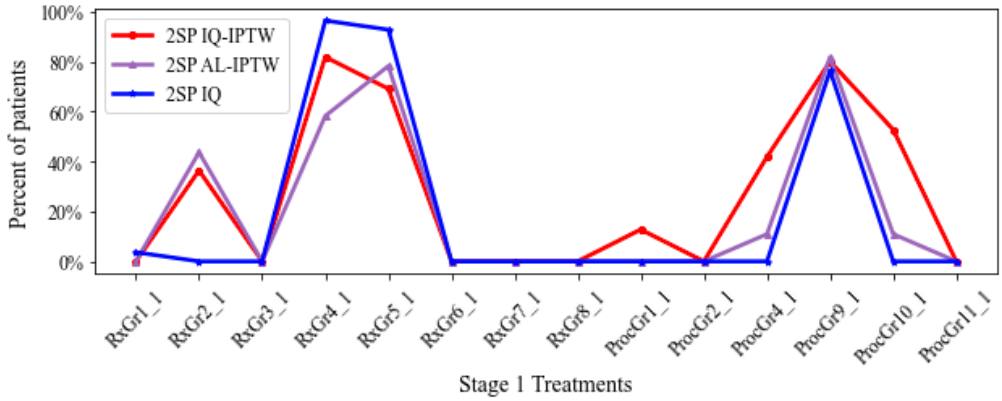


Figure 33(b): Stage 1 treatment assignment comparison for high IPTW weight patients

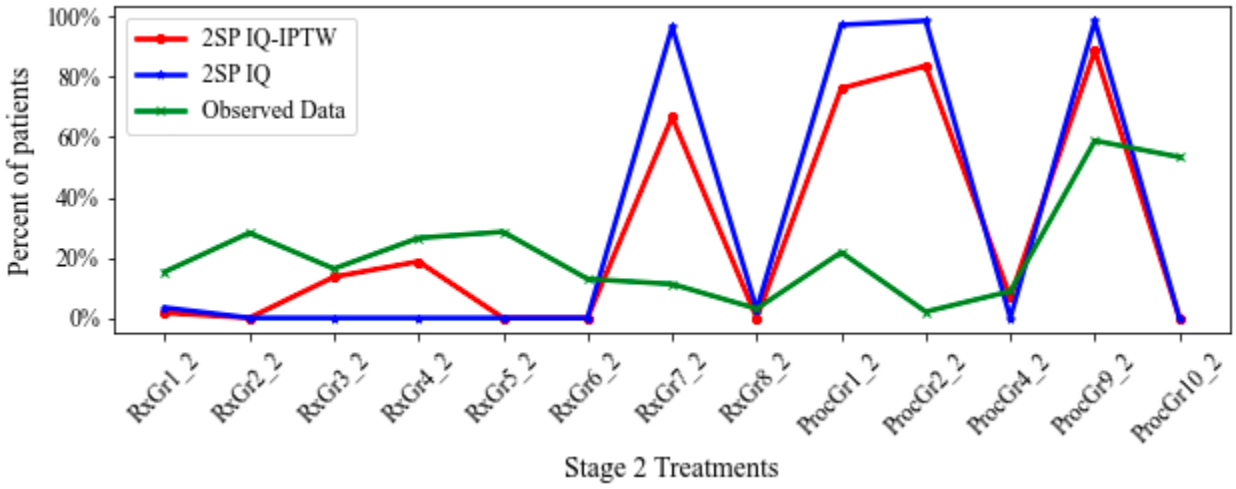
4.3.2 Stage 2 Treatment Analysis

The stage 2 treatments selected by the optimization approaches are analyzed in this section.

4.3.2.1 Stage 2 2SP IQ-IPTW vs 2SP IQ Treatment Analysis

The stage 2 treatment usage from 2SP IQ-IPTW, 2SP IQ, and observed data are shown in Figure 34. The procedural treatments ProcGr9 (Cognitive Behavioral Therapy) and ProcGr10 (Physical Therapy) are the most recommended treatments in the observed dataset, prescribed to 59% and 53% of the patients

respectively. ProcGr9 is recommended to 99% of the patients in 2SP IQ and 89% of the patients in 2SP IQ-IPTW, while ProcGr10 is not recommended in either. The 2SP IQ approach recommends treatments RxGr7 (Sleeping Pills), ProcGr1 (Injection), and ProcGr2 (Block Procedure) to 97%, 97%, and 99% of the patients, while 2SP IQ-IPTW recommends these treatments to 67%, 76% and 84% of the patients respectively. In the observed dataset, RxGr7 is recommended to 11% of the patients, while ProcGr1 and ProcGr2 are prescribed to 22% and 2% of the patients respectively. This analysis shows that the 2SP IQ approach recommends a few select treatments to the majority of the patients, while the 2SP IQ-IPTW approach recommends a wide range of treatments in stage 2.



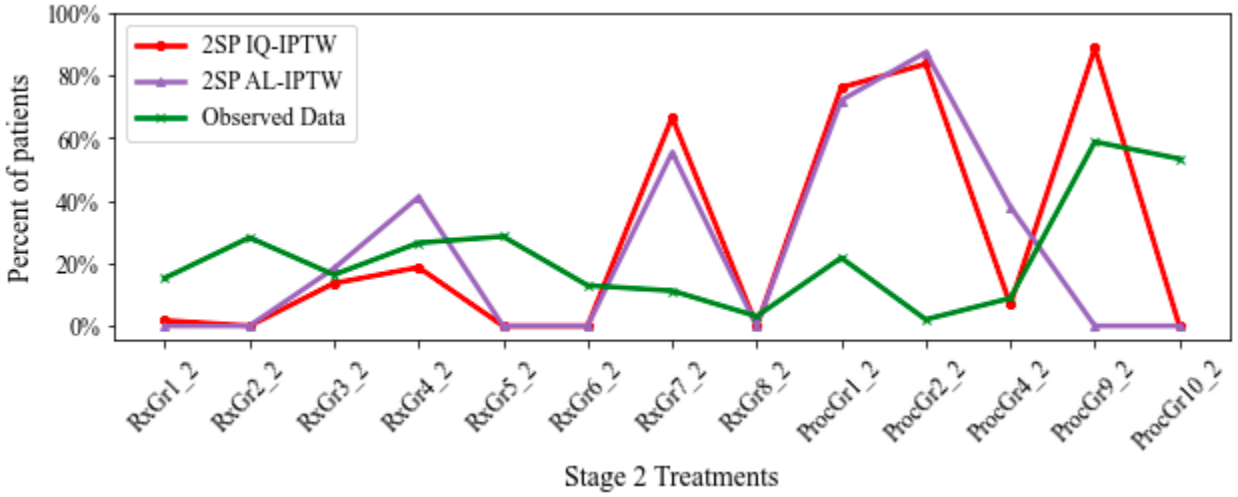
Treatments-Stage 2	RxGr1	RxGr2	RxGr3	RxGr4	RxGr5	RxGr6	RxGr7	RxGr8	ProcGr1	ProcGr2	ProcGr4	ProcGr9	ProcGr10
2SP IQ-IPTW	2%	0%	14%	19%	0%	0%	67%	0%	76%	84%	7%	89%	0%
2SP IQ	3%	0%	0%	0%	0%	0%	97%	3%	97%	99%	0%	99%	0%
Observed Data	15%	28%	16%	27%	29%	13%	11%	3%	22%	2%	9%	59%	53%

Figure 34: Stage 2 Treatment Usage 2SP IQ-IPTW, 2SP IQ, and observed data

4.3.2.2 Stage 2 2SP IQ-IPTW vs 2SP AL-IPTW Treatment Analysis

The stage 2 treatment recommendations from 2SP IQ-IPTW, 2SP AL-IPTW, and observed data are shown in Figure 35. The major difference between the two approaches is in the recommendation of RxGr4 (Muscle Relaxant), ProcGr4 (Stimulation Procedure), and ProcGr9 (Cognitive Behavioral Therapy). While 2SP AL-IPTW recommends the treatments RxGr4 and ProcGr4 to 41% and 38% of the patients, 2SP IQ-IPTW uses

them in 19% and 7% of the patients respectively. ProcGr9 is not recommended to any patients in 2SP AL-IPTW while it is recommended to 89% of the patients in 2SP IQ-IPTW. The treatments RxGr3, RxGr7, ProcGr1, and ProcGr2 followed similar recommendation patterns with both approaches.



Treatments-Stage 2	RxGr1	RxGr2	RxGr3	RxGr4	RxGr5	RxGr6	RxGr7	RxGr8	ProcGr1	ProcGr2	ProcGr4	ProcGr9	ProcGr10
2SP IQ-IPTW	2%	0%	14%	19%	0%	0%	67%	0%	76%	84%	7%	89%	0%
2SP AL-IPTW	0%	0%	18%	41%	0%	0%	55%	0%	72%	87%	38%	0%	0%
Observed Data	15%	28%	16%	27%	29%	13%	11%	3%	22%	2%	9%	59%	53%

Figure 35: Stage 2 Treatment Usage 2SP IQ-IPTW, 2SP AL-IPTW, and observed data

4.3.2.3 Stage 2 Treatment Analysis between low and high IPTW weight patients

The treatment recommendation patterns with 2SP IQ for low and high weight patients are similar as shown in Figure 36. The treatments RxGr7 (Sleeping Pill), ProcGr1 (Injection), and ProcGr9 (Cognitive Behavioral Therapy) are recommended slightly more often to low weight patients than high weight patients while RxGr1 (Tramadol) and RxGr8 (Other) are recommended marginally more to high weight patients. The treatment ProcGr4 (Stimulation Therapy) is similarly recommended to low and high weight patients.

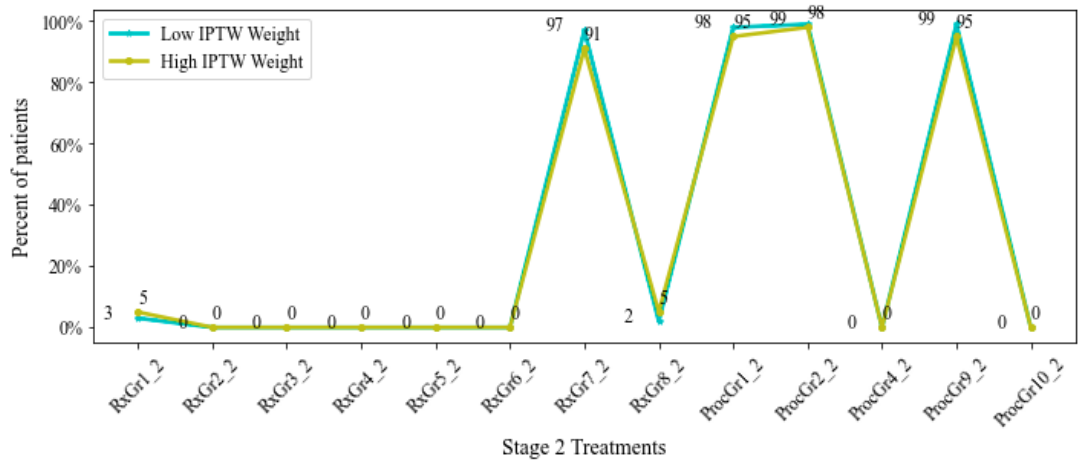


Figure 36: Stage 2 Treatment Usage 2SP IQ for Low and High IPTW weight patients

The treatment recommendations with the 2SP AL-IPTW approach are shown in Figure 37. The pharmaceutical treatments RxGr3 (Narcotic), RxGr4 (Muscle Relaxant), and RxGr7 (Sleeping Pill) are used similarly in low and high weight patients. The procedural treatments ProcGr1 (Injection) and ProcGr4 (Stimulation Procedure) are recommended more to low weight patients, while ProcGr2 (Block Procedure) is recommended more to high weight patients.

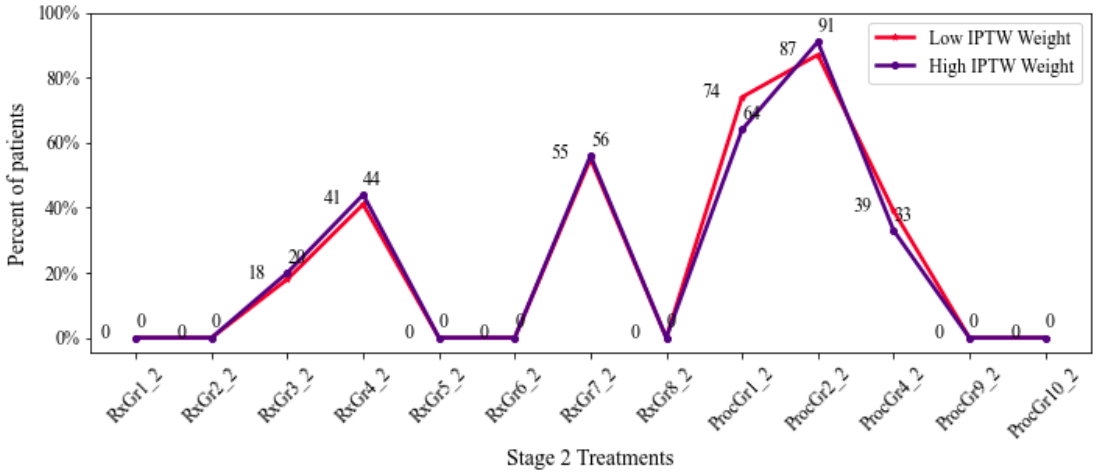


Figure 37: Stage 2 Treatment Usage 2SP AL-IPTW for Low and High IPTW weight patients

The treatment recommendations using the 2SP IQ-IPTW approach are shown in Figure 38. ProcGr4 (Stimulation Procedure) is predominantly recommended to high weight patients. The treatments RxGr1

(Tramadol), RxGr3 (Narcotic), RxGr4 (Muscle Relaxant), RxGr7 (Sleeping Pill), ProcGr2 (Block Therapy), and ProcGr9 (Cognitive Behavioral Therapy) are recommended slightly more frequently to the high weight patients. The treatment ProcGr1 (Injection) is recommended similarly to both low and high weight patients.

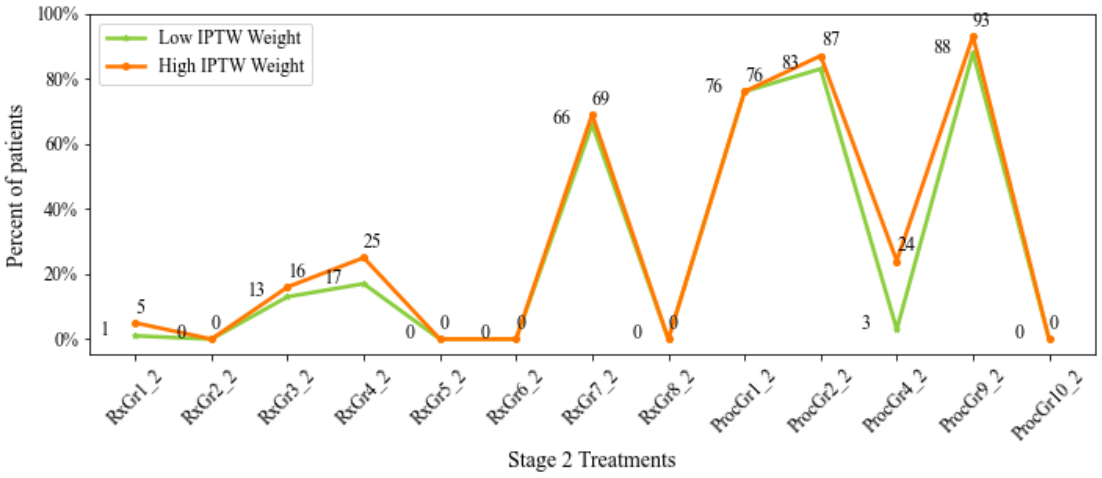


Figure 38: Stage 2 Treatment Usage 2SP IQ-IPTW for Low and High IPTW weight patients

The comparison of the treatment assignment to low and high IPTW patients between 2SP IQ-IPTW, 2SP AL-IPTW, and 2SP IQ is shown in Figures 39(a) and 39(b), respectively. The treatment ProcGr4 (Stimulation Therapy) is recommended to a greater percentage of high weight patients, while it is recommended to a very small percentage of low weight patients by 2SP IQ-IPTW. The treatment ProcGr1 (Injection) is recommended to the same percentage of low weight patients by 2SP AL-IPTW and 2SP IQ-IPTW, while it is recommended to a greater percentage of high weight patients by 2SP IQ-IPTW. The recommendation of treatment RxGr4 (Muscle Relaxant) is also greater to the high weight patients by 2SP IQ-IPTW. This analysis shows that the 2SP IQ-IPTW approach recommends more treatments to the high weight patients than the other two methods. The 2SP IQ-IPTW approach is more adaptive than the other two approaches and by recommending more treatments to high weight patients than the low weight patients, so it reduces over-medication to more common patients with low IPTW weights.

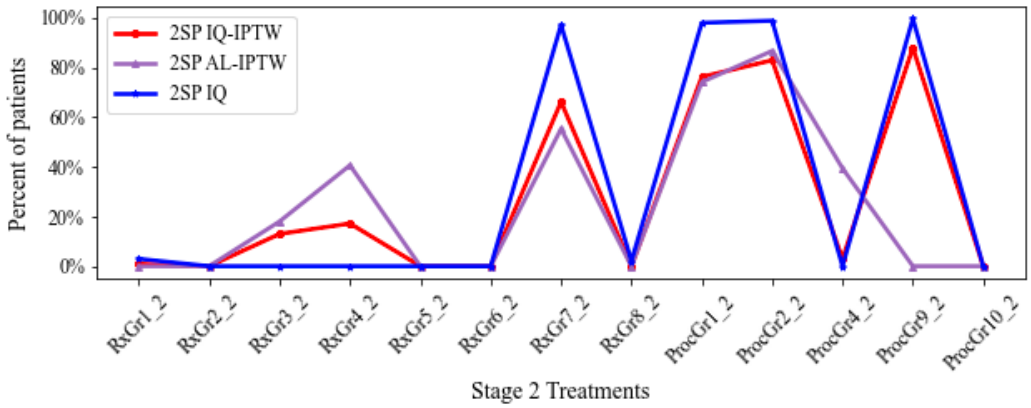


Figure 39(a): Stage 2 treatment assignment comparison for low IPTW weight patients

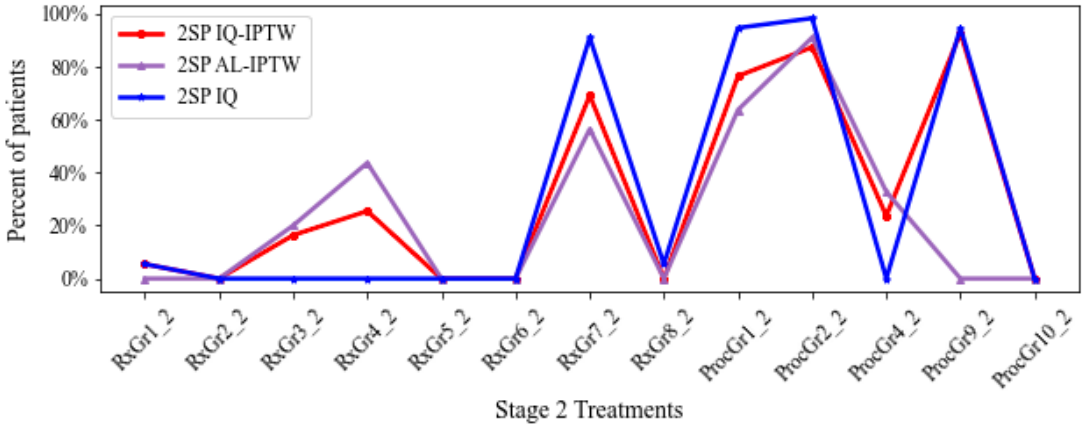


Figure 39(b): Stage 2 treatment assignment comparison for high IPTW weight patients

4.4 Conclusion and Future Work

This research proposes the 2SP IQ-IPTW MIQCP optimization approach using the state transition models developed using the HierNet-IPTW method, which uses IPTW to address time-varying confounding and models the state treatment interaction effects. The treatment decisions recommended by the proposed 2SP IQ-IPTW approach are compared against treatments recommended by the 2SP IQ approach, which does not use IPTW, and those from the 2SP AL-IPTW approach, which does not model interaction effects.

The analysis of stage 1 and stage 2 treatment recommendations show that the 2SP IQ approach recommends fewer treatments to most of the patients, and the recommendation pattern is similar between

low and high IPTW weight patients. The 2SP AL-IPTW approach recommends more treatments than 2SP IQ recommends, and the patterns are similar between low and high IPTW weight patients. The pharmaceutical treatments recommended by the 2SP IQ-IPTW approach are similar to those of the 2SP AL-IPTW approach, but the 2SP IQ-IPTW approach recommends more procedural treatments. The 2SP IQ-IPTW approach also recommends certain treatments to a greater percentage of the high IPTW weight patients than the low IPTW weight patients.

The inclusion of IPTW and state treatment interaction effects in building the state transition models results in more adaptive treatment decisions being made by the optimization module. The optimization recommends more treatments to the high IPTW weight patients that represent rare patient instances in the dataset. This avoids overprescribing certain medications to all the patients.

The difference in treatment recommendations between the optimization and the observed dataset needs to be studied further. The proposed 2SP IQ-IPTW approach uses LASSO-based HierNet modeling to build state transition models with interactions. We can consider other modeling techniques to build state transition models with interactions and compare them against the 2SP IQ-IPTW treatment recommendations. In preliminary research, we considered Glinternet, another LASSO-based modeling method with interactions. This method can be explored in future work. The results from the simulation case study in the preliminary research are included in Appendix V.

Chapter 5

Discussion and Future Work

The main aim of this research is to develop state transition models on the pain management dataset while addressing the challenge of including state-treatment interactions that are critical to enable personalized treatment plan for patients. In Chapter 3, we propose a method named HierNet-IPTW, where HierNet, a LASSO based method is combined with IPTW from Ohol [26] to build state transition and outcome models that enable feature selection and modeling of interaction effects in the presence of time-varying confounding. This approach is studied using a simulated case study structured based on the McDermott Center data with time varying confounding and various correlation structures between the variables. The performance of the proposed HierNet-IPTW approach is compared against the baseline method that does not address time varying confounding. The evaluation was done on the feature and interaction selection metrics. The ANOVA conducted on the experimental data showed that the proposed HierNet-IPTW approach performed better than the baseline in correctly identifying the true feature and interaction terms.

In Chapter 4, we build state transition models on the pain management dataset. These models are used in a multi-objective two stage stochastic optimization problem formulated as a MIQCP problem named 2SP IQ-IPTW. The treatment recommendations from the proposed 2SP IQ-IPTW approach were compared against the treatments recommended from the 2SP AL-IPTW and 2SP IQ approach. The results showed that the inclusion of IPTW and state treatment interaction effects in building the state transition models results in more adaptive treatment decisions being made by the optimization module. The proposed approach recommends more treatments to the high IPTW weight patients that represent rare patient instances in the dataset. This avoids overprescribing certain medications to all the patients.

The comparison of the final patient pain outcome measures using the different first stage treatment policies generated from the optimization is a subject of future work. This comparison is difficult since the

policies were generated under different state transition and outcome models. This can be done using an evaluation framework that is closer to the ground truth. In this study, we cannot conclusively tell that a particular state transition model is closer to the ground truth. We can assume that the state transition models used in the 2SP IQ-IPTW framework generated using the proposed HierNet-IPTW are closer to the ground truth since it captures more information from the given data. In future work, the different treatment policies shall be evaluated using the 2SP IQ-IPTW evaluation framework.

The difference in treatment recommendations between the optimization and the observed dataset needs to be studied further. The proposed 2SP IQ-IPTW approach uses LASSO-based HierNet modeling to build state transition models with interactions. In future work, we can consider other modeling techniques to build state transition models with interactions and compare them against the 2SP IQ-IPTW treatment recommendations.

Appendix I: Modifications in HierNet to force treatments and incorporate IPTW weights

The optimization problem shown in Equation (15) is solved in HierNet

$$\text{argmin}_{\mu,\beta,\theta} \frac{1}{2}\sum_{i=1}^{n}(y_i - \mu - x_i^T\beta - \frac{1}{2}x_i^T\Theta x_i)^2 + \lambda 1^T(\beta^+ + \beta^-) + \frac{\lambda}{2}\|\Theta\|_1, \quad (15)$$

$$\text{subject to } \Theta = \Theta^T, \left\|\Theta_j\right\|_1 \leq \left(\beta_j^+ + \beta_j^-\right),\ \beta_j^+ \geq 0,\ \beta_j^- \geq 0,$$

In order to force the treatments in the model, the constraint $\beta_{jT}^+ + \beta_{jT}^- \geq \varepsilon$ is added where β_{jT}^+ and β_{jT}^- are the treatment main effect co-efficients. ε is set to 0.05 in this study. In the HierNet R package, this modification is incorporated in the ONEROW function.

The loss function is modified as shown in Equation (16), where w_i are the weights obtained from IPTW

$$\text{argmin}_{\mu,\beta,\theta} \frac{1}{2}\sum_{i=1}^{n} w_i(y_i - \mu - x_i^T\beta - \frac{1}{2}x_i^T\Theta x_i)^2 \quad (16)$$

Appendix II: Residual analysis: Checking model assumptions

Feature Selection Sensitivity

The error terms have constant variance: The residual vs experimental treatments plot in Figure 40 does not show the presence of a funnel shape, and we can conclude that the error terms have constant error variance.

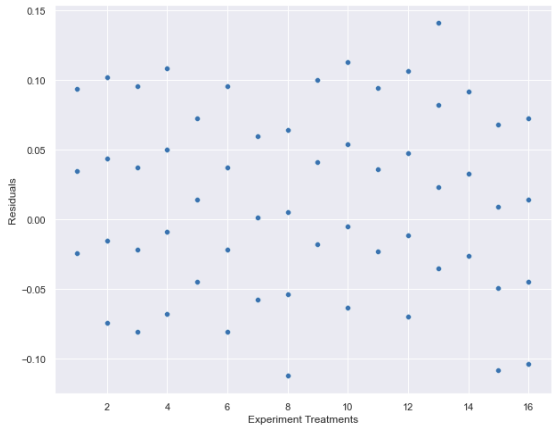


Figure 40: Residuals versus experiment treatments feature selection sensitivity

The error terms are normally distributed: The normality plot shown in Figure 41, shows that the residuals are along the normal distribution line. We can conclude that the error terms are normally distributed.

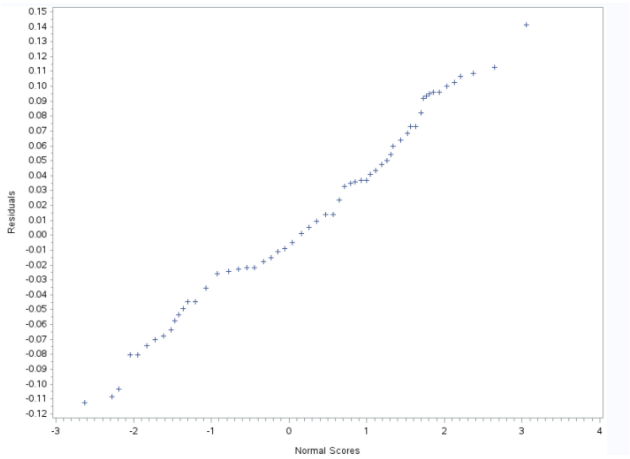


Figure 41: Normal Probability Plot (Q-Q Plot) feature selection sensitivity

Feature Selection Specificity

The error terms have constant variance: The residual vs experimental treatments plot in Figure 42 does not show the presence of a funnel shape, and we can conclude that the error terms have constant error variance.

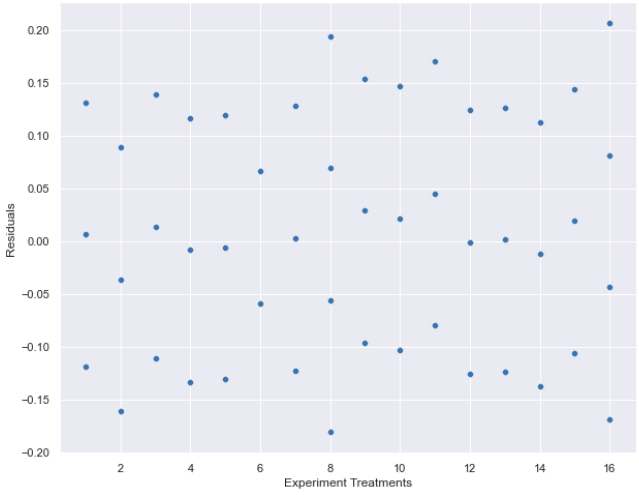


Figure 42: Residuals versus experimental treatments for feature selection specificity

The error terms are normally distributed: The normality plot shown in Figure 43, shows that the residuals are along the normal distribution line. We conclude that the error terms are normally distributed.

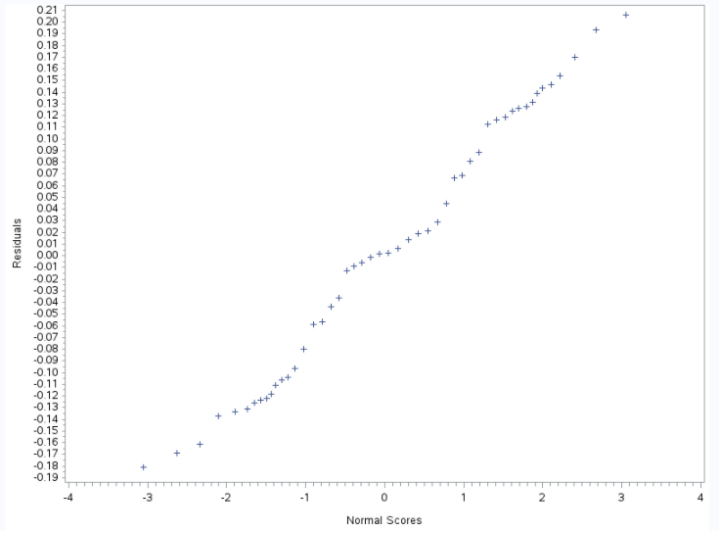


Figure 43: Normal Probability Plot (Q-Q Plot) feature selection specificity

Interaction Selection Sensitivity

The error terms have constant variance: The residual vs experimental treatments plot in Figure 44 does not show the presence of a funnel shape, and we conclude that the error terms have constant error variance

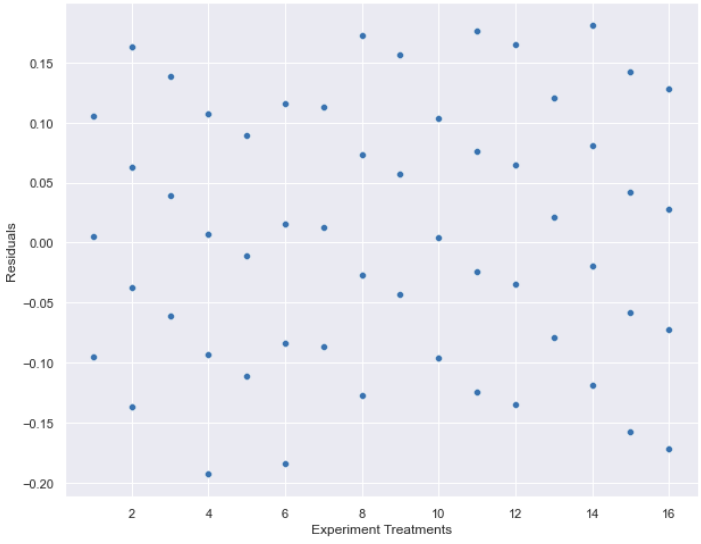


Figure 44: Residuals versus experimental treatments for interaction selection sensitivity

The error terms are normally distributed: The normality plot (Figure 45) shows that the residuals are symmetrical with shorter upper and lower tails. We conclude that the normality assumption is not satisfied.

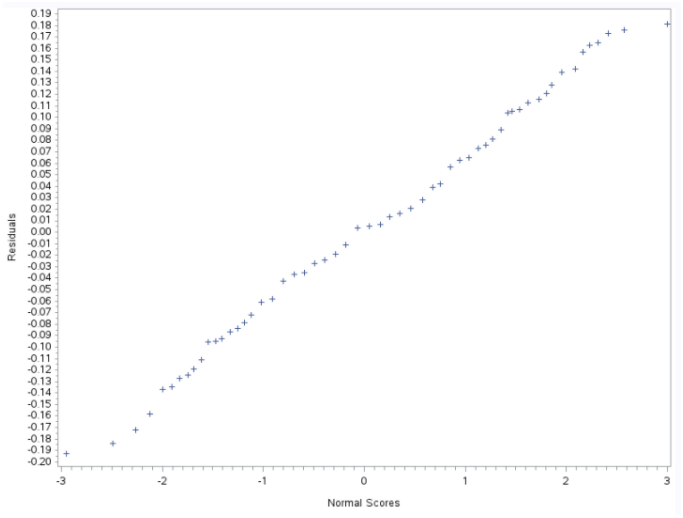


Figure 45: Normal Probability Plot (Q-Q Plot) interaction selection sensitivity

Interaction Selection FDR

The error terms have constant variance: The residual vs experimental treatments plot in Figure 46 does not show the presence of a funnel shape, and we conclude that the error terms have constant error variance.

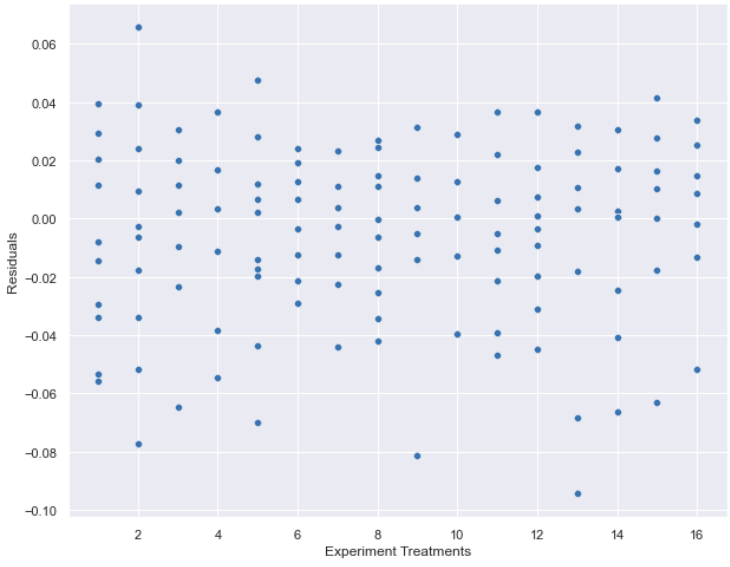


Figure 46: Residuals versus experimental treatments for interaction selection FDR

The error terms are normally distributed: The normality plot (Figure 47) shows that the residuals are symmetrical with longer upper and lower tail. We conclude that the normality assumption is not satisfied.

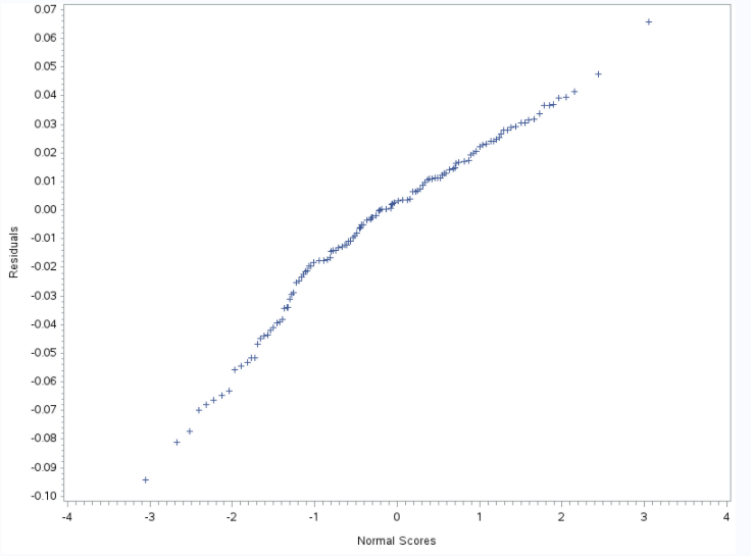
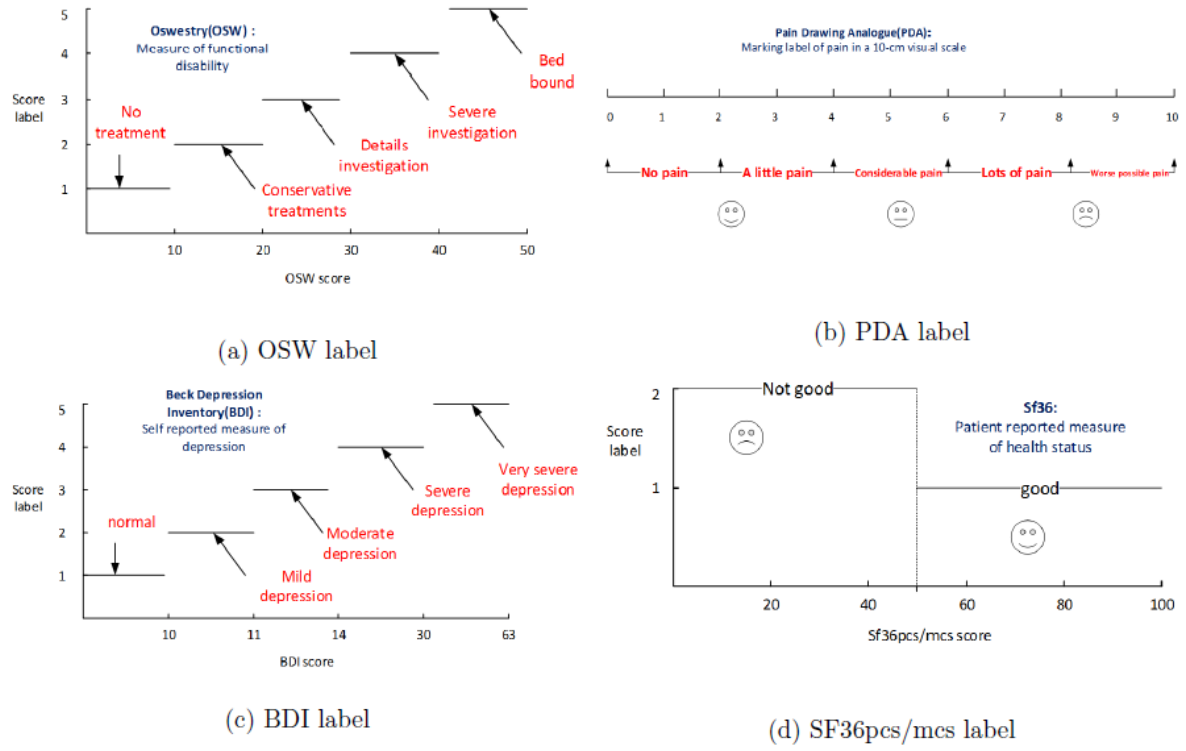


Figure 47: Normal Probability Plot (Q-Q Plot) interaction selection FDR

Appendix III

The different levels of the pain outcome measures indicating the severity of pain, depression, and general health status.



(a) OSW label

(b) PDA label

(c) BDI label

(d) SF36pcs/mcs label

Figure 48: Different Pain Outcome Measures and their levels [27]

The description and type of the state, treatment and outcome variables in the pain management study is shown in Table 35.

Table 35: Description of state, treatment, and outcome variables in pain management dataset [27]

Variable type	Variable Name	Description	Values
State Variables	pre_PDA	PDA measure at the initial point	Continuous
	pre_OSW	OSW measure at the initial point	Continuous
	pre_BDI	BDI measure at the initial point	Continuous
	pre_SF36-pcs	SF36-pcs measure at the initial point	Continuous
	pre_SF36-mcs	SF36-mcs measure at the initial point	Continuous
	Age	Patient's age	Continuous
	Children	Children	Continuous
	Onset	Time (in months) since the first onset of pain	Continuous
	Duration	Duration	Continuous
	Status	Status of condition	1: acute (< 3 months), 2:acute (< 6 months), 3:acute (< 9 months)
	Race_1	Race of Patient	0:no, 1:Caucasian
	Race_2	Race of Patient	0:no, 1:African American
	Litigate	Pending litigation related to pain?	0:no, 1:yes
	Gender	Patient's gender	0:male, 1:female
	phydx1	Physical Dx1/Facial 784.0	0:no, 1:yes
	phydx3	Physical Dx3/Headache 784.0	0:no, 1:yes
	phydx4	Physical Dx4/Cervical 723.1	0:no, 1:yes
	phydx5	Physical Dx5/Thoracic 724.1	0:no, 1:yes
	phydx6	Physical Dx6/Lumbar 724.2	0:no, 1:yes
	phydx7	Physical Dx7/Myofascial-Fibromyalgia 729.1	0:no, 1:yes
	phydx8	Physical Dx8/Abdominal 789.0	0:no, 1:yes
	phydx11	Physical Dx11/Upper Extremity 729.5	0:no, 1:yes
	phydx12	Physical Dx12/Low Extremity 729.5	0:no, 1:yes
	phydx14	Physical Dx14/Osteoarthritis 716.9	0:no, 1:yes
	phydx15	Physical Dx15/Sacro-illitis 724.6	0:no, 1:yes
	phydx20	Physical Dx20/Neuralgia, Neuritis, Unspecified	0:no, 1:yes

Variable type	Variable Name	Description	Values
	phydx31	Physical Dx/Cervical Spondylosis W/O Myelopathy (721.0)	0:no, 1:yes
	ProcGr1_0	Injection in stage 0	0:no, 1:yes
	ProcGr2_0	Block Procedure in stage 0	0:no, 1:yes
	ProcGr4_0	Stimulation Procedure in stage 0	0:no, 1:yes
	ProcGr9_0	Psychotherapy in stage 0	0:no, 1:yes
	ProcGr10_0	Physical Therapy in stage 0	0:no, 1:yes
	ProcGr11_0	Number of Additional Procedures in stage 0	0:no, 1:yes
	pastdx3	Past Dx3/Headache 784.0	0:no, 1:yes
	pastdx4	Past Dx4/Cervical 723.1	0:no, 1:yes
	pastdx5	Past Dx5/Thoracic 724.1	0:no, 1:yes
	pastdx6	Past Dx6/Lumbar 724.2	0:no, 1:yes
	pastdx7	Past Dx7/Myofascial-Fibromyalgia 729.1	0:no, 1:yes
	pastdx8	Past Dx8/Abdominal 789.0	0:no, 1:yes
	pastdx11	Past Dx11/Upper Extremity 729.5	0:no, 1:yes
	pastdx12	Past Dx12/Low Extremity 729.5	0:no, 1:yes
	pastdx14	Past Dx14/Osteoarthritis 716.9	0:no, 1:yes
	pastdx15	Past Dx15/Sacro-illitis 724.6	0:no, 1:yes
	pastdx20	Past Dx20/Neuralgia, Neuritis, Unspecified	0:no, 1:yes
	pastdx32	Past Dx/Number of Additional Diagnoses	0:no, 1:yes
	SghxGr1	Surgical History/Unspecified discectomy	0:no, 1:yes
	SghxGr3	Surgical History/Percutaneous discectomy	0:no, 1:yes
	SghxGr5	Surgical History/Unspecified fusion	0:no, 1:yes
	SghxGr6	Surgical History/Anterior fusion	0:no, 1:yes
	SghxGr11	Surgical History/Hardware removal	0:no, 1:yes
	RxGr1_0	Tramadol in stage 0	0:no, 1, 2,3
	RxGr2_0	NSAIDs in stage 0	0:no, 1, 2,3
	RxGr3_0	Narcotic in stage 0	0:no, 1, 2,3
	RxGr4_0	Muscle Relaxant in stage 0	0:no, 1, 2,3
	RxGr5_0	Antidepressant in stage 0	0:no, 1, 2,3
	RxGr6_0	Tranquilizer in stage 0	0:no, 1, 2,3
	RxGr7_0	Sleeping Pills in stage 0	0:no, 1, 2,3
	RxGr8_0	Others in stage 0	0:no, 1, 2,3
	marital_1	Marital Status of Patient	0:no, 1:single
	marital_2	Marital Status of Patient	0:no, 1:married

Variable type	Variable Name	Description	Values
	marital_3	Marital Status of Patient	0:no, 1:divorced
	marital_4	Marital Status of Patient	0:no, 1:widow
Decision Variables	ProcGr1_1	Injection in stage 1	0:no, 1:yes
	ProcGr2_1	Block Procedure in stage 1	0:no, 1:yes
	ProcGr4_1	Stimulation Procedure in stage 1	0:no, 1:yes
	ProcGr9_1	Psychotherapy in stage 1	0:no, 1:yes
	ProcGr10_1	Physical Therapy in stage 1	0:no, 1:yes
	ProcGr11_1	Number of Additional Procedures in stage 1	0:no, 1:yes
	RxGr1_1	Tramadol in stage 1	0:no, 1, 2
	RxGr2_1	NSAIDs in stage 1	0:no, 1, 2, 3
	RxGr3_1	Narcotic in stage 1	0:no, 1, 2, 3
	RxGr4_1	Muscle Relaxant in stage 1	0:no, 1, 2, 3
	RxGr5_1	Antidepressant in stage 1	0:no, 1, 2, 3
	RxGr6_1	Tranquilizer in stage 1	0:no, 1, 2, 3
	RxGr7_1	Sleeping Pills in stage 1	0:no, 1, 2
	RxGr8_1	Others in stage 1	0:no, 1, 2
	ProcGr1_2	Injection in stage 2	0:no, 1:yes
	ProcGr2_2	Block Procedure in stage 2	0:no, 1:yes
	ProcGr4_2	Stimulation Procedure in stage 2	0:no, 1:yes
.	ProcGr9_2	Psychotherapy in stage 2	0:no, 1:yes
	ProcGr10_2	Physical Therapy in stage 2	0:no, 1:yes
	RxGr1_2	Tramadol in stage 2	0:no, 1, 2, 3
	RxGr2_2	NSAIDs in stage 2	0:no, 1, 2, 3
	RxGr3_2	Narcotic in stage 2	0:no, 1, 2, 3
	RxGr4_2	Muscle Relaxant in stage 2	0:no, 1, 2, 3
	RxGr5_2	Antidepressant in stage 2	0:no, 1, 2, 3
	RxGr6_2	Tranquilizer in stage 2	0:no, 1, 2, 3
	RxGr7_2	Sleeping Pills in stage 2	0:no, 1, 2
	RxGr8_2	Others in stage 2	0:no, 1, 2,3

The two-way adverse treatment interaction constraints are shown in Table 36.

Table 36: Adverse Treatment Interaction Constraints [27]

	RxGr1	RxGr2	RxGr3	RxGr4	RxGr5	RxGr6	RxGr7	RxGr8	ProcGr1	ProcGr2
RxGr1		✓	✗	✗	✗	✗	✗	✓	✗	✓
RxGr2			✓	✓	✓	✗	✓	✗	✗	✗
RxGr3				✗	✗	✗	✓	✓	✓	✓
RxGr4					✓	✗	✗	✗	✓	✓
RxGr5						✗	✗	✓	✗	✗
RxGr6							✗	✓	✓	✓
RxGr7								✗	✓	✓
RxGr8									✗	✗

Appendix IV

The stage 1 and stage 2 state transition models for five pain outcome measures are shown in the following tables.

Table 37: Stage 1 Pain Outcome Transition Models 2 SP AL-IPTW

Variable	mid_PDA	mid_OSW	mid_BDI	mid_sf36pcs	mid_sf36mcs
Intercept	4.312	10.983	11.212	27.569	34.82
RxGr1_1	0.027	0.507	0.011	-0.205	-0.475
RxGr2_1	-0.156	-0.561	-0.171	0.539	0.329
RxGr3_1	-0.003	-0.061	0.373	-0.2	-0.255
RxGr4_1	-0.322	-1.52	-0.281	0.582	0.626
RxGr5_1	-0.526	-2.24	-0.645	1.141	0.903
RxGr6_1	0.038	-0.048	-0.054	-0.073	0.047
RxGr7_1	0.055	0.012	0.008	-0.353	-0.02
RxGr8_1	0.007	0.012	-0.055	0.003	-0.135
ProcGr1_1	-0.006	-0.056	-0.021	-0.049	-0.208
ProcGr2_1	0.029	0.122	-0.015	-0.436	-0.009
ProcGr4_1	-0.081	-0.273	-0.123	0.319	0.117
ProcGr9_1	-0.371	-1.853	-0.962	1.422	1.102
ProcGr10_1	-0.0618	-0.206	-0.107	0.275	0.073
ProcGr11_1	-0.041	0.033	0.052	0.141	-0.033
ProcGr1_0	0.318	0	0	0	-0.422
RxGr2_0	0.685	0	0.894	0	-1.398
RxGr3_0	0.196	0.431	0	-2.678	0

phydx6	0.344	0	0	-1.286	0
phydx12	0.459	0	0	-0.66	-1.728
phydx14	0.777	0	0	0	0
pastdx4	0.268	0	0	-1.167	0
pastdx12	0.104	0	0	0	-1.715
pre_PDA	0.297	0	0	0	0
race_1	0.07	0	0	0	0
age	0.025	0	0	0	-0.06
ProcGr4_0	0	0.293	0.101	0	0
ProcGr10_0	0	-0.119	0	0	0
phydx4	0	1.875	0	0	0
phydx31	0	2.009	0	0	-2.35
pastdx6	0	0.769	1.19	0	0
pastdx7	0	1.457	0	0	0
pastdx11	0	1.274	0	0	0
SghxGr3	0	2.686	0	-2.349	0
pre_OSW	0	0.291	0	0.103	0
marital_1	0	1.141	0	0	0
onset	0	0.094	0	0	0
gender	0	0.79	0	0	0
ProcGr9_0	0	0	-0.179	0	0.275
phydx7	0	0	1.357	0	0
phydx20	0	0	2.268	0	0
pastdx32	0	0	0.834	0	0
SghxGr6	0	0	1.259	0	-2.423
pre_BDI	0	0	0.221	0	0
status	0	0	0.834	0	0
ProcGr11_0	0	0	0	-0.017	0
RxGr4_0	0	0	0	-0.017	0
pre_sf36pcs	0	0	0	0.18	0
pre_sf36mcs	0	0	0	0	0.214
duration	0	0	0	0	-0.07

Table 38: Stage 2 Pain Outcome Transition Models 2 SP AL-IPTW

Variable	post_PDA	post_OSW	post_BDI	post_sf36pcs	post_sf36mcs
Intercept	4.31	9.548	10.1	24.623	30.461
RxGr1_2	0.259	0.077	0.071	-0.149	-0.033
RxGr2_2	0.207	0.039	0.056	-0.076	-0.043
RxGr3_2	-0.113	-0.66	-0.956	0.835	0.438
RxGr4_2	-0.374	-1.119	-1.08	1.236	0.671
RxGr5_2	0.079	-0.047	-0.183	-0.029	0.211
RxGr6_2	0.201	-0.024	0.436	0.046	-0.254
RxGr7_2	-0.707	-1.876	-1.378	1.805	1.016
RxGr8_2	0.066	0.047	-0.053	-0.513	-0.671
ProcGr1_2	-0.438	-1.849	-1.881	2.04	1.523
ProcGr2_2	-0.609	-2.177	-2.134	2.488	2.215
ProcGr4_2	-0.146	-1.39	-0.927	0.782	1.161
ProcGr9_2	-0.005	-0.131	-0.107	-0.385	0.208
ProcGr10_2	0.179	-0.02	0.475	-0.261	0.091
ProcGr9_1	-0.376	-1.137	-0.55	0.655	1.029
RxGr2_1	-0.19	0	0	0	0
RxGr4_1	-0.218	-0.546	0	0.436	0
RxGr2_0	0.723	0	0.697	0	-0.314
ProcGr10_0	-0.002	0	-0.066	0	0.208
ProcGr4_0	-0.007	0	0	-0.115	0
marital_2	-0.034	0	0	0	0
mid_PDA	0.463	0.119	0	0	0
gender	0.423	0	0	0	0
age	0.037	0	0.007	0.012	0
phydx6	0.518	0	0	0	-1.343
phydx12	1.033	0	1.349	-1.054	0
pastdx4	0.115	0	0	-0.63	-1.211
pastdx6	0.228	0	0	0	0
SghxGr3	0.123	1.332	1.211	-0.697	-1.883
onset	0.018	0	0	0.083	0
race_1	0.116	0	0.347	0	0
RxGr5_1	0	-0.729	-0.624	0.504	0.651
ProcGr1_0	0	-0.133	0	0.16	0
ProcGr11_0	0	-0.206	0.765	0	0
RxGr3_0	0	0.269	0	-2.018	0
RxGr5_0	0	0.584	0.365	0	-0.254
phydx4	0	1.172	0	0	0
phydx7	0	0.827	0.966	0	0

phydx31	0	2.365	0	0	-0.607
mid_OSW	0	0.252	0.021	0.172	0
pre_OSW	0	0.129	0	0	0
pastdx7	0	0.711	2.038	0	0
pastdx11	0	0.535	0	0	0
SghxGr6	0	0.868	0	0	-0.565
duration	0	0.074	0	0	0.054
litigat	0	0.007	0	0	0
ProcGr10_1	0	0	-0.142	0	0
ProcGr4_1	0	0	-0.203	0	0
phydx20	0	0	1.069	0	0
mid_sf36mcs	0	0	0.082	0	0.124
mid_BDI	0	0	0.248	0	0
RxGr1_0	0	0	0	-1.131	0
RxGr6_0	0	0	0	-2.209	0
phydx11	0	0	0	-2.082	0
mid_sf36pcs	0	0	0	0.308	0.017
pastdx12	0	0	0	0	-2.204

Table 39: Stage 1 Pain Outcome Transition Models 2 SP IQ-IPTW

Variable	mid_PDA	mid_OSW	mid_BDI	mid_sf36pcs	mid_sf36mcs
Intercept	4.017	8.615	10.406	25.216	30.612
RxGr1_1	0.06	0.86	-0.078	-0.05	-0.119
RxGr2_1	-0.197	-0.379	-0.05	0.317	-0.05
RxGr3_1	0.244	0.952	-0.05	-1.428	-0.05
RxGr4_1	-0.547	-2.667	-0.752	1.883	2.895
RxGr5_1	-0.399	-1.62	-0.414	0.764	1.554
RxGr6_1	-0.05	-0.05	-0.05	-2.739	-0.05
RxGr7_1	-0.05	-0.05	-0.05	-0.05	-0.05
RxGr8_1	0.205	-0.05	0.772	-0.05	-0.05
ProcGr1_1	-0.125	-0.05	-0.102	0.159	0.454
ProcGr2_1	-0.05	-0.05	-0.05	-0.05	-0.138
ProcGr4_1	-0.236	-0.265	-0.158	0.623	0.027
ProcGr9_1	-0.581	-1.792	-0.87	1.291	1.81
ProcGr10_1	-0.05	-0.412	-0.255	0.36	0.07
ProcGr11_1	-0.05	-0.05	-0.05	-0.632	-0.05
phydx14	0.581	0	0	0	0
pastdx7	0.142	0	1.08	0	0
onset	0.049	0	0	-0.019	0

phydx6	0.134	0	1.073	0	0
pastdx4	0.13	0	0	0	0
pre_PDA	0.383	0	0	0	0
age	0.061	0.021	0	0	0
RxGr5_0	-0.155	0	0	0	0
ProcGr10_0	-0.186	-1.129	-0.296	0	0
RxGr3_0	0.458	0	0	0	0
ProcGr4_1:RxGr5_0	-0.258	0	0	0	0
ProcGr9_1:ProcGr10_0	-0.443	0	0.12	0	0
RxGr2_1:pastdx7	-0.17	0	0	0	0
RxGr3_1:onset	-0.109	0	0	0	0
RxGr4_1:phydx6	-0.32	0	0	0	0.375
RxGr5_1:RxGr3_0	-0.203	0	0	0	0
RxGr8_1:pre_PDA	0.452	0	0	0	0
age:phydx14	0.17	0	0	0	0
onset:pastdx7	0.227	0	0	0	0
age:onset	0.104	0	0	0	0
RxGr5_0:phydx6	-0.067	0	0	0	0
age:ProcGr10_0	-0.005	0	0	0	0
RxGr3_0:pre_PDA	0.003	0	0	0	0
age:pre_PDA	0.004	0	0	0	0
RxGr5_0:pastdx4	-0.078	0	0	0	0
phydx14:pastdx7	0.134	0	0	0	0
RxGr3_0:ProcGr10_0	0.065	0	0	0	0
phydx6:pre_PDA	0.002	0	0	0	0
phydx4	0	3.92	0	0	0
pastdx6	0	0.996	0	0	0
race_1	0	0.076	0	0	0
marital_4	0	-0.125	0	0	0
SghxGr3	0	-0.014	0	-0.423	0
RxGr2_0	0	0.012	1.478	0	0
pre_OSW	0	0.314	0	0.024	0
ProcGr9_1:RxGr2_0	0	-1.43	0	0	0
ProcGr10_1:SghxGr3	0	-0.246	0	0	0
RxGr1_1:race_1	0	0.434	0	0	0
RxGr2_1:marital_4	0	-0.311	0	0	0
RxGr4_1:ProcGr10_0	0	-1.209	0	0	0
RxGr4_1:pastdx6	0	-0.072	0	0	0
RxGr4_1:pre_OSW	0	-0.071	0	0	0
RxGr5_1:phydx4	0	-0.128	0	0	0
race_1:pre_OSW	0	0.009	0	0	0

phydx4:pastdx6	0	0.149	0	0	0
RxGr2_0:pastdx6	0	0.721	0	0	0
ProcG10_0:pastdx6	0	1.979	0	0	0
RxGr2_0:SghxGr3	0	0.388	0	0	0
pastdx6:SghxGr3	0	0.234	0	-0.437	0
RxGr2_0:phydx31	0	0.256	0	0	0
ProcGr10_0:race_1	0	-0.056	0	0	0
phydx4:pre_OSW	0	0.011	0	0	0
age:pre_OSW	0	0.01	0	0	0
phydx4:SghxGr3	0	0.952	0	0	0
SghxGr3:marital_4	0	-0.44	0	0	0
age:phydx4	0	0.038	0	0	0
age:SghxGr3	0	0.078	0	0	0
phydx20	0	0	0.214	-1.36	0
marital_1	0	0	-0.034	0	0.615
SghxGr11	0	0	0.118	0	0
ProcGr11_0	0	0	-0.102	0.443	-1.002
pre_BDI	0	0	0.223	0	0
status	0	0	0.052	0	0
ProcGr4_1:phydx20	0	0	-0.443	0	0
ProcGr9_1:pre_BDI	0	0	-0.125	0	0
ProcGr10_1:phydx12	0	0	-0.079	0	0
RxGr4_1:RxGr2_0	0	0	0.142	0	0
RxGr8_1:pastdx7	0	0	1.414	0	0
pastdx7:SghxGr11	0	0	0.926	0	0
ProcGr10_0:phydx6	0	0	-0.626	0	0
phydx20:SghxGr11	0	0	0.312	0	0
phydx20:pastdx7	0	0	0.104	0	0
ProcGr11_0:pastdx7	0	0	0.824	0	0
RxGr6_0:SghxGr11	0	0	0.217	0	0
phydx20:pastdx5	0	0	0.409	0	0
ProcGr2_0:phydx6	0	0	-0.659	0	0
RxGr2_0:ProcGr11_0	0	0	0.045	0	0
status:ProcGr10_0	0	0	0.01	0	0
phydx12	0	0	0	-0.348	-1.687
pastdx5	0	0	0	-1.035	0
gender	0	0	0	-0.014	0
pre_sf36pcs	0	0	0	0.216	0
ProcGr1_1:pastdx20	0	0	0	0.792	0
ProcGr4_1:phydx12	0	0	0	0.377	0
ProcGr11_1:pastdx6	0	0	0	0.192	0

RxGr1_1:SghxGr3	0	0	0	-0.714	0
RxGr5_1:ProcGr1_0	0	0	0	0.94	0
onset:SghxGr3	0	0	0	0.016	0
phydx12:SghxGr3	0	0	0	-0.048	0
ProcGr1_0:pastdx6	0	0	0	0.952	0
pre_OSW:pre_sf36pcs	0	0	0	0.01	0
gender:phydx12	0	0	0	-0.984	0
ProcGr1_0:phydx12	0	0	0	1.048	0
ProcGr1_0:phydx20	0	0	0	0.236	0
phydx31	0	0	0	0	-0.349
pastdx3	0	0	0	0	-1.027
SghxGr5	0	0	0	0	-1.266
pastdx12	0	0	0	0	-0.345
pre_sf36mcs	0	0	0	0	0.238
ProcGr9_0	0	0	0	0	-0.375
ProcGr2_0	0	0	0	0	-1.402
ProcGr9_1:phydx31	0	0	0	0	1.965
RxGr1_1:pastdx12	0	0	0	0	-1.402
phydx12:pre_sf36mcs	0	0	0	0	-0.466
phydx31:SghxGr5	0	0	0	0	-0.375
ProcGr9_0:SghxGr5	0	0	0	0	0.266
ProcGr2_0:ProcGr11_0	0	0	0	0	-0.14
pastdx12:marital_1	0	0	0	0	-0.375
ProcGr2_0:marital_1	0	0	0	0	-0.615
RxGr8_0:phydx5	0	0	0	0	-1.1

Table 40: Stage 2 Pain Outcome Transition Models 2 SP IQ-IPTW

Variable	post_PDA	post_OSW	post_BDI	post_sf36pcs	post_sf36mcs
Intercept	4.798	8.721	11.627	28.216	32.063
RxGr1_2	-0.081	-0.102	0.206	-0.05	-0.05
RxGr2_2	-0.05	0.211	-0.05	-0.05	-0.05
RxGr3_2	-0.162	-0.279	-0.866	1.069	1.437
RxGr4_2	-0.153	-0.633	-1.054	1.215	1.298
RxGr5_2	-0.05	-0.05	-0.05	-0.05	-0.254
RxGr6_2	-0.05	-0.05	-0.05	-0.079	-0.05
RxGr7_2	-0.5	-1.19	-1.803	2.586	2.446
RxGr8_2	-0.05	-0.05	-0.05	-0.217	-0.05
ProcGr1_2	-0.339	-1.328	-2.252	0.675	1.499
ProcGr2_2	-0.467	-1.692	-2.217	1.154	2.344

ProcGr4_2	-0.05	-0.175	-1.27	0.079	0.575
ProcGr9_2	-0.268	-2.08	-3.627	1.871	2.475
ProcGr10_2	-0.05	0.66	-0.05	-0.05	0.108
RxGr4_1	-0.249	-0.58	-0.715	0.967	1.14
RxGr5_1	-0.162	0	0	0.699	0
ProcGr10_0	-0.062	0.441	0.596	0	-0.312
RxGr3_0	0.201	0	0	0	0
phydx12	0.228	0	0	-1.183	-2.263
phydx4	0.147	1.871	0	0	0
pastdx7	0.228	0	1.006	2.04	0
SghxGr3	0.304	0	0.884	-1.182	-4.476
pre_PDA	0.004	0	0	0	0
mid_PDA	0.58	0.129	0	0	0
age	0.008	0	0	0	0
race_1	0.258	0.41	0	0	0
ProcGr1_2:ProcGr2_2	-0.339	0	0	0	0
ProcGr1_2:ProcGr10_0	-0.112	0	0	0	0
ProcGr2_2:SghxGr3	-0.203	0	0	0	0
ProcGr9_2:RxGr4_1	0.013	-1.138	0	0	0
RxGr3_2:phydx4	-0.146	0	0	0	0
RxGr4_2:pastdx7	-0.142	0	0	0	0
RxGr7_2:mid_PDA	-0.033	0	0	0	0
RxGr4_1:pre_PDA	0.008	0	0	0	0
phydx4:pastdx4	0.354	0	0	0	0
SghxGr3:pre_PDA	0.013	0	0	0	0
mid_PDA:pre_PDA	0.003	0	0	0	0
RxGr5_1:SghxGr3	0.126	0	0	0	0
ProcGr10_0:phydx12	0.339	0	0	0	0
pastdx7:SghxGr3	0.213	0	0	0	0
ProcGr9_1	0	-0.121	-1.043	1.161	1.455
RxGr2_1	0	-0.188	0	0	0
RxGr2_0	0	0.843	0.469	0	0
pre_OSW	0	0.11	0	0	0
mid_OSW	0	0.35	0	0.045	0
pastdx6	0	1.311	0	1.242	-1.093
mid_sf36pcs	0	0.02	0	0.34	0.074
pastdx12	0	1.381	0.384	1.176	0
phydx6	0	1.46	1.567	0	-1.173
SghxGr5	0	3.898	0	0	0
onset	0	0.006	0	0	-0.122
ProcGr2_2:ProcGr9_1	0	-1.493	0	0	0

ProcGr4_2:phydx4	0	-0.987	0	0	0
RxGr1_2:SghxGr5	0	-0.544	0	0	0
RxGr2_2:race_1	0	0.089	0	0	0
RxGr3_2:RxGr2_1	0	-0.386	0	0	0
RxGr4_2:RxGr2_0	0	-0.608	0	0	0
RxGr7_2:phydx6	0	-1.283	0	0	0
mid_OSW:pre_OSW	0	0.007	0	0	0
phydx4:mid_OSW	0	0.076	0	0	0
pastdx12:SghxGr5	0	0.069	0	0	0
RxGr4_1:RxGr2_0	0	0.484	0	0	0
ProcGr10_0:phydx6	0	0.055	0.659	0	0
phydx6:SghxGr5	0	2.915	0	0	0
race_1:mid_OSW	0	0.009	0	0	0
onset:race_1	0	0.017	0	0	0
onset:pastdx12	0	0.004	0	0	0
phydx6:pre_OSW	0	0.05	0	0	0
RxGr2_0:pre_OSW	0	0.097	0	0	0
ProcGr10_0:pastdx12	0	0.634	0	0	0
ProcGr10_1	0	0	-0.355	0	0.649
RxGr8_0	0	0	0.335	0	0
phydx20	0	0	0.39	-1.524	0
SghxGr6	0	0	3.556	0	0
pre_BDI	0	0	0.071	0	0
status	0	0	0.044	0	0
phydx8	0	0	0.441	0	0
mid_BDI	0	0	0.26	0	0
ProcGr2_2:phydx6	0	0	-1.293	0	0
ProcGr9_2:ProcGr4_0	0	0	-1.469	0	0
ProcGr9_2:phydx20	0	0	-0.138	0	0
RxGr1_2:RxGr8_0	0	0	0.052	0	0
RxGr7_2:RxGr2_1	0	0	-1.199	0	0
RxGr8_2:RxGr7_0	0	0	0.527	0	0
ProcGr9_1:mid_BDI	0	0	-0.067	0	0
phydx8:SghxGr3	0	0	0.367	0	0
RxGr2_0:pastdx12	0	0	1.221	0	0
phydx6:pastdx7	0	0	0.543	0	0
ProcGr10_0:mid_BDI	0	0	-0.002	0	0
status:phydx6	0	0	0.075	0	0
pastdx12:SghxGr3	0	0	0.949	0	0
ProcGr11_0	0	0	0	-1.601	0
phydx31	0	0	0	-0.258	0

pastdx5	0	0	0	-2.496	0
gender	0	0	0	-0.191	0
pre_sf36pcs	0	0	0	0.12	0
ProcGr1_2:pastdx6	0	0	0	1.177	1.327
ProcGr2_2:phydx12	0	0	0	1.446	0
ProcGr9_2:SghxGr3	0	0	0	1.31	0
RxGr4_2:ProcGr4_1	0	0	0	0.953	0
RxGr7_2:mid_OSW	0	0	0	0.158	0
phydx12:pastdx6	0	0	0	-0.689	0
pastdx6:SghxGr3	0	0	0	-0.346	2.947
pastdx12:mid_sf36pcs	0	0	0	-0.158	0
ProcGr11_0:gender	0	0	0	-0.36	0
RxGr4_1:mid_sf36pcs	0	0	0	0.023	0
phydx20:pastdx12	0	0	0	-1.128	0
ProcGr4_1	0	0	0	0	0.505
ProcGr2_0	0	0	0	0	-1.366
pastdx4	0	0	0	0	-1.326
marital_1	0	0	0	0	-0.311
mid_sf36mcs	0	0	0	0	0.116
ProcGr2_2:ProcGr10_0	0	0	0	0	1.09
ProcGr9_2:mid_sf36mcs	0	0	0	0	0.073
RxGr7_2:phydx12	0	0	0	0	1.263
phydx12:pastdx4	0	0	0	0	-0.283
ProcGr10_0:onset	0	0	0	0	-0.054
mid_sf36mcs:marital_1	0	0	0	0	0.061
ProcGr9_1:phydx6	0	0	0	0	1.855
ProcGr2_0:ProcGr10_0	0	0	0	0	-1.137
pastdx4:pastdx6	0	0	0	0	-3.284

Table 41: Stage 1 Pain Outcome Transition Models 2 SP IQ

Variable	mid_PDA	mid_OSW	mid_BDI	mid_sf36pcs	mid_sf36mcs
Intercept	4.738	13.222	11.696	29.158	36.25
RxGr1_1	-0.179	-0.05	-0.05	-0.05	-0.05
RxGr2_1	-0.05	1.034	-0.05	-0.05	-0.05
RxGr3_1	-0.05	-0.05	-0.05	-0.05	-0.05
RxGr4_1	-0.852	-3.851	-1.329	3.315	2.75
RxGr5_1	-0.561	-2.977	-0.662	2.436	1.359
RxGr6_1	-0.05	-0.05	-0.05	-0.05	-0.05
RxGr7_1	-0.05	-0.05	0.136	-0.05	-1.443

RxGr8_1	-0.05	-0.05	-0.05	-3.607	-0.05
ProcGr1_1	-0.05	-0.05	0.125	-0.05	-1.295
ProcGr2_1	0.259	-0.05	-0.05	-0.378	-0.05
ProcGr4_1	-0.05	-0.05	-0.05	-0.152	-0.05
ProcGr9_1	-0.469	-2.352	-0.959	2.671	2.024
ProcGr10_1	0.103	-0.052	-0.05	-0.05	-0.05
ProcGr11_1	-0.05	-0.05	-0.05	-0.05	-0.05
phydx6	0.499	0	0	0	0
pastdx7	0.138	0	0	0	0
pre_PDA	0.34	0	0	0	0
age	0.002	0.042	0	0	0
RxGr3_0	0.103	0	0	0	0
RxGr4_0	-0.076	0	0	0	0
pastdx12	0.084	0	0	-0.902	-0.861
ProcGr9_1:ProcGr10_0	-0.079	0	-0.457	0	0
RxGr4_1:phydx6	-0.208	0	0	0	0
RxGr5_1:RxGr3_0	-0.151	0	0	0	0
RxGr5_1:pastdx7	-0.076	0	0	0	0
age:phydx6	0.048	0	0	0	0
RxGr3_0:pre_PDA	0.003	0	0	0	0
age:pre_PDA	0.004	0	0	0	0
RxGr4_0:pastdx12	0.013	0	0	0	0
RxGr4_0:pastdx7	-0.048	0	0	0	0
age:RxGr4_0	-0.003	0	0	0	0
RxGr3_0:pastdx12	0.172	0	0	0	0
pastdx12:pre_PDA	-0.025	0	0	0	0
phydx4	0	1.157	0	0	0
pastdx6	0	0.52	1.913	-1.42	0
marital_4	0	0.136	0	0	0
SghxGr3	0	0.132	0	-0.822	0
RxGr2_0	0	-0.198	0	0	0
pre_OSW	0	0.285	0	0.057	0
phydx31	0	1.634	0	0	-1.043
pastdx20	0	0.301	0	0	0
ProcGr9_1:RxGr2_0	0	-0.868	0	0	0
ProcGr10_1:SghxGr3	0	0.113	0	0	0
RxGr2_1:phydx4	0	0.276	0	0	0
RxGr4_1:pastdx6	0	-0.634	0	0.367	0
RxGr4_1:SghxGr3	0	-0.388	0	0	0
RxGr5_1:RxGr2_0	0	-0.289	0	0	0
RxGr2_0:pastdx6	0	0.775	0	0	0

pastdx6:SghxGr3	0	3.012	0	0	0
ProcGr10_0:race_1	0	-1.375	0	0	0
age:pre_OSW	0	0.091	0	0	0
SghxGr3:marital_4	0	1.97	0	0	0
age:phydx4	0	0.118	0	0	0
phydx31:SghxGr3	0	-0.023	0	0	0
pastdx20:marital_4	0	-0.141	0	0	0
RxGr2_0:pastdx20	0	0.207	0	0	0
phydx31:pastdx20	0	0.369	0	0	0
age:phydx31	0	0.392	0	0	0
RxGr1_0	0	0	0.893	0	0
marital_1	0	0	2.829	0	0
SghxGr11	0	0	0.49	0	0
ProcGr11_0	0	0	0	-0.78	0.043
ProcGr10_0	0	0	-0.372	0	0
pre_BDI	0	0	0.175	0	0
status	0	0	0.667	0	0
phydx20	0	0	2.307	0	0
ProcGr4_1:phydx20	0	0	-0.35	0	0
ProcGr9_1:pastdx6	0	0	-1.368	0	0
RxGr4_1:RxGr1_0	0	0	-0.242	0	0
ProcGr10_0:pastdx6	0	0	0.986	0	0
phydx20:SghxGr11	0	0	1.52	0	0
ProcGr11_0:pastdx6	0	0	1.155	0	0
phydx20:pastdx6	0	0	0.398	0	0
RxGr1_0:ProcGr11_0	0	0	0.202	0	0
status:ProcGr10_0	0	0	-0.043	0	0
phydx20:status	0	0	0.745	0	0
ProcGr10_0:phydx20	0	0	-0.243	0	0
RxGr1_0:marital_1	0	0	1.566	0	0
phydx12	0	0	0	-1.296	-0.089
gender	0	0	0	-0.408	0
ProcGr1_0	0	0	0	0.572	0
pre_sf36pcs	0	0	0	0.189	0
ProcGr11_1:pastdx6	0	0	0	-0.136	0
RxGr4_1:ProcGr1_0	0	0	0	0.562	0
RxGr5_1:phydx12	0	0	0	0.219	1.043
RxGr8_1:onset	0	0	0	-0.75	0
ProcGr11_0:SghxGr3	0	0	0	-0.948	0
pastdx12:SghxGr3	0	0	0	-0.482	0
ProcGr1_0:pastdx6	0	0	0	0.204	0

pre_OSW:pre_sf36pcs	0	0	0	0.007	0
gender:phydx12	0	0	0	-0.761	0
phydx12:SghxGr3	0	0	0	-0.279	0
ProcGr1_0:phydx20	0	0	0	0.196	0
gender:pastdx6	0	0	0	-0.045	0
pastdx6:pre_OSW	0	0	0	0.007	0
SghxGr5	0	0	0	0	-1.056
ProcGr9_0	0	0	0	0	0.053
pre_sf36mcs	0	0	0	0	0.144
race_1	0	0	0	0	-0.548
ProcGr9_1:race_1	0	0	0	0	1.407
phydx12:pre_sf36mcs	0	0	0	0	-0.618
ProcGr9_0:SghxGr5	0	0	0	0	0.789
pastdx12:marital_1	0	0	0	0	-0.829
RxGr8_0:phydx5	0	0	0	0	-0.83
ProcGr11_0:phydx31	0	0	0	0	-0.953
pastdx12:pre_sf36mcs	0	0	0	0	0.003
phydx12:SghxGr5	0	0	0	0	-0.23

Table 42: Stage 2 Pain Outcome Transition Models 2 SP IQ

Variable	post_PDA	post_OSW	post_BDI	post_sf36pcs	post_sf36mcs
Intercept	5.098	9.822	10.344	25.115	29.063
RxGr1_2	-0.05	-0.182	-0.05	-0.429	-0.05
RxGr2_2	-0.05	-0.05	-0.05	-0.05	-0.05
RxGr3_2	-0.05	0.254	-0.05	-0.05	-0.05
RxGr4_2	-0.05	-0.05	-0.05	-0.755	-0.05
RxGr5_2	-0.05	-0.05	-0.144	-0.05	-0.05
RxGr6_2	-0.05	0.013	-0.05	-0.05	-0.722
RxGr7_2	-1.117	-1.491	-2.758	2.029	2.163
RxGr8_2	0.195	-0.05	0.256	-0.05	-0.847
ProcGr1_2	-0.94	-1.027	-2.123	1.773	2.487
ProcGr2_2	-0.958	-1.331	-2.744	3.726	2.178
ProcGr4_2	-0.05	-0.05	-0.05	0.113	-0.05
ProcGr9_2	-1.058	-1.412	-2.981	2.749	2.923
ProcGr10_2	-0.05	-0.05	-0.05	-0.05	-0.684
RxGr1_1	-0.204	0	0	0.262	0
RxGr4_1	-0.387	-0.972	0	2.009	1.844
RxGr9_1	1.281	0	0	0	0
ProcGr10_0	-0.124	0	0.255	0	-0.297

RxGr2_0	0.794	0.013	0	0	-0.631
phydx6	-0.072	0	1.428	0	-1.923
phydx12	0.458	0	0	-1.867	0.772
pastdx7	0.773	0	2.648	0	-0.504
pastdx11	0.367	0	0	0	0
SghxGr3	0.716	0	0	-1.5	-0.235
mid_PDA	0.63	0.113	0	0.015	0
race_1	0.522	0	0	0	0
ProcGr1_2:ProcGr2_2	-0.259	0	0	0	0
ProcGr2_2:RxGr2_0	-0.361	0	0	0	0
ProcGr9_2:phydx6	-0.173	0	0	0	1.39
RxGr7_2:mid_PDA	-0.008	0	0	0	0
SghxGr3:mid_PDA	0.022	0	0	0	0
phydx12:pastdx7	0.301	0	0	0	0
RxGr1_1:pastdx11	-0.366	0	0	0	0
RxGr2_0:ProcGr10_0	0.779	0	0	0	-0.576
RxGr4_1:phydx6	-0.154	0	0	0	0
phydx6:SghxGr3	0.557	0	0	0	0
race__1:mid_PDA	0.018	0	0	0	0
ProcGr9_1	0	-1.7	-1.067	0	1.208
RxGr5_1	0	-0.839	0	1.197	0
ProcGr1_0	0	-0.042	0.313	0	0
mid_OSW	0	0.382	0	0	0
phydx4	0	1.07	0	0	0
phydx7	0	2.129	1.907	0	0
mid_sf36pcs	0	0.084	0	0.395	0
pastdx12	0	1.425	0	-0.918	0
pastdx4	0	0.043	0.837	-0.326	-1.622
SghxGr5	0	0.587	0	0	0
age	0	0.019	0	0	0
ProcGr1_2:ProcGr1_0	0	-0.503	-0.497	0	0
ProcGr2_2:ProcGr9_1	0	-0.609	0	0	0
ProcGr9_2:phydx4	0	-0.77	0	0	0
RxGr1_2:SghxGr5	0	-0.434	0	0	0
RxGr7_2:age	0	-0.306	0	0	0
pastdx4:pastdx12	0	0.315	0	0	0
phydx4:SghxGr5	0	1.237	0	0	0
age:mid_PDA	0	0.025	0	0	0
RxGr4_1:RxGr2_0	0	-0.909	0	0	0
ProcGr9_1:phydx7	0	-0.551	0	0	0
phydx4:mid_OSW	0	0.16	0	0	0

phydx7:pastdx4	0	2.375	0	0	0
age:SghxGr5	0	0.194	0	0	0
RxGr2_0:pastdx4	0	1.075	0	0	0
ProcGr10_1	0	0	-0.464	0	0.684
RxGr1_0	0	0	0.563	0	0
phydx20	0	0	1.49	-1.986	0
phydx31	0	0	2.634	0	-2.004
SghxGr1	0	0	2.281	0	0
SghxGr6	0	0	0.668	0	0
mid_BDI	0	0	0.383	0	0
status	0	0	0.036	0	0
ProcGr2_2:phydx7	0	0	-0.656	0	0
ProcGr9_2:ProcGr10_0	0	0	-0.513	0	0
RxGr7_2:phydx6	0	0	-0.483	0	0
phydx6:pastdx7	0	0	0.966	0	0
SghxGr1:mid_BDI	0	0	0.011	0	0
phydx7:SghxGr6	0	0	0.597	0	0
ProcGr10_0:mid_BDI	0	0	0.102	0	0
status:phydx6	0	0	0.033	0	0
ProcGr10_0:phydx31	0	0	0.917	0	0
pastdx7:SghxGr1	0	0	0.535	0	0
ProcGr11_0	0	0	0	-0.629	0
pastdx6	0	0	0	-0.629	0
pre_sf36pcs	0	0	0	0.186	0
onset	0	0	0	-0.041	0
ProcGr1_2:pastdx6	0	0	0	1.327	0
ProcGr9_2:SghxGr3	0	0	0	0.732	0
RxGr7_2:ProcGr11_0	0	0	0	1.297	0
RxGr7_2:phydx12	0	0	0	0.833	0
mid_sf36pcs:pre_sf36pcs	0	0	0	0.004	0
pastdx6:SghxGr3	0	0	0	-1.418	0
RxGr4_1:mid_sf36pcs	0	0	0	0.015	0
phydx20:pastdx12	0	0	0	-0.988	0
onset:SghxGr3	0	0	0	-0.218	0
pastdx4:pre_sf36pcs	0	0	0	0.013	0
marital_1	0	0	0	0	-0.103
pre_sf36mcs	0	0	0	0	0.046
mid_sf36mcs	0	0	0	0	0.143
ProcGr1_2:pastdx7	0	0	0	0	0.629
ProcGr2_2:ProcGr10_0	0	0	0	0	1.049
RxGr7_2:pre_sf36mcs	0	0	0	0	0.258

pastdx4:SghxGr3	0	0	0	0	-1.299
mid_sf36mcs:marital_1	0	0	0	0	0.023
ProcGr10_1:phydx12	0	0	0	0	-0.838
phydx6:pastdx4	0	0	0	0	-0.722
phydx31:SghxGr3	0	0	0	0	-0.932
phydx12:marital_1	0	0	0	0	-0.302

Appendix V

A section of the preliminary simulation case study conducted as part of the dissertation proposal is included here. We compare the modeling methods Glinternet, Elastic Net and HierNet on the feature selection metrics. Since Elastic Net builds only linear models without interactions, we compare Glinternet and HierNet on the interaction selection metrics. The experimental factors included the number of observations, number of variables, proportion of causal variables, correlation between causal variables, correlation between causal and spurious features, magnitude of coefficients, and the signal to noise ratio.

Elastic Net

Elastic Net is a regularization and variable selection method, where the elastic net penalty function is a convex combination of the LASSO and ridge penalty, shown in Equation (17) [144].

$$L(\lambda_1, \lambda_2, \beta) = |y - X\beta|^2 + \alpha|\beta|^2 + (1-\alpha)|\beta|_1, \quad (17)$$

where y is the response vector, X is the input matrix, β is elastic net estimator, $\alpha|\beta|^2 + (1-\alpha)|\beta|_1$ is the elastic net penalty, $\alpha \in [0, 1)$. When $\alpha = 1$, the elastic net penalty becomes a simple ridge regression penalty, and at $\alpha = 0$, it becomes a LASSO penalty. Elastic Net can only build linear models without interactions and is mainly used for feature selection in a correlated input space. The quadratic part of the penalty encourages the grouping effect, which means it can select groups of correlated variables in the model unlike LASSO, which selects just one variable from the correlated group and drops the others. Elastic Net is considered in this case study to compare the feature selection performance against the other two methods that model interactions.

Glinternet

Glinternet is a LASSO based method developed by Lim and Hastie [163] for learning pairwise interactions in a linear regression or logistic regression model that satisfies strong hierarchy. The formulation of Glinternet shown in Equation (18) is equivalent to a constrained overlapped group LASSO. Glinternet can

build models on continuous and categorical variables. As an example, consider two categorical variables F_1 and F_2 with L_1 and L_2 levels respectively. The indicator matrices for F_1 and F_2 are represented as X_1 and X_2 respectively.

$$\text{argmin}_{\mu,\alpha,\tilde{\alpha}} \frac{1}{2}\left\| Y - \mu - X_1\alpha_1 - X_2\alpha_2 - [X_1\, X_2\, X_{1:2}] \begin{bmatrix} \tilde{\alpha}_1 \\ \tilde{\alpha}_2 \\ \alpha_{1;2} \end{bmatrix} \right\|_2^2$$

$$+\lambda \left(\|\alpha_1\|_2 + \|\alpha_2\|_2 + \sqrt{L_2\|\tilde{\alpha}_1\|_2^2 + L_1\|\tilde{\alpha}_2\|_2^2 + \|\alpha_{1:2}\|_2^2} \right), \tag{18}$$

$$\text{subject to} \sum_{i=1}^{L_1} \alpha_1^i = 0, \quad \sum_{j=1}^{L_2} \alpha_2^j = 0, \quad \sum_{i=1}^{L_1} \tilde{\alpha}_1^i = 0, \quad \sum_{j=1}^{L_2} \tilde{\alpha}_2^i = 0$$

$$\text{and} \sum_{i=1}^{L_1} \alpha_{1:2}^{ij} = 0 \; for \; fixed \; j, \quad \sum_{j=1}^{L_2} \alpha_{1:2}^{ij} = 0 \; for \; fixed \; i,$$

where Y is the response vector, μ is the fixed intercept term, the main effect of variable F_1 and F_2 are $(\alpha_1 + \tilde{\alpha}_1)$ and $(\alpha_2 + \tilde{\alpha}_2)$ respectively. The interaction effect is $\alpha_{1;2}$, λ is the regularization parameter. The constrained overlapped group lasso is solved by using an equivalent unconstrained group lasso formulation. The term $\sqrt{L_2\|\tilde{\alpha}_1\|_2^2 + L_1\|\tilde{\alpha}_2\|_2^2 + \|\alpha_{1:2}\|_2^2}$ in the Glinternet penalty ensures that the estimates satisfy strong hierarchy. The main difference between Glinternet and HierNet is that Glinternet can handle both continuous and categorical variables, while HierNet can only accommodate continuous and binary variables [163]. Glinternet has a faster computation time and can handle large problems compared to HierNet. Glinternet identifies interactions strictly under strong hierarchy while HierNet can identify interactions under both strong and weak hierarchy.

Simulation Case Study

The simulation case study is designed to evaluate the performance of the modeling methods under different conditions. The true response variable, Y is generated by the Equation (19).

$$Y = \sum_{j=1}^{p} \beta_j X_j + \sum_{j=1}^{p-1} \sum_{k>j}^{p} \gamma_{jk} X_j X_k + \epsilon_0 \,, \tag{19}$$

where β_j is the main effect of input variable X_j, γ_{jk} is the interaction effect between variables X_j and X_k, p is the number of input variables, ϵ_0 is the error term, which is selected such that $Var(\epsilon_0) = \frac{Var(f(X))}{SNR}$, SNR is the signal to noise ratio and $f(X) = \sum_{j=1}^{p} \beta_j X_j + \sum_{j=1}^{p-1} \sum_{k>j}^{p} \gamma_{jk} X_j X_k$

The following are the factors considered in this simulation case study

1. **Number of input variables**: {10,20}
2. **Proportion of causal variables among all the input variables**: {0.3, 0.5}. The interactions simulated in the case study are based on strong hierarchy. Depending on the number of causal variables, a fixed number of interactions between randomly selected causal variables is generated. If the number of causal variables is less than or equal to three, one interaction term is generated and in all other cases three interactions are generated.
3. **Number of observations:** {150, 375}. The number of observations is further divided into training and testing dataset, with 2/3 of the observations used for training, and 1/3 for testing.
4. **Magnitude of co-efficients $\boldsymbol{\beta_j, \gamma_{jk}}$**: {[0.4, 0.6] and [0.8, 1.0]}.The magnitude of the co-efficients is a randomly generated number between 0.4 and 0.6 for level [0.4,0.6] and between 0.8 and 1 for level [0.8, 1.0].
5. **Correlation between causal and spurious variables:** {High, Medium, Low}. High correlation level has a correlation randomly generated between [0.7, 0.9]. Medium and low correlation has correlation randomly generated between [0.4, 0.6] and (0.0, 0.3] respectively.
6. **Correlation between causal variables:** {High, Medium, Low}.
7. **Signal to Noise Ratio (SNR):** {3, 9}. SNR levels are selected so that the proportion of variance explained by $f(X)$, referred to as $PVE(f)$ is 0.75 and 0.9. The relation between SNR and $PVE(f)$ is shown in Equation (20) [172].

$$PVE(f) = \frac{SNR}{(1+SNR)} \quad (20)$$

At SNR values of 3 and 9, the proportion of variance explained by $f(X)$ is 0.75 and 0.9 respectively. With the above factors and their corresponding levels, we have $\underline{\mathbf{2 \times 2 \times 2 \times 2 \times 3 \times 3 \times 2 = 288}}$ cases. For each simulated case, we generate 100 replications.

Simulation Results

Feature Selection sensitivity and specificity

High Correlation between causal and spurious variables

The average sensitivity, shown in Figure 49, can be used to compare the different methods on how accurately they classify the causal variables. The sensitivity is higher for low and medium correlation between causal variables but drops slightly at high correlation for Glinternet and HierNet. With Elastic Net, sensitivity increases from low to high correlation. HierNet Strong and HierNet Weak perform almost similarly and better than the other methods.

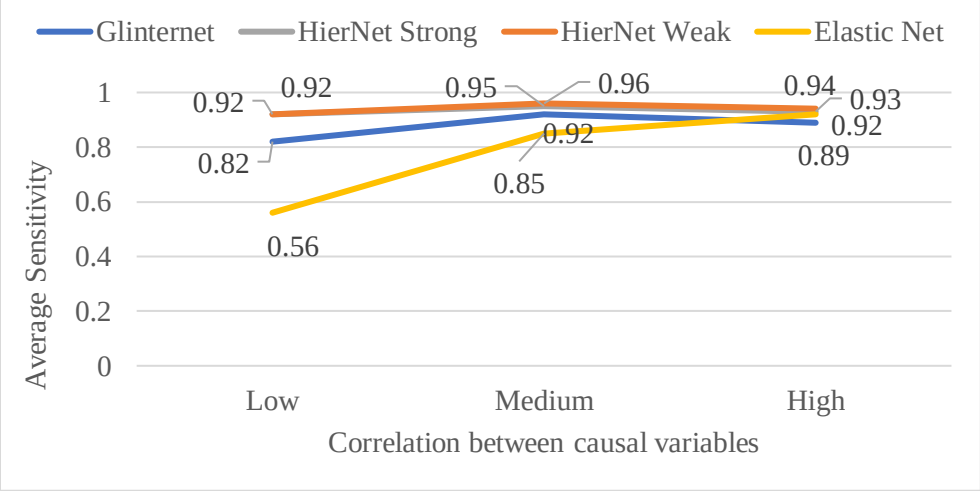


Figure 49: Average feature selection Sensitivity for high correlation between causal and spurious variables

The comparison of average specificity is shown in Figure 50. This shows if the methods are correctly classifying spurious variables. The specificity is low for the low and medium correlation cases indicating that due to the high correlation between causal and spurious variables, more spurious variables are selected in the predicted models. Glinternet performs better than the others in correctly classifying spurious variables in the predicted model.

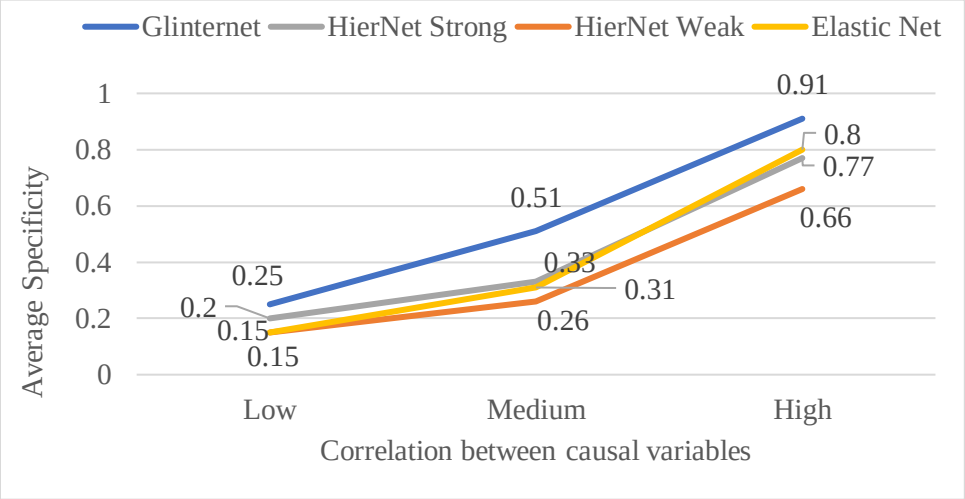


Figure 50 : Average feature selection Specificity for high correlation between causal and spurious variables

Medium Correlation between causal and spurious variables

The comparison of average sensitivity is shown in Figure 51. The average sensitivity for HierNet Strong and HierNet Weak is slightly better than the other methods at medium and high correlation, but is significantly better at low correlation. This implies HierNet based methods are better at classifying causal variables compared to the other methods. The average sensitivity for all the methods is lowest at low correlation and highest at medium correlation.

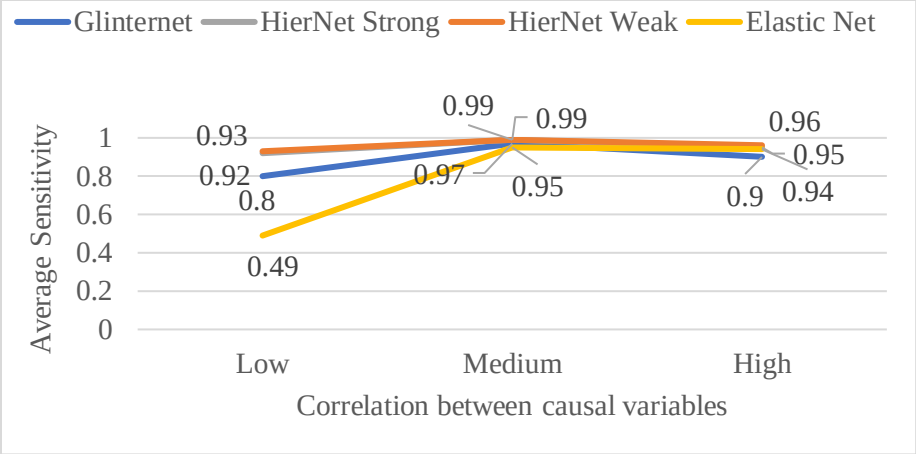


Figure 51: Average feature selection Sensitivity for medium correlation between causal and spurious variables

The comparison of average specificity in Figure 52, shows that Glinternet performs slightly better than the other methods on this metric. For high correlation between causal variables, the specificity is almost close to 1 for Glinternet and HierNet Strong, implying these methods correctly classify almost all the spurious variables. The specificity values are low at around 0.2 for low correlation between causal variables, implying spurious variables are selected in the predicted model due to the correlation between causal and spurious variables.

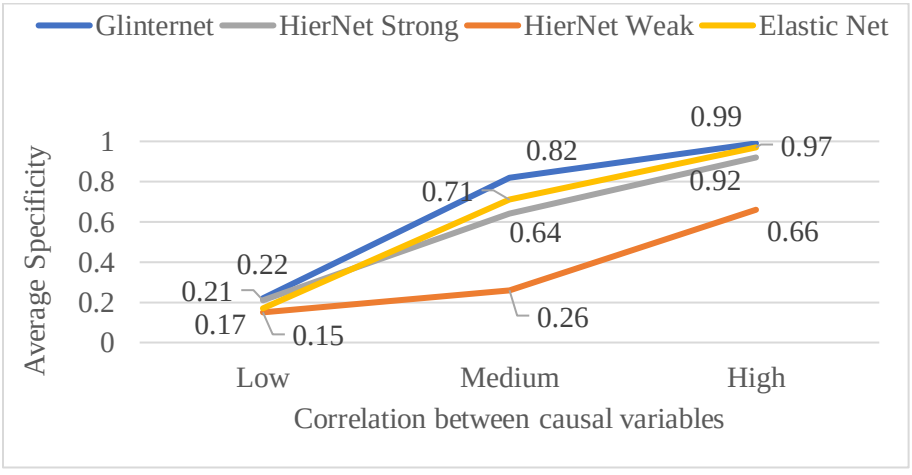


Figure 52: Average feature selection Specificity for medium correlation between causal and spurious variables

Low Correlation between causal and spurious variables

The average sensitivity is shown in Figure 53. The different methods correctly classify almost all the causal variables in the predicted model, with the sensitivity values ranging between 0.9 – 0.99. The sensitivity for Glinternet drops marginally when there is high correlation between the causal variables.

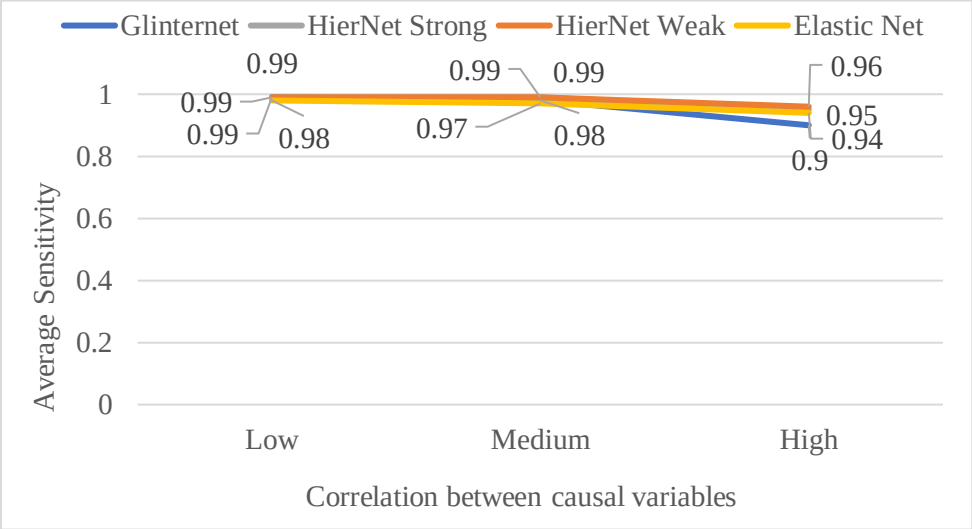


Figure 53: Average feature selection Sensitivity for low correlation between causal and spurious variables

The average specificity is shown in figure 54. The average specificity is very high and almost similar for all the methods when there is high correlation between causal variables. At low and medium correlation between causal variables, HierNet based methods perform poorly with low specificity values compared to Glinternet and Elastic Net, indicating they are selecting more spurious variables in the predicted model.

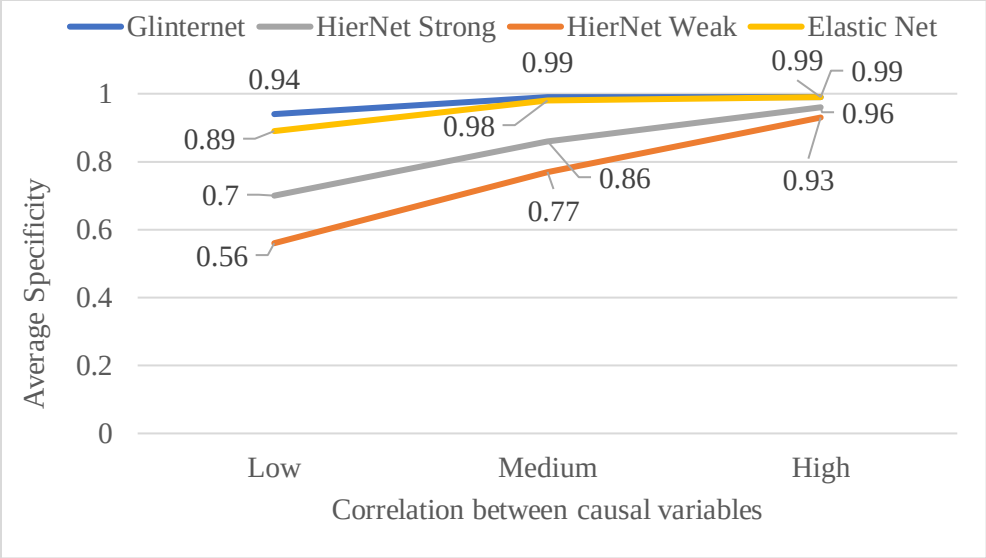


Figure 54: Average feature selection Specificity for low correlation between causal and spurious variables

Interaction Selection sensitivity and FDR

The interaction selection sensitivity and FDR are compared on the LASSO based methods that model interactions, namely Glinternet and HierNet. Elastic Net is not included in this analysis since it handles only linear additive terms.

High Correlation between causal and spurious variables

The average sensitivity plot is shown in Figure 55. HierNet based methods perform better than Glinternet at classifying true interactions. The sensitivity for all the methods drops at high correlation between causal variables.

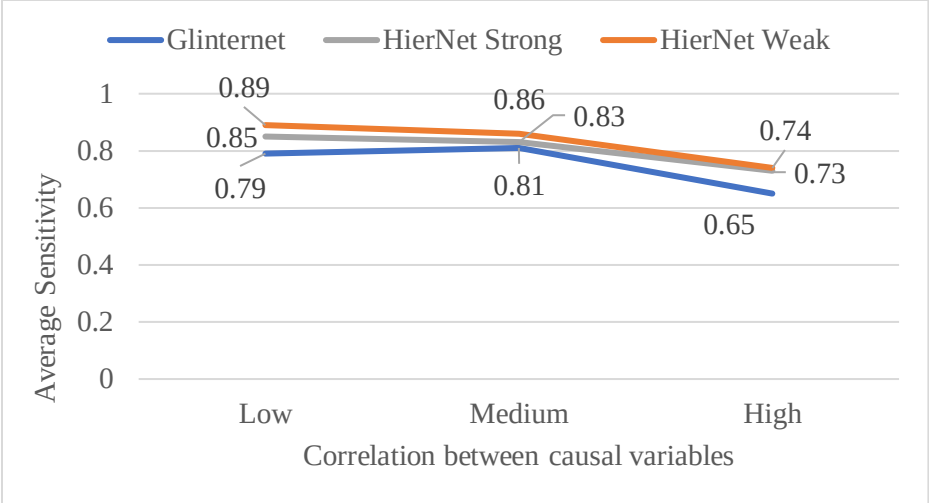


Figure 55: Average interaction selection Sensitivity for high correlation between causal and spurious variables

The False Discovery Rate (FDR) gives a picture of the number of spurious interactions in the predicted model. It is the proportion of spurious interactions among all the interactions in the predicted model. A lower average FDR value is better, as this indicates that the predicted model does not have many spurious interactions. From the Figure 56, we see that Glinternet has lower FDR values than the HierNet based methods. The FDR analysis is consistent with the specificity analysis since specificity is a measure of the number of spurious interactions correctly identified as spurious in the predicted model.

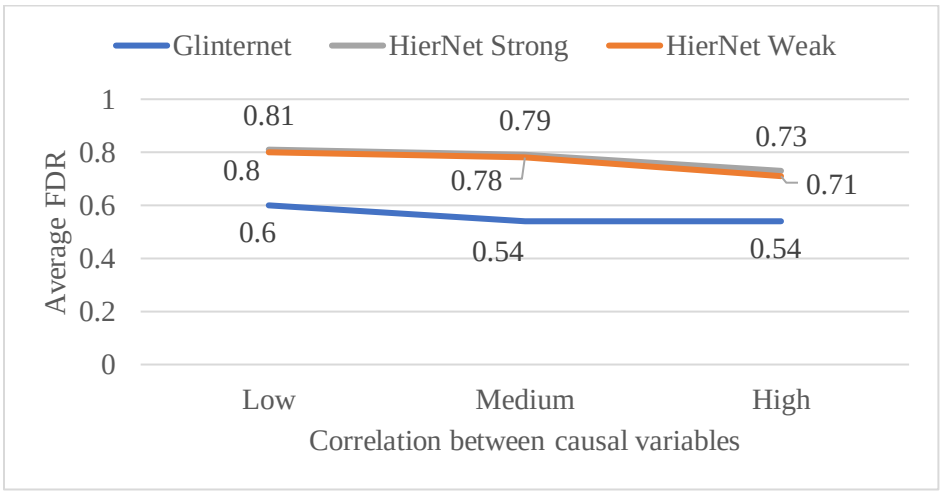


Figure 56: Average FDR for high correlation between causal and spurious variables

Medium Correlation between causal and spurious variables

The average sensitivity, shown in Figure 57, shows that HierNet Weak and HierNet Strong have better sensitivity than Glinternet. The sensitivity is least when correlation between causal variables is high. This shows that high correlation between the variables in the interaction term reduces the chances of correctly classifying them.

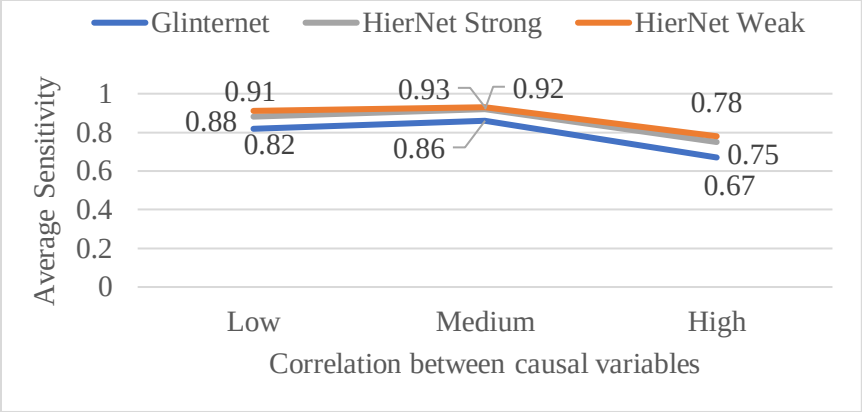


Figure 57: Average interaction selection Sensitivity for medium correlation between causal and spurious variables

The average FDR plot in Figure 58, shows that Glinternet performs better than HierNet. It has higher FDR value at high correlation than medium correlation. This can be explained by the sensitivity being low at high correlation, resulting in the proportion of spurious interactions in the predicted model to go up.

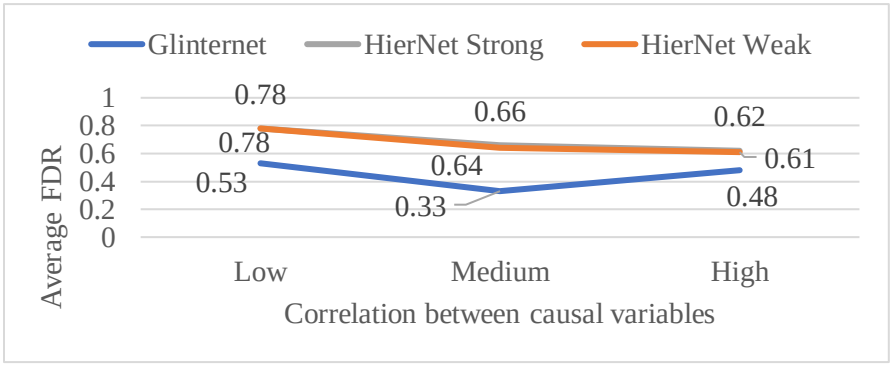


Figure 58: Average FDR for medium correlation between causal and spurious variables

Low Correlation between causal and spurious variables

The average sensitivity plot in Figure 59 shows similar trends as the average G-mean plot shown above, with the sensitivity values dropping at high correlation between causal variables, and HierNet based methods performing better than Glinternet.

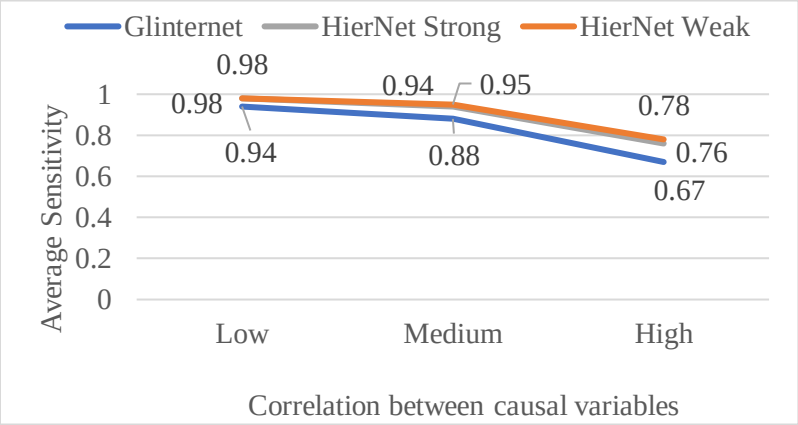


Figure 59: Average interaction selection Sensitivity for low correlation between causal and spurious variables

The average FDR in Figure 60 shows that Glinternet is better than the HierNet based methods, but its False Discovery Rate goes up from low to high correlation between causal variables, while HierNet based methods have almost similar FDR across the correlation levels.

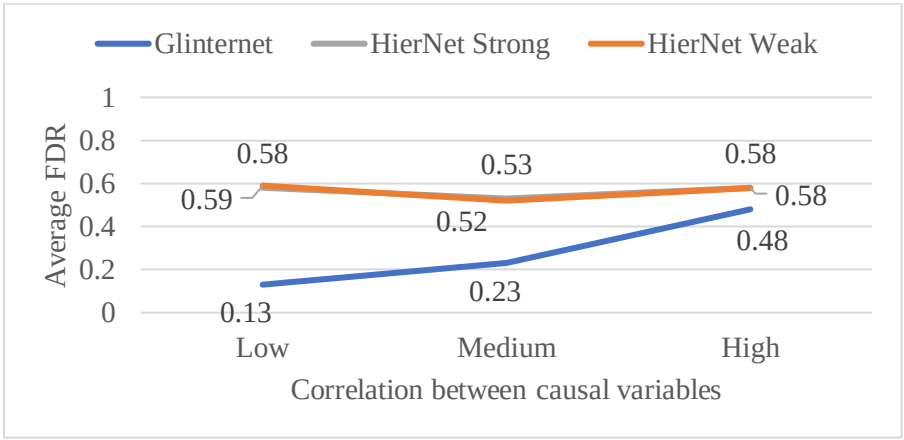


Figure 60: Average FDR for low correlation between causal and spurious variables

Discussion

The simulated case study had a number of factors, with the correlation structure between the variables being the important one. The feature selection performance of both Glinternet and HierNet were comparable or better than Elastic Net, which is the preferred feature selection method under multi collinearity. HierNet was better at correctly identifying the causal features and causal interactions, while Glinternet was better at correctly identifying the spurious features and spurious interactions. HierNet can handle both strong and weak hierarchy while Glinternet can handle only strong hierarchy. If the application places more importance on identifying the causal features and interactions than identifying the spurious ones, HierNet would be the better option, but if the application demands not including spurious terms in the model, Glinternet would be the better modeling method.

References

[1] Merskey H, Bogduk N. *Classification of chronic pain*. 2nd ed. Seattle: IASP Press, 1994. p. 1.

[2] Schappert SM, Burt CW. Ambulatory care visits to physician offices, hospital outpatient departments, and emergency departments: United States, 2001–02. *Vital Health Stat 13* 2006;13:1–66

[3] National Pain Strategy A comprehensive population health-level strategy for pain National Academies Press, Washington, DC, (2016)

[4] Von Korff, M., Scher, A. I., Helmick, C., Carter-Pokras, O., Dodick, D. W., Goulet, J. ... & Mackey, S. (2016). United States national pain strategy for population research: concepts, definitions, and pilot data. *The Journal of Pain*, *17*(10), 1068-1080.

[5] Dahlhamer, J., Lucas, J., Zelaya, C., Nahin, R., Mackey, S., DeBar, L., ... & Helmick, C. (2018). Prevalence of chronic pain and high-impact chronic pain among adults—United States, 2016. *Morbidity and Mortality Weekly Report*, *67*(36), 1001.

[6] Bevers, K., Watts, L., Kishino, N. D., & Gatchel, R. J. (2016). The biopsychosocial model of the assessment, prevention, and treatment of chronic pain. *US Neurol*, *12*(2), 98-104.

[7] Ojeda, B., Salazar, A., Dueñas, M., Torres, L. M., Micó, J. A., & Failde, I. (2014). The impact of chronic pain: The perspective of patients, relatives, and caregivers. *Families, Systems, & Health*, *32*(4), 399.

[8] Dueñas, M., Ojeda, B., Salazar, A., Mico, J. A., & Failde, I. (2016). A review of chronic pain impact on patients, their social environment and the health care system. *Journal of pain research*, *9*, 457.

[9] George, E., & Engel, L. (1980). The clinical application of the biopsychosocial model. *American journal of Psychiatry*, *137*(5), 535-544.

[10] Krashin, D., Sullivan, M., & Ballantyne, J. (2013). What are we treating with chronic opioid therapy?. *Current rheumatology reports*, *15*(3), 311.

[11] Deyo, R. A., Smith, D. H., Johnson, E. S., Donovan, M., Tillotson, C. J., Yang, X., ... & Dobscha, S. K. (2011). Opioids for back pain patients: primary care prescribing patterns and use of services. *The Journal of the American Board of Family Medicine*, *24*(6), 717-727.

[12] Eriksen, J., Sjøgren, P., Bruera, E., Ekholm, O., & Rasmussen, N. K. (2006). Critical issues on opioids in chronic non-cancer pain: An epidemiological study. *Pain*, *125*(1-2), 172-179.

[13] Potter, J. S., & Marino, E. N. (2013). How to avoid opioid misuse. *The Journal of family practice*, *62*(3 Suppl), S2-7.

[14] Rajaee, S. S., Bae, H. W., Kanim, L. E., & Delamarter, R. B. (2012). Spinal fusion in the United States: analysis of trends from 1998 to 2008. *Spine*, *37*(1), 67-76.

[15] Tarnanen, S., Neva, M. H., Dekker, J., Häkkinen, K., Vihtonen, K., Pekkanen, L., & Häkkinen, A. (2012). Randomized controlled trial of postoperative exercise rehabilitation program after lumbar spine fusion: study protocol. *BMC musculoskeletal disorders*, *13*(1), 1-7.

[16] Gatchel, R. J., McGeary, D. D., McGeary, C. A., & Lippe, B. (2014). Interdisciplinary chronic pain management: past, present, and future. *American psychologist*, *69*(2), 119.

[17] Younger, J., McCue, R., & Mackey, S. (2009). Pain outcomes: a brief review of instruments and techniques. *Current pain and headache reports*, *13*(1), 39-43.

[18] Fairbank, J. C., Couper, J., Davies, J. B., & O'brien, J. P. (1980). The Oswestry low back pain disability questionnaire. *Physiotherapy*, *66*(8), 271-273.

[19] Haefeli, M., & Elfering, A. (2006). Pain assessment. *European Spine Journal*, *15*(1), S17-S24.

[20] Beck, A. T., Steer, R. A., & Brown, G. (1996). Beck depression inventory–II. *Psychological Assessment*.

[21] Cech, D. J., & Martin, S. T. (2011). *Functional Movement Development Across the Life Span-E-Book*. Elsevier Health Sciences.

[22] Lin, C. F., LeBoulluec, A. K., Zeng, L., Chen, V. C., & Gatchel, R. J. (2014). A decision-making framework for adaptive pain management. *Health care management science*, *17*(3), 270-283.

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

[24] Robbins, H., Gatchel, R. J., Noe, C., Gajraj, N., Polatin, P., Deschner, M., ... & Adams, L. (2003). A prospective one-year outcome study of interdisciplinary chronic pain management: compromising its efficacy by managed care policies. *Anesthesia & Analgesia*, *97*(1), 156-162.

[25] LeBoulluec, A., Ohol, N., Chen, V., Zeng, L., Rosenberger, J., & Gatchel, R. (2018). Handling time-varying confounding in state transition models for dynamic optimization of adaptive interdisciplinary pain management. *IISE Transactions on Healthcare Systems Engineering*, *8*(1), 83-92.

[26] N. Ohol. "Adjusting for Time Varying Confounding in Adaptive Interdisciplinary Pain Management Program." Ph.D. dissertation, Dept IMSE, The University of Texas at Arlington, Arlington, TX, 2018.

[27] Iqbal, G. M. D., Rosenberger, J., Chen, V., Gatchel, R., & Noe, C. (2021). Multi-objective two-stage stochastic programming for adaptive interdisciplinary pain management with piecewise linear network transition models. *IISE Transactions on Healthcare Systems Engineering*, *11*(3), 240-254.

[28] Wang, N., Rosenberger, J., Iqbal, G. M. D., Chen, V., Gatchel, R. J., Noe, C., & LeBoulluec, A. K. (2019). Two-stage stochastic programming for interdisciplinary pain management. *IISE Transactions on Healthcare Systems Engineering*, *9*(2), 131-145.

[29] Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, *58*(1), 267-288.

[30] Bien, J., Taylor, J., & Tibshirani, R. (2013). A lasso for hierarchical interactions. *Annals of statistics*, *41*(3), 1111.

[31] Melzack, R., & Wall, P. D. (1965). Pain mechanisms: a new theory. *Science*, *150*(3699), 971-979.

[32] Simon, L. S. (2012). Relieving pain in America: A blueprint for transforming prevention, care, education, and research. *Journal of pain & palliative care pharmacotherapy*, *26*(2), 197-198.

[33] Seal, K., Becker, W., Tighe, J., Li, Y., & Rife, T. (2017). Managing chronic pain in primary care: it really does take a village. *Journal of general internal medicine*, *32*(8), 931-934.

[34] Centers for Disease Control and Prevention. CDC Guideline for Prescribing Opioids for Chronic Pain — United States, MMWR: Recommendations and Reports. March 18, 2016; 65(1):1-49.

[35] United States Department of Health and Human Services Interagency Pain Research Coordinating Committee. National Pain Strategy: A Comprehensive Population Health-Level Strategy for Pain. March 2016.

[36] Turk, D. C., Stanos, S. P., Palermo, T. M., Paice, J. A., Jamison, R. N., Gordon, D. B., & Clark, M. E. (2010). Interdisciplinary pain management. *Glenview, IL: American Pain Society*.

[37] Stanos, S. (2012). Focused review of interdisciplinary pain rehabilitation programs for chronic pain management. *Current pain and headache reports*, *16*(2), 147-152.

[38] American Pain Society White Paper: Interdisciplinary Pain Management, 2008.

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

[40] Turk, D. C. (2002). Clinical effectiveness and cost-effectiveness of treatments for patients with chronic pain. *The Clinical journal of pain*, *18*(6), 355-365.

[41] Gagnon, C. M., Stanos, S. P., van der Ende, G., Rader, L. R., & Harden, R. N. (2013). Treatment outcomes for workers compensation patients in a US-based interdisciplinary pain management program. *Pain Practice*, *13*(4), 282-288.

[42] JOHNSON, R., & MOORES, L. (2013). Pain management: integrating physiotherapy and clinical psychology practice. *Topical issues in pain 5*, 311.

[43] Collins, L. M., Murphy, S. A., & Strecher, V. (2007). The multiphase optimization strategy (MOST) and the sequential multiple assignment randomized trial (SMART): new methods for more potent eHealth interventions. *American journal of preventive medicine*, *32*(5), S112-S118.

[44] Pereira, M. V., & Pinto, L. M. (1991). Multi-stage stochastic optimization applied to energy planning. *Mathematical programming*, *52*(1), 359-375.

[45] Hamdy, M., Hasan, A., & Siren, K. (2013). A multi-stage optimization method for cost-optimal and nearly-zero-energy building solutions in line with the EPBD-recast 2010. *Energy and Buildings, 56*, 189-203.

[46] Mulvey, J. M., & Shetty, B. (2004). Financial planning via multi-stage stochastic optimization. *Computers & Operations Research, 31*(1), 1-20.

[47] Nickel, S., Saldanha-da-Gama, F., & Ziegler, H. P. (2012). A multi-stage stochastic supply network design problem with financial decisions and risk management. *Omega, 40*(5), 511-524.

[48] Zahiri, B., Torabi, S. A., Mohammadi, M., & Aghabegloo, M. (2018). A multi-stage stochastic programming approach for blood supply chain planning. *Computers & Industrial Engineering*, *122*, 1-14.

[49] Nagar, L., & Jain, K. (2008). Supply chain planning using multi-stage stochastic programming. *Supply Chain Management: An International Journal*, *13*(3), 251-256.

[50] Kazemi Zanjani, M., Nourelfath, M., & Ait-Kadi, D. (2010). A multi-stage stochastic programming approach for production planning with uncertainty in the quality of raw materials and demand. *International Journal of Production Research, 48*(16), 4701-4723.

[51] Yin, X., & Büyüktahtakın, İ. E. (2021). A multi-stage stochastic programming approach to epidemic resource allocation with equity considerations. *Health Care Management Science*, *24*(3), 597-622.

[52] Denardo, E. V. (2012). *Dynamic programming: models and applications*. Courier Corporation.

[53] Bellman, R. (1966). Dynamic programming. *Science*, *153*(3731), 34-37.

[54] Pardalos, P. M. (2009). Approximate dynamic programming: solving the curses of dimensionality.

[55] Powell, W. B. (2014). Clearing the jungle of stochastic optimization. In *Bridging data and decisions* (pp. 109-137). Informs.

[56] Yang, Z., Chen, V. C., Chang, M. E., Sattler, M. L., & Wen, A. (2009). A decision-making framework for ozone pollution control. *Operations Research*, *57*(2), 484-498.

[57] Chen, V. C., Ruppert, D., & Shoemaker, C. A. (1999). Applying experimental design and regression splines to high-dimensional continuous-state stochastic dynamic programming. *Operations Research*, *47*(1), 38-53.

[58] Nikzad, E., Bashiri, M., & Oliveira, F. (2019). Two-stage stochastic programming approach for the medical drug inventory routing problem under uncertainty. *Computers & Industrial Engineering*, *128*, 358-370.

[59] Fulton, L. V., Lasdon, L. S., McDaniel Jr, R. R., & Coppola, M. N. (2010). Two-stage stochastic optimization for the allocation of medical assets in steady-state combat operations. *The Journal of Defense Modeling and Simulation*, *7*(2), 89-102.

[60] Kim, K., & Mehrotra, S. (2015). A two-stage stochastic integer programming approach to integrated staffing and scheduling with application to nurse management. *Operations Research*, *63*(6), 1431-1451.

[61] Alfieri, A., Tolio, T., & Urgo, M. (2012). A two-stage stochastic programming project scheduling approach to production planning. *The International Journal of Advanced Manufacturing Technology*, *62*(1), 279-290.

[62] Zhou, Z., Zhang, J., Liu, P., Li, Z., Georgiadis, M. C., & Pistikopoulos, E. N. (2013). A two-stage stochastic programming model for the optimal design of distributed energy systems. *Applied Energy*, *103*, 135-144.

[63] Birge, J. R., & Louveaux, F. (1997). Two-Stage Linear Recourse Problems. *Introduction to Stochastic Programming*, 155-197.

[64] Shapiro, A., & Philpott, A. (2007). A tutorial on stochastic programming. *Manuscript. Available at www2. isye. gatech. edu/ashapiro/publications. html*, *17*.

[65] Mitra, S. (2006). A white paper on scenario generation for stochastic programming. *Optirisk systems: white paper series, ref. No. OPT004*.

[66] Mitra, S., & Domenica, N. D. (2010). A review of scenario generation methods. *International Journal of Computing Science and Mathematics*, *3*(3), 226-244.

[67] Beck, J. R., & Pauker, S. G. (1983). The Markov process in medical prognosis. *Medical decision making*, *3*(4), 419-458.

[68] Sonnenberg, F. A., & Beck, J. R. (1993). Markov models in medical decision making: a practical guide. *Medical decision making*, *13*(4), 322-338.

[69] Schaefer, A. J., Bailey, M. D., Shechter, S. M., & Roberts, M. S. (2005). Modeling medical treatment using Markov decision processes. In *Operations research and health care* (pp. 593-612). Springer, Boston, MA.

[70] Alagoz, O., Maillart, L. M., Schaefer, A. J., & Roberts, M. S. (2004). The optimal timing of living-donor liver transplantation. *Management Science*, *50*(10), 1420-1430.

[71] Shechter, S. M., Bailey, M. D., Schaefer, A. J., & Roberts, M. S. (2008). The optimal time to initiate HIV therapy under ordered health states. *Operations Research*, *56*(1), 20-33.

[72] Denton, B. T., Kurt, M., Shah, N. D., Bryant, S. C., & Smith, S. A. (2009). Optimizing the start time of statin therapy for patients with diabetes. *Medical Decision Making*, *29*(3), 351-367.

[73] Groot Koerkamp, B., Weinstein, M. C., Stijnen, T., Heijenbrok-Kal, M. H., & Hunink, M. M. (2010). Uncertainty and patient heterogeneity in medical decision models. *Medical Decision Making*, *30*(2), 194-205.

[74] Weinstein, M. C. (2006). Recent developments in decision-analytic modelling for economic evaluation. *Pharmacoeconomics*, *24*(11), 1043-1053.

[75] Rutter, C. M., Zaslavsky, A. M., & Feuer, E. J. (2011). Dynamic microsimulation models for health outcomes: a review. *Medical Decision Making*, *31*(1), 10-18.

[76] Siebert, U., Alagoz, O., Bayoumi, A. M., Jahn, B., Owens, D. K., Cohen, D. J., & Kuntz, K. M. (2012). State-transition modeling: a report of the ISPOR-SMDM modeling good research practices task force–3. *Medical Decision Making*, *32*(5), 690-700.

[77] Miller, A. B., Gribble, S., Nadeau, C., Asakawa, K., Flanagan, W. M., Wolfson, M., ... & Popadiuk, C. (2015). Evaluation of the natural history of cancer of the cervix, implications for prevention. The

Cancer Risk Management Model (CRMM)–Human papillomavirus and cervical components. *Journal of Cancer Policy*, *4*, 1-6.

[78] Van der Heijden, A. A. W. A., Feenstra, T. L., Hoogenveen, R. T., Niessen, L. W., de Bruijne, M. C., Dekker, J. M., ... & Nijpels, G. (2015). Policy evaluation in diabetes prevention and treatment using a population-based macro simulation model: the MICADO model. *Diabetic Medicine*, *32*(12), 1580-1587.

[79] Lay-Yee, R., & Cotterell, G. (2015). The role of microsimulation in the development of public policy. In *Policy Practice and Digital Science* (pp. 305-320). Springer, Cham.

[80] Rizopoulos, D. (2012). *Joint models for longitudinal and time-to-event data: With applications in R*. CRC press.

[81] Schulz, J., Kvaløy, J. T., Engan, K., Eftestøl, T., Jatosh, S., Kidanto, H., & Ersdal, H. (2020). State transition modeling of complex monitored health data. *Journal of Applied Statistics*, *47*(11), 1915-1935.

[82] Qian, T., Klasnja, P., & Murphy, S. A. (2020). Linear mixed models with endogenous covariates: modeling sequential treatment effects with application to a mobile health study. *Statistical science: a review journal of the Institute of Mathematical Statistics*, *35*(3), 375.

[83] Laird, N. and Ware, J. (1982). Random-effects models for longitudinal data. *Biometrics* **38,** 963‑974.

[84] Liang, K. Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, *73*(1), 13-22.

[85] Hedeker, D., & Gibbons, R. D. (2006). *Longitudinal data analysis* (Vol. 451). John Wiley & Sons.

[86] Ghisletta, P., & Spini, D. (2004). An introduction to generalized estimating equations and an application to assess selectivity effects in a longitudinal study on very old individuals. *Journal of Educational and Behavioral Statistics*, *29*(4), 421-437.

[87] Kvaløy, J. T., Skogvoll, E., Eftestøl, T., Gundersen, K., Kramer-Johansen, J., Olasveengen, T. M., & Steen, P. A. (2009). Which factors influence spontaneous state transitions during resuscitation?. *Resuscitation*, *80*(8), 863-869.

[88] Nordseth, T., Bergum, D., Edelson, D. P., Olasveengen, T. M., Eftestøl, T., Wiseth, R., ... & Skogvoll, E. (2013). Clinical state transitions during advanced life support (ALS) in in-hospital cardiac arrest. *Resuscitation*, *84*(9), 1238-1244.

[89] Huang, C. (2011). *Joint mean-covariance modelling and variable selection for longitudinal data analysis* (Doctoral dissertation, The University of Manchester (United Kingdom)).

[90] Dong, M., & He, D. (2007). Hidden semi-Markov model-based methodology for multi-sensor equipment health diagnosis and prognosis. *European Journal of Operational Research*, *178*(3), 858-878.

[91] Liu, Q., Dong, M., Lv, W., Geng, X., & Li, Y. (2015). A novel method using adaptive hidden semi-Markov model for multi-sensor monitoring equipment health prognosis. *Mechanical Systems and Signal Processing*, *64*, 217-232.

[92] Liu, Z., Wu, L., & Hauskrecht, M. (2013, May). Modeling clinical time series using gaussian process sequences. In *Proceedings of the 2013 SIAM International Conference on Data Mining* (pp. 623-631). Society for Industrial and Applied Mathematics.

[93] Imai, K., & Kim, I. S. (2019). When should we use unit fixed effects regression models for causal inference with longitudinal data?. *American Journal of Political Science*, *63*(2), 467-490.

[94] Guo, S., & Fraser, M. W. (2014). *Propensity score analysis: Statistical methods and applications* (Vol. 11). SAGE publications.

[95] Weitzen, S., Lapane, K. L., Toledano, A. Y., Hume, A. L., & Mor, V. (2004). Principles for modeling propensity scores in medical research: a systematic literature review. *Pharmacoepidemiology and drug safety*, *13*(12), 841-853.

[96] Imai, K., & Ratkovic, M. (2014). Covariate balancing propensity score. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *76*(1), 243-263.

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

[98] Daniel, R. M., Cousens, S. N., De Stavola, B. L., Kenward, M. G., & Sterne, J. A. C. (2013). Methods for dealing with time-dependent confounding. *Statistics in medicine*, *32*(9), 1584-1618.

[99] Fewell, Z., Hernán, M. A., Wolfe, F., Tilling, K., Choi, H., & Sterne, J. A. (2004). Controlling for time-dependent confounding using marginal structural models. *The Stata Journal*, *4*(4), 402-420.

[100] Bodnar, L. M., Davidian, M., Siega-Riz, A. M., & Tsiatis, A. 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.

[101] Sterne, J. A., & Tilling, K. (2002). G-estimation of causal effects, allowing for time-varying confounding. *The Stata Journal*, *2*(2), 164-182.

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

[103] Robins, J. M. (2002). Analytic methods for estimating HIV-treatment and cofactor effects. In *Methodological issues in AIDS behavioral research* (pp. 213-288). Springer, Boston, MA.

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

[105] Austin, P. C., & Stuart, E. A. (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*(28), 3661-3679.

[106] J. M. Robins, "Association, causation, and marginal structural models," *Synthese*, vol. 121, no. 1/2, pp: 151-179, 1999

[107] 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*(9-12), 1393-1512.

[108] Bray, B. C., Almirall, D., Zimmerman, R. S., Lynam, D., & Murphy, S. A. (2006). Assessing the total effect of time-varying predictors in prevention research. *Prevention Science*, *7*(1), 1-17.

[109] Thoemmes, F., & Ong, A. D. (2016). A primer on inverse probability of treatment weighting and marginal structural models. *Emerging Adulthood*, *4*(1), 40-59.

[110] Cole, S. R., & Hernán, M. A. (2008). Constructing inverse probability weights for marginal structural models. *American journal of epidemiology*, *168*(6), 656-664.

[111] De Bonet, J., Isbell Jr, C., & Viola, P. (1997). Mimic: Finding optima by estimating probability densities, in 'Advances in Neural Information Processing Systems-9'. *S*, *424*, 431.

[112] Chandrashekar, G., & Sahin, F. (2014). A survey on feature selection methods. *Computers & Electrical Engineering*, *40*(1), 16-28.

[113] Sivarajah, U., Kamal, M. M., Irani, Z., & Weerakkody, V. (2017). Critical analysis of Big Data challenges and analytical methods. *Journal of business research*, *70*, 263-286.

[114] Liu, H., & Motoda, H. (Eds.). (1998). *Feature extraction, construction and selection: A data mining perspective* (Vol. 453). Springer Science & Business Media.

[115] Matheus, C. J. (1991). The need for constructive induction. In *Machine Learning Proceedings 1991* (pp. 173-177). Morgan Kaufmann.

[116] Wnek, J., & Michalski, R. S. (1994). Hypothesis-driven constructive induction in AQ17-HCI: A method and experiments. *Machine Learning*, *14*(2), 139-168.

[117] Liu, H., & Motoda, H. (Eds.). (2007). *Computational methods of feature selection*. CRC Press.

[118] Sánchez-Marono, N., Alonso-Betanzos, A., & Tombilla-Sanromán, M. (2007, December). Filter methods for feature selection–a comparative study. In *International Conference on Intelligent Data Engineering and Automated Learning* (pp. 178-187). Springer, Berlin, Heidelberg.

[119] Langley, P. (1994, November). Selection of relevant features in machine learning. In *Proceedings of the AAAI Fall symposium on relevance* (Vol. 184, pp. 245-271).

[120] Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of machine learning research*, *3*(Mar), 1157-1182.

[121] Biesiada, J., & Duch, W. (2007). Feature selection for high-dimensional data—a Pearson redundancy based filter. In *Computer recognition systems 2* (pp. 242-249). Springer, Berlin, Heidelberg.

[122] Jin, X., Xu, A., Bie, R., & Guo, P. (2006, April). Machine learning techniques and chi-square feature selection for cancer classification using SAGE gene expression profiles. In *International Workshop on Data Mining for Biomedical Applications* (pp. 106-115). Springer, Berlin, Heidelberg.

[123] Liao, C., Li, S., & Luo, Z. (2006, November). Gene selection using Wilcoxon rank sum test and support vector machine for cancer classification. In *International Conference on Computational and Information Science* (pp. 57-66). Springer, Berlin, Heidelberg.

[124] Battiti, R. (1994). Using mutual information for selecting features in supervised neural net learning. *IEEE Transactions on neural networks*, *5*(4), 537-550.

[125] Liu, H., Sun, J., Liu, L., & Zhang, H. (2009). Feature selection with dynamic mutual information. *Pattern Recognition*, *42*(7), 1330-1339.

[126] John, G. H., Kohavi, R., & Pfleger, K. (1994). Irrelevant features and the subset selection problem. In *Machine Learning Proceedings 1994* (pp. 121-129). Morgan Kaufmann.

[127] Tangaro, S., Amoroso, N., Brescia, M., Cavuoti, S., Chincarini, A., Errico, R., ... & Bellotti, R. (2015). Feature selection based on machine learning in MRIs for hippocampal segmentation. *Computational and mathematical methods in medicine*, *2015*.

[128] Remeseiro, B., Bolón-Canedo, V., Alonso-Betanzos, A., & Penedo, M. G. (2015). Learning features on tear film lipid layer classification. In *European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning* (pp. 195-200).

[129] Park, C., & Cho, S. B. (2003, July). Genetic search for optimal ensemble of feature-classifier pairs in DNA gene expression profiles. In *Proceedings of the International Joint Conference on Neural Networks, 2003.* (Vol. 3, pp. 1702-1707). IEEE.

[130] Li, T., Zhang, C., & Ogihara, M. (2004). A comparative study of feature selection and multiclass classification methods for tissue classification based on gene expression. *Bioinformatics*, *20*(15), 2429-2437.

[131] Hernández-Pereira, E., Bolón-Canedo, V., Sánchez-Maroño, N., Álvarez-Estévez, D., Moret-Bonillo, V., & Alonso-Betanzos, A. (2016). A comparison of performance of K-complex classification methods using feature selection. *Information Sciences*, *328*, 1-14.

[132] Phinyomark, A., N Khushaba, R., & Scheme, E. (2018). Feature extraction and selection for myoelectric control based on wearable EMG sensors. *Sensors*, *18*(5), 1615.

[133] Kohavi, R., & John, G. H. (1997). Wrappers for feature subset selection. *Artificial intelligence*, *97*(1-2), 273-324.

[134] Wang, H., Das, S. R., Suh, J. W., Altinay, M., Pluta, J., Craige, C., ... & Alzheimer's Disease Neuroimaging Initiative. (2011). A learning-based wrapper method to correct systematic errors in automatic image segmentation: consistently improved performance in hippocampus, cortex and brain segmentation. *NeuroImage*, *55*(3), 968-985.

[135] González, J., Ortega, J., Damas, M., Martín-Smith, P., & Gan, J. Q. (2019). A new multi-objective wrapper method for feature selection–Accuracy and stability analysis for BCI. *Neurocomputing*, *333*, 407-418.

[136] Muštra, M., Grgić, M., & Delač, K. (2012). Breast density classification using multiple feature selection. *automatika*, *53*(4), 362-372.

[137] Subashini, T. S., Ramalingam, V., & Palanivel, S. (2010). Automated assessment of breast tissue density in digital mammograms. *Computer Vision and Image Understanding*, *114*(1), 33-43.

[138] Jović, A., Brkić, K., & Bogunović, N. (2015, May). A review of feature selection methods with applications. In *2015 38th international convention on information and communication technology, electronics and microelectronics (MIPRO)* (pp. 1200-1205). IEEE.

[139] Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1993). Classification and Regression Trees. Belmont, CA: Wadsworth Inc.

[140] Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5-32.

[141] Meinshausen, N., & Buhlmann, P. (2004, May). Consistent neighbourhood selection for sparse high-dimensional graphs with the Lasso. Seminar für Statistik, Eidgenössische Technische Hochschule (ETH), Zürich.

[142] Tibshirani, R. (2011). Regression shrinkage and selection via the lasso: a retrospective. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *73*(3), 273-282.

[143] Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American statistical association*, *101*(476), 1418-1429.

[144] Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the royal statistical society: series B (statistical methodology)*, *67*(2), 301-320.

[145] Yuan, M., & Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *68*(1), 49-67.

[146] Wang, H., & Leng, C. (2008). A note on adaptive group lasso. *Computational statistics & data analysis*, *52*(12), 5277-5286.

[147] Jacob, L., Obozinski, G., & Vert, J. P. (2009, June). Group lasso with overlap and graph lasso. In *Proceedings of the 26th annual international conference on machine learning* (pp. 433-440).

[148] Zhang, H., Wang, J., Sun, Z., Zurada, J. M., & Pal, N. R. (2019). Feature selection for neural networks using group lasso regularization. *IEEE Transactions on Knowledge and Data Engineering*, *32*(4), 659-673.

[149] Wang, J., Zhang, H., Wang, J., Pu, Y., & Pal, N. R. (2020). Feature selection using a neural network with group lasso regularization and controlled redundancy. *IEEE transactions on neural networks and learning systems*.

[150] Meinshausen, N., & Bühlmann, P. (2006). High-dimensional graphs and variable selection with the lasso. *Annals of statistics*, *34*(3), 1436-1462.

[151] Friedman, J., Hastie, T., & Tibshirani, R. (2010). *Applications of the lasso and grouped lasso to the estimation of sparse graphical models* (pp. 1-22). Technical report, Stanford University.

[152] Farahani, A. A., Viswanatha, A., Chen, V. C., & Rosenberger, J. M. (2021). LASSO Based Causal Variable Selection Addressing Time-Varying Confounding in Interdisciplinary Pain Management. In *IIE Annual Conference. Proceedings* (pp. 1028-1033). Institute of Industrial and Systems Engineers (IISE).

[153] Hao, N., & Zhang, H. H. (2014). Interaction screening for ultrahigh-dimensional data. *Journal of the American Statistical Association*, *109*(507), 1285-1301.

[154] Zhao, J., & Leng, C. (2016). An analysis of penalized interaction models. *Bernoulli*, *22*(3), 1937-1961.

[155] Gunter, L., Zhu, J., & Murphy, S. (2007, July). Variable selection for optimal decision making. In *Conference on Artificial Intelligence in Medicine in Europe* (pp. 149-154). Springer, Berlin, Heidelberg.

[156] Nelder, J. A. (1977). A reformulation of linear models. *Journal of the Royal Statistical Society: Series A (General)*, *140*(1), 48-63.

[157] Peixoto, J. L. (1987). Hierarchical variable selection in polynomial regression models. *The American Statistician*, *41*(4), 311-313.

[158] Chipman, H. (1996). Bayesian variable selection with related predictors. *Canadian Journal of Statistics*, *24*(1), 17-36.

[159] Wu, J., Devlin, B., Ringquist, S., Trucco, M., and Roeder, K. (2010), "Screen and Clean: A Tool for Identifying Interactions in Genome-Wide Association Studies," *Genetic Epidemiology*, 34, 275–285. [1286, 1290]

[160] Park, M. Y., & Hastie, T. (2008). Penalized logistic regression for detecting gene interactions. *Biostatistics*, *9*(1), 30-50.

[161] Zhao, P., Rocha, G., & Yu, B. (2009). The composite absolute penalties family for grouped and hierarchical variable selection. *The Annals of Statistics*, *37*(6A), 3468-3497.

[162] Choi, N. H., Li, W., & Zhu, J. (2010). Variable selection with the strong heredity constraint and its oracle property. *Journal of the American Statistical Association*, *105*(489), 354-364.

[163] Lim, M., & Hastie, T. (2015). Learning interactions via hierarchical group-lasso regularization. *Journal of Computational and Graphical Statistics*, *24*(3), 627-654.

[164] Chipman, H. (1996). Bayesian variable selection with related predictors. *Canadian Journal of Statistics*, *24*(1), 17-36.

[165] George, E. I., & McCulloch, R. E. (1993). Variable selection via Gibbs sampling. *Journal of the American Statistical Association*, *88*(423), 881-889.

[166] Yuan, M., Joseph, V. R., & Lin, Y. (2007). An efficient variable selection approach for analyzing designed experiments. *Technometrics*, *49*(4), 430-439.

[167] Radchenko, P., & James, G. M. (2010). Variable selection using adaptive nonlinear interaction structures in high dimensions. *Journal of the American Statistical Association*, *105*(492), 1541-1553.

[168] A.A. Farahani. "Uncovering Underlying Features for State Transition Modeling." Ph.D. dissertation, Dept IMSE, The University of Texas at Arlington, Arlington, TX, 2019.

[169] Tan, P. N., Steinbach, M., & Kumar, V. (2016). *Introduction to data mining*. Pearson Education India.

[170] Kubat, M., Holte, R. C., & Matwin, S. (1998). Machine learning for the detection of oil spills in satellite radar images. *Machine learning*, *30*(2), 195-215.

[171] Chong, I. G., & Jun, C. H. (2005). Performance of some variable selection methods when multicollinearity is present. *Chemometrics and intelligent laboratory systems*, *78*(1-2), 103-112.

[172] Hastie, T., Tibshirani, R., & Tibshirani, R. J. (2017). Extended comparisons of best subset selection, forward stepwise selection, and the lasso. *arXiv preprint arXiv:1707.08692*.