Investigating the Impact of Multilevel Treatment Options on Statin Patient Outcomes: A Counterfactual Analysis

A THESIS
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Yue Liang (Lynnc)

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
Doctor of Philosophy

Committee Chair: Dr. Sisi Ma.
Committee Members: Dr. Chih-Lin Chi, Dr. Julian Wolfson, Dr. Terrence Adam

November, 2024

Acknowledgements

There are many people that have earned my gratitude for their contribution to my time in graduate school. I'm really thankful to my primary advisor Dr. Chih-Lin Chi for all of the support and understanding through the past several years. I also want to thank all of the committee members including Dr. Sisi Ma, Dr. Julian Wolfson, and Dr. Terrence Adam for their invaluable advice and support throughout the development of this dissertation.

Dedication

To those who held me up over the years.

Abstract

Statins are among the most commonly prescribed medications for lowering cardiovascular disease risk by reducing LDL-C levels. However, statin treatments can lead to several clinically significant side effects, such as myopathy and liver problems. Therefore, a machine learning (ML) system integrated with electronic health records (EHR) is essential for making personalized treatment recommendations. This system would consider medication side effects, treatment efficacy, and the risk of discontinuation, while addressing confounding factors inherent in observational data. Confounding occurs when conclusions like "Drug A is better than Drug B for alleviating symptoms" are drawn without fully accounting for factors that influence both treatment and outcomes. By mitigating these biases, ML models can assist doctors in making more informed prescription decisions, such as predicting what might happen if a patient receives a lower dosage.

This thesis explores the combination of machine learning and causal inference techniques to develop counterfactual (CF) predictions for statin-related outcomes. Specifically, it focuses on predicting statin-associated symptoms (SAS) and other relevant outcomes in real-world settings. Additionally, the thesis includes a counterfactual outcome simulation for both baseline CF models and two-stage CF model, which is used to evaluate the accuracy of these predictions. Finally, the thesis presents discussions on the broader applications of counterfactual predictions in clinical decision-making.

Contents

Acknowledgements i

Dedication ii

Abstract iii

List of Tables vii

List of Figures viii

1 Introduction 1

1.1 Background 1

1.2 Motivations of Counterfactual Prediction under Multiple Treatment Settings 4

1.3 Significance of Counterfactual Predictions 5

2 Literature Review 7

2.0.1 Counterfactual Prediction in Medical Research 8

2.0.2 Applications of Counterfactual Prediction in Medicine 8

2.0.3 Overlap Weights and Doubly Robust Estimation 9

2.0.4 Two-Stage Models in Treatment Effect Estimation 9

2.0.5 Challenges and Future Directions 10

2.0.6 Gaps in Current Literature 10

2.0.7 Need for Two-Stage Models in Complex Treatment Scenarios 11

2.0.8 Future Research Directions 12

2.0.9 Significance of the Thesis 12

2.0.10 Challenges and Future Directions 13

3 Methodology 14

3.1 Introduction 14

3.1.1 Data Extraction and Preprocessing 16

3.1.2 Model Development and Prediction . . . 16
3.1.3 Optimization and Clinical Trial Simulation . . . 16
3.1.4 Data Simulation for Model Evaluation . . . 17
3.2 Data . . . 17
3.3 Model . . . 19
3.3.1 Baseline Counterfactual Model: Counterfactual prediction using Overlap Weights 19
3.3.2 Two-stage Counterfactual Model: Propensity for Statin agent and counterfactual propensity for Statin intensity level . . . 21
3.3.3 Reversed Order Two-stage (TSRO) counterfactual model: Reverse the order of Statin agents and Statin dosage/intensity levels . . . 22
3.4 Clinical Trial Simulation . . . 23
3.5 Simulation . . . 24

4 Simulation **25**
4.1 Introduction . . . 25
4.1.1 Baseline Model Overview . . . 26
4.1.2 Two-Stage Model Overview . . . 27
4.1.3 Simulation Goals and Evaluation . . . 27
4.2 Methods . . . 27
4.2.1 Simulation Design . . . 27
4.2.2 Simulation steps . . . 28
4.2.3 Simulation Scenarios . . . 35
4.2.4 Goals and Evaluation Metrics . . . 36

5 Results **37**
5.1 Result from Optum Labs data . . . 37
5.1.1 Data . . . 37
5.1.2 Consort diagram . . . 37
5.1.3 Balancing check . . . 37
5.1.4 Overlap check . . . 38
5.1.5 Counterfactual prediction . . . 38
5.1.6 Optimization showcase . . . 38
5.1.7 Clinical Trial Simulation . . . 39
5.2 Result from AHC-IE data . . . 39
5.2.1 Data . . . 39
5.2.2 Clinical Trial Simulation . . . 40
5.3 Result from Simulations . . . 40

5.3.1 Comparison between ground Truth and predicted treatment effects 42

6 Discussion **43**

6.1 Model: Baseline model 43
6.2 Model: Two-stage model 45
6.3 Optum Labs data 46
6.4 AHC-IE data 48
6.5 Simulation 49
6.6 Significance 51
6.7 Limitation 52

7 Conclusion **54**

References **56**

Appendix A. Supplementary Figures and Tables **60**

A.1 Tables 60
A.2 Figures 80
A.2.1 Consort Diagram 80
A.2.2 Balancing check 81
A.2.3 Overlap check 82
A.2.4 SAS/Discontinuation counterfactual survival prediction 83
A.2.5 3d Multi-Objective Optimization showcase 84
A.3 Algorithms 93
A.3.1 Algorithm: Data Simulation for Additive Treatment Effects (No Interaction Terms) 93
A.3.2 Algorithm: Data Simulation for Overlapped Treatment Effects (No Interaction Terms) 96
A.3.3 Algorithm: Data Simulation for Overlapped Treatment Effects with Interaction Terms 98

List of Tables

A.1 OptumLabs data cohort: Subset of Baseline covariates across different Statin types(each cell is mean(SD) or proportion of each column;Total N is defined in first row) 61
A.2 AHC-IE data cohort: Baseline covariates across different Statin types 62
A.3 AHC-IE data cohort: Baseline covariates across different Statin intensity 73

List of Figures

A.1 Consort Diagram of OptumLabs data with inclusion/exclusion criteria 80
A.2 Consort Diagram of UMN Fairview data with inclusion/exclusion criteria 81
A.3 ASD and PSD of the sorted covariates 81
A.4 ASD and PSD of the sorted covariates 82
A.5 Weighted and unweighted distribution for age variale at index date 82
A.6 SAS/Discontinuation counterfactual survival prediction 83
A.6 Example plot of 3D Pareto Optimization for selecting out PSTP treatment for a simulated patient for 30 days to 360 days survival risks of SAS, discontinuation, and LDL-C reduction. Purple dot means that it is pareto front options. Orange dot is worse choice . . 85
A.7 Summarization of the CTS result with multiple survival days for PSTP predictions result 86
A.8 Summarised SAS/Disc/LDL reduction mean change overtime 86
A.9 Summarization of the CTS result using AHC-IE data with baseline model PSTP 87
A.10 Summarization of the CTS result using AHC-IE data with two-stage model PSTP 87
A.11 Summarised SAS/Disc/LDL reduction mean change overtime 88
A.12 Mean of model performance across all treatment options, for Additive Treatment effects with NO interactions terms 88
A.13 Mean of model performance across all treatment options, for Additive Treatment effects with interactions terms 89
A.14 Mean of model performance across all treatment options, for Overlapped Treatment effects with NO interactions terms 89
A.15 Mean of model performance across all treatment options, for Overlapped Treatment effects with interactions terms 90
A.16 4 treatments examples of model performance across all treatment options, for Additive Treatment effects with No interactions terms 90
A.17 4 treatments examples of model performance across all treatment options, for Additive Treatment effects with interactions terms 91
A.18 4 treatments examples of model performance across all treatment options, for Overlapped Treatment effects with No interactions terms 91

A.19 4 treatments examples of model performance across all treatment options, for Overlapped Treatment effects with interactions terms 92

Chapter 1

Introduction

1.1 Background

Statins are a class of drugs that inhibit cholesterol synthesis by blocking hydroxymethylglutaryl coenzyme-A (HMG-CoA) reductase, an enzyme involved in cholesterol production. They are widely prescribed for patients with high cholesterol levels or at risk of cardiovascular disease, as they have been shown to reduce the incidence of myocardial infarction, stroke, and mortality [1]. However, statins have the potential to cause muscle pain, liver damage, renal failure, and rhabdomyolysis, collectively referred to as statin-associated symptoms (SAS). Some other lieratures pointed out some other specific statin related side effects including myositis [2]. While recent work has called into question the degree to which SAS are due to the medication versus nocebo effects [3], it remains true that concerns about SAS can reduce adherence and lead to discontinuation of statin therapy [1]. About 50 % of patients tend to discontinue their statin prescription within one year of treatment initiation. [4]. Such discontinuation lost the ASCVD-prevention for patients. To minimize the impact of this problem, in our previous study [3], we proposed a proof-of-concept tool to produce a proactive strategy called Personalized Statin Treatment Plan (PSTP), which utilizes a machine-learning approach to personalize the treatment plan (specific statin agent and dosage) with the minimal SAS and therapy discontinuation risks. Chi et al. [5] also utilized the machine learning techniques with EHR data to try to predict the Statin medications outcomes, including SAS and Statin discontinuation. After the prediction is done, the prediction results are incorporated together to select the optimal personal treatment plan for each patient using 2-dimensional optimization.

When it comes to clinical outcome predictions of statin medications, relying solely on machine learning for predicting statin outcomes using observational data may pose risks if the issue of confounding under various treatment settings is not adequately addressed. Moreover, considering the multitude of statin treatment options and dosage levels available, tailored methodologies for counterfactual predictions may be necessary. There are three crucial aspects that were overlooked in our previous PSTP

model: the failure to account for confounding bias, the absence of balance between benefits and risks in the optimization process—including the balance between LDL-C reduction benefits and risks of adverse events and discontinuation—and the lack of evaluation of the proactive prescribing strategy (PSTP) in retrospective data. Addressing these aspects is pivotal for advancing the PSTP toward practical implementation, and this part of this thesis aims to tackle them.

Hence, employing both machine learning techniques and causal inference methods is imperative for making counterfactual predictions for Statin patients. We know that machine learning (ML) models have gained widespread adoption in medical research over the past decade, ranging from basic models like K-nearest neighbor to advanced ones such as recurrent neural networks and generative adversarial neural networks. These models play significant roles in analyzing medical data, whether structured or unstructured. Additionally, beyond their application, it's vital to consider the assumptions, definitions, explainability, and interpretability of these models in medical research. Relying solely on machine learning or black-box models for treatment/exposure predictions may necessitate stronger assumptions and interpretations. Hence, researchers must remain cognizant of confounding issues when applying machine learning models to observational data and drawing conclusions like "Drug A is superior to Drug B in alleviating disease pain." This underscores the importance of embracing the concept of Causal Inference.

Causal inference, a statistical discipline, focuses on inferring treatment effects using observational data. When asserting the superiority of one drug (treatment) over another (control), the gold standard method is conducting a randomized control trial (RCT). In an RCT, subjects have an equal chance of being assigned to either treatment or control groups, thus minimizing confounding and selection biases induced by propensity. This allows for causal inference of treatment effects without these issues. However, RCTs can be costly and sometimes impractical. In such cases, researchers may turn to observational data and medical records to gather real-world evidence to support their findings. Utilizing electronic health record (EHR) data, causal inference enables researchers to infer the causal effects of policies or treatments without the need for an RCT.

Furthermore, in many cases, the treatment arm isn't simply binary. The Variational Sample Reweighting algorithm [6], designed for treatment bundle counterfactual prediction, is particularly relevant in situations where there are more than two treatment options. Multilevel treatment choices involving various dosages, brands of drugs, and combinations of drugs or auxiliary medications are quite common. A prime example is the Statin drug, which offers numerous types and dosage combinations of medications. It also provides multiple dosage levels corresponding to varying intensities of the drug. With the combination of different drug types, Statin users have access to dozens of specific drug options. Hence, a causal inference or counterfactual prediction model capable of accommodating scenarios with multiple treatment arms becomes essential.

Besides the causal inference models, we also need to tailor the estimand when making counterfactual

predictions. The estimand of the causal inference model is typically the Average Treatment Effect (ATE), also known as uplift, or the Conditional Average Treatment Effect (CATE), also known as conditional uplift. These terms represent the increase or decrease of a certain measure of the effect on the outcome variable [7]. Recently, the popular uplift model primarily focuses on predicting the uplift of an outcome if some treatment or exposure changes. It seeks to answer questions such as, "If we implement smoking cessation at 65, how much longer will people live?" However, this model doesn't provide the predicted years of longevity for individuals who cease smoking.

Instead, counterfactual (CF) prediction aims to predict the absolute value of the outcome under the "what if" scenario. Observational data only allows us to observe one of the potential outcomes - the factual outcome if an individual ceased smoking or the factual outcome if they did not. The CF question, "If this person did (or did not) cease smoking, how many years of longevity would they have?" remains unanswered. The absolute CF prediction result provides a framework for patients to choose better treatment options based on the predicted CF clinical endpoints in this era of big data.

From table 1 below, it is a illustration of the what CF prediction is doing. we use question mark to represent the CF results for each patients:

PatID	T	Demographics	Actual Outcome	CF T_A	CF T_B	CF T_C
0	A	X_1	1	1	?	?
1	B	X_2	0	?	0	?
2	C	X_3	1	?	?	1
3	B	X_4	1	?	0	?
4	A	X_5	1	1	?	?
5	C	X_6	0	?	?	0

The question mark stands for the unanswered CF result under "what if" condition. CF prediction model focuses on making predictions for those questions marks based on the observed values in reality.

Some basic methods like K-nearest neighbor (KNN) matching and propensity score matching (PSM) can be useful for predictions and inferences. KNN groups data points with similar characteristics, allowing outcomes from one patient group to inform another. This aligns with case-control studies where all factors, except treatments, are identical, eliminating confounding issues. On the other hand, PSM focuses on the probability of receiving certain treatments, balancing covariates to reduce confounding. Data points with similar propensity scores are grouped, partitioned, and cross-generalized to generate results. Other methods, like PS weighting and doubly robust estimators, also calculate treatment group contrasts. However, in this case, we aim to use causal inference to control confounding and make predictions for Statin patients, rather than focusing on contrasts. This is achieved by combining propensity score weights from each treatment cohort to predict others.

One convenient way to make the CF prediction under multiple treatment setting is using the outcome supervised ML model. We can expand the treatment options as a dummy variable matrix, and concatenate it with all other covariates for each data point, then feed that into regression, forest-based ML, or Neural Network-based models, as Chih-Lin et al. [5] did in his paper. Suppose we have M number of treatments, then we use the trained model to predict the augmented data matrix where each data point is expanded into M entries with only the treatment indicator switched to all the other treatment options. The CF prediction can be directly generated by making the model prediction for this augmented data set where the model is trained using the original data. Here we call this an outcome model that $Y \sim T + X$. This method has some significant flaws: It might be very model based, in which case some predictions are made from extrapolation of data points that are far away from the data cloud. Another problem is that: Regularization of the ML model is always implemented, and it might cause the loss of the Treatment variable if the treatment variable tends to be insignificant for the model. Therefore, the specialty of the treatment variables is also an essential factor to consider because investigators always want to see which treatment is superior comparing to another. The third problem would be the lack of consideration of confounding where outcome model cannot eliminate the confounding issue when making the CF prediction. Therefore, prediction result might be unreliable and biased because the CF prediction for each treatment might be using different data due to the regularization of the ML model.

1.2 Motivations of Counterfactual Prediction under Multiple Treatment Settings

Counterfactual predictions based on Electronic Health Records (EHR) data offer significant advancements in personalized medicine and treatment optimization for statin therapy. By exploring the treatment space for statins, including variations in agents and dosages, counterfactual predictions allow for more precise tailoring of treatments, ensuring maximum efficacy and minimizing adverse effects. This personalized approach enhances patient adherence, optimizes dosage levels, and assesses the potential for combining statins with other lipid-lowering agents.

Counterfactual prediction models enable clinicians to simulate different treatment scenarios, offering valuable insights for decision-making. These models are essential in comparative effectiveness research, providing real-world evidence on the efficacy of statin therapies, and supporting the development of clinical guidelines and policy decisions. By incorporating data-driven insights from these models, healthcare providers can improve patient outcomes and enhance healthcare efficiency.

Counterfactual predictions in EHR data significantly advance personalized medicine. These models forecast how individual patients might respond to different statin therapies, enabling clinicians to create customized treatment plans that maximize efficacy while minimizing side effects. By accounting for genetic, physiological, and lifestyle factors, counterfactual predictions improve the precision of treatment

strategies. Research has demonstrated that counterfactual models can identify the most effective statin regimen, reducing cardiovascular risk and improving patient adherence by mitigating side effects [8]. Furthermore, real-time data integration allows for dynamic treatment adjustments, a cornerstone of personalized medicine [9].

Counterfactual models are also pivotal in optimizing statin therapy. These models simulate unimplemented treatment paths, allowing a comprehensive evaluation of how alternative regimens might impact patient outcomes. Research has shown that counterfactual predictions personalize treatment plans by balancing the benefits of LDL-C reduction with the risks of statin-associated symptoms or therapy discontinuation. This data-driven approach minimizes trial-and-error prescribing, enhances patient satisfaction, and ultimately improves treatment efficacy while reducing healthcare costs [10].

Counterfactual predictions also provide valuable insights into the long-term effects of statin therapy. By examining potential outcomes over extended periods, researchers can assess the impact of statins on cardiovascular risk and mortality. These models also highlight the safety profiles of different statin regimens, shedding light on adverse effects such as muscle symptoms or new-onset diabetes from prolonged use [11]. EHR-based predictions refine treatment strategies and help clinicians make informed decisions about the risks and benefits of long-term statin therapy [12].

Additionally, counterfactual predictions allow healthcare providers and policymakers to evaluate the cost-effectiveness of different statin therapies before implementation. These models can assess both direct medication costs and potential savings from reduced cardiovascular events and hospitalizations, contributing to better resource allocation and sustainable healthcare practices [13].

In summary, counterfactual prediction models play a critical role in personalized medicine, optimizing therapeutic outcomes, understanding long-term effects, and assessing cost-effectiveness. These models significantly improve patient care, refine clinical guidelines, and contribute to the efficient use of healthcare resources.

1.3 Significance of Counterfactual Predictions

Predictive Modeling: Counterfactual predictions derived from EHR data represent significant advancements in predictive modeling for statin therapy. These models allow clinicians to simulate various treatment scenarios, enabling the creation of customized treatment plans based on individual risk profiles and genetic predispositions. By reducing trial-and-error prescribing, counterfactual predictions enhance patient care by providing accurate forecasts of LDL-C reduction and the likelihood of adverse reactions.

Comparative Effectiveness Research: Counterfactual predictions are essential for comparative effectiveness research, offering insights into how different statin therapies perform in real-world settings. By analyzing what might have happened if a patient received an alternative treatment, these models clarify the effectiveness of various statins for different patient demographics. This evidence-based approach

ensures that healthcare providers can prescribe the most appropriate and effective statin therapy for each patient.

Adverse Event Analysis: Counterfactual predictions are invaluable for identifying and managing risks associated with statin therapy. By simulating different treatment scenarios, researchers can identify adverse reactions specific to certain statins, helping clinicians avoid high-risk medications. These proactive analyses improve patient safety and encourage adherence by minimizing side effects.

Improving Clinical Guidelines: Finally, counterfactual predictions support the refinement of clinical guidelines for statin use. These models offer clearer insights into how treatments perform across diverse patient populations, enabling the development of more effective and nuanced guidelines. The insights gained from counterfactual analyses help optimize treatment strategies and ensure that healthcare resources are used efficiently, ultimately improving patient outcomes.

Overall, counterfactual predictions play a critical role in multiple treatment settings, supporting personalized medicine, improving efficacy, and guiding evidence-based clinical decisions.

Chapter 2

Literature Review

There are tons of literatures focusing on causal inference in past decades. From the beginning, Hernan & Robins [14] has been talking about esitmating causal target through standardization, propensity related modeling, and time-varying treatment. Hernan & Robins mainly involves binary treatment setting and the target of estimation is focusing on causal effect (This can be defined as the causal difference or causal quotient). While in reality, there are many scenarios involing multiple treatments instead of dichotomous treatment setting. Fan Li et al. [15] proposed Overlap Weighting mechanism to accomodate situations with more than 2 treatment options, by focusing on the clinical equipoise of the patient cohort. Patients with larger uncertainty of being assigned to certain treatment option will be up-weighted, while patients with larger cerntainty of being assigned to a specific treatment option will be down-weighted. Overlap Weights became an important stake for us to develop the counterfactual prediction for statin treatment cohort.

While casual inference methods are important for causal effect estimation while eliminating confounding issue, Machine Learning became another important topic in past decade. There are tons of application of combining Machine Learning models with Causal inference methods for Causal target estimation. Chernozhukov et al. [16] introduced several Causal Learner techinques for causal estimation such as Double Machine learning. Doubly Machine Learning constructs one model for Treatment/Exposure variables in terms of confounding variables, and constructs the second model for Outcome model using both treatment variables and confounding variables.

Previously Chi et al. [5] proposed an machine learning framework using neural networks for making predictions of three clinical endpoints for Statin cohort. It was named as Personalized Statin Treatment Plan (PSTP) generated using Pareto optimal of the three predicted clinical endpoints, by minimizing Statin Associated Syndromes, Discontinuation of the Statin Treatment, and maximizing the LDL-C reduction. However, there are three limitations, and they can influence PSTP usability: (1) Not taking the counterfactual predictions and confounding bias into account. (2) The balance between multiple drug-prescribing objectives (especially trade-off objectives), such as tradeoff between benefits and risks.

(3) Evaluating PSTP in retrospective data.

2.0.1 Counterfactual Prediction in Medical Research

Counterfactual prediction can be a powerful tool in context of causal inference, allowing researchers to estimate what would have happened to individuals under different treatment scenarios. This method is particularly valuable in medical research, where randomized controlled trials (RCTs) are often impractical or unethical. Instead, observational data can be leveraged to draw causal conclusions using counterfactual frameworks.

The foundational theory of counterfactual prediction is deeply rooted in the Rubin Causal Model (RCM), which formalizes the concept of potential outcomes [17]. This framework posits that each individual has a set of potential outcomes, one for each possible treatment. The observed outcome is determined by the treatment actually received, while the counterfactual outcomes (those corresponding to treatments not received) remain unobserved. This approach has been instrumental in developing methods for causal inference from observational data.

Besides, counterfactual prediction is increasingly recognized as a valuable tool for personalizing medical care by simulating the outcomes of different treatment options for individual patients, allowing for more optimized decision-making in clinical practice. This approach extends beyond the traditional use of statistical models by incorporating multiple potential scenarios, which enhances the ability to navigate uncertainties inherent in patient care. By predicting how various interventions might impact individual outcomes, counterfactual prediction supports more tailored treatment strategies, ultimately aiming to improve patient care and inform healthcare policies. This broader application underscores the importance of adopting counterfactual methods more widely in medical research, highlighting their potential to significantly enhance clinical and policy-related decisions. [18]

Similarly, the work published in BMC Medical Research Methodology [19] highlights the importance of counterfactual thinking in understanding causal relationships, particularly in observational studies where the control of confounding factors is crucial. By estimating the average causal effects under different treatment scenarios, these models provide valuable insights that guide the development of more effective and individualized treatment strategies. The broader adoption of counterfactual approaches in medical research not only improves patient outcomes but also informs healthcare policies by integrating predictions from multiple potential scenarios, addressing uncertainties in clinical practice.

2.0.2 Applications of Counterfactual Prediction in Medicine

In recent years, there has been a surge in the application of counterfactual prediction methods in medical research. These methods help to estimate the effects of medical treatments, inform clinical decision-making, and guide healthcare policy.

One notable application is in the evaluation of drug efficacy and safety. Hernán and Robins [20] demonstrated the use of causal inference methods to emulate target trials using observational data, providing insights into the long-term effects of treatments that would be challenging to capture in RCTs. Similarly, Shadish, Cook, and Campbell [21] highlighted the use of propensity score methods to adjust for confounding variables, thus enabling more accurate estimates of treatment effects.

Machine learning techniques have further enhanced the capability of counterfactual prediction models. For instance, Collier et al. [22] used a neural network-based approach to estimate individualized treatment effects, showing improved accuracy over traditional methods. These advancements are crucial in personalized medicine, where treatments can be tailored to the individual characteristics of patients.

2.0.3 Overlap Weights and Doubly Robust Estimation

Overlap weights are a recent innovation in the realm of causal inference, offering a solution to some of the limitations of traditional propensity score methods. Li, Morgan, and Zaslavsky [23] introduced overlap weights to minimize the influence of extreme propensity scores, thereby enhancing the balance between treated and untreated groups and improving the robustness of causal estimates.

Doubly robust estimation is another significant advancement in this field. This method combines propensity score weighting with outcome regression, ensuring that if either the model for the treatment assignment or the model for the outcome is correctly specified, unbiased treatment effect estimates can still be obtained [24]. This dual robustness makes it a preferred choice in medical studies where model misspecification is a concern.

2.0.4 Two-Stage Models in Treatment Effect Estimation

Two-stage models have been proposed to address the complexity of treatments that vary in both type and intensity. These models decompose the treatment assignment process into two stages: first, determining the treatment agent, and second, determining the treatment intensity. This approach is particularly useful in scenarios with multiple treatment options and varying dosages, such as in Statin therapy for cholesterol management.

Recent studies have applied two-stage models to improve the accuracy of counterfactual predictions. For example, counterfactual prediction for bundle treatment [6] discusses a variational sample re-weighting algorithm to improve counterfactual outcome estimation. It emphasizes the importance of accurately predicting counterfactual by adjusting for treatment variations at different stages, which can be applied to dosage levels after initial treatment agent matching. Such models are instrumental in optimizing treatment strategies and enhancing patient outcomes in clinical practice. But it didn't involve Overlapping Weights into the data and simulation for multiple treatment options.

There are some other articles talking about two-stage model applications, which really motivated my

work on spliting the treatment space into agents and dosages/intensities: Pellegrini et ak, [25] describes a model specifically designed to handle variability in treatment effects among patients. The first stage involves predicting baseline risk, while the second stage uses this risk score as a factor in network meta-regression, which is particularly useful for personalizing treatment decisions based on individual patient characteristics. This approach demonstrates how separating the stages allows for more precise estimations of treatment effects across multiple options.

2.0.5 Challenges and Future Directions

Despite the advancements, several challenges remain in the application of counterfactual prediction in medical research. High-dimensional data, the presence of unmeasured confounders, and the need for large sample sizes are persistent issues. Addressing these challenges requires ongoing methodological innovations and rigorous validation studies.

Future research directions include the integration of counterfactual prediction methods with real-time clinical decision support systems, the development of more robust algorithms to handle high-dimensional and sparse data, and the exploration of causal inference techniques in new medical domains such as genomics and personalized healthcare.

2.0.6 Gaps in Current Literature

Despite significant advancements in counterfactual prediction and causal inference, several gaps remain in the current literature, particularly regarding their application in medical research:

High-Dimensional Data: Many existing methods struggle with high-dimensional datasets, which are common in medical research due to the large number of covariates involved. While machine learning approaches have been developed to handle high-dimensional data, their integration with causal inference methods requires further exploration [26].

Unmeasured Confounders: A persistent challenge in observational studies is the presence of unmeasured confounders, which can bias the estimated treatment effects. While methods such as instrumental variables and sensitivity analysis exist, they are not always applicable or sufficiently robust in complex medical datasets [20].

Scalability: Many counterfactual prediction models face scalability issues when applied to large-scale datasets. Efficient algorithms and computational techniques are needed to ensure that these models can be applied to big data without compromising accuracy or computational feasibility [27].

Validation and Generalizability: While numerous methods have been proposed, rigorous validation and testing on diverse medical datasets are often lacking. Ensuring that these methods generalize well across different populations and settings is crucial for their broader adoption in clinical practice [28].

Complex Treatment Regimes: There is limited research on methods that can handle complex treatment regimens involving multiple agents and varying dosages. Existing models typically focus on single treatment effects, whereas real-world medical treatments often involve combinations of drugs and varying intensities [29].

2.0.7 Need for Two-Stage Models in Complex Treatment Scenarios

The need for two-stage models becomes particularly evident in scenarios with a large number of treatments or complex treatment combinations:

Multiple Treatment Agents and Dosages: In many medical contexts, such as Statin therapy for cholesterol management, patients may receive different types of Statins at varying dosages. Paraskevas et al. [30] discussed the differences between different Statin agents in terms of potency and lipid-lowering strength. This paper aims to find the optimal Statin agent and dosage combination for vascular patients, which enlightened us how personalized statin is important for different clinical outcomes and need. Traditional single-stage models may struggle to accurately estimate treatment effects due to the high dimensionality and complexity of the treatment space [31]. Wang et al. discussed multiple different scenarios for defining different treatment cases under different situations. Especically for multi-valued treatment case or multi-leveled treatment, we can use methods like generalized propensity score. But it also didn't mention anything possible methods that split the dimensions of the treatment space into agents and dosages. While at the same time, the dosages can be continuous in terms of the treatment form. The combination of categorical treatment agents and continuous treatment dosages adds more complexity to treatment space.

Improved Treatment Effect Estimation: Two-stage models can provide more precise estimates by decomposing the treatment process into manageable stages. The first stage can focus on predicting the probability of receiving a particular treatment agent, while the second stage can address the intensity or dosage of the treatment. This approach reduces the complexity of each individual model, potentially leading to more accurate and robust estimates [23].

Mitigating Sample Size Issues: When treatments are broken down into numerous combinations, some treatment cohorts may have very small sample sizes, making it challenging to train reliable models. Two-stage models help mitigate this problem by aggregating information across stages, thereby enhancing the stability and reliability of the estimates. There are also other literatures talking about the sample size/treatment cohort size issue [32]. For small sample sizes, simulations can be used to estimate the power of the study and detecting significant effect. Therefore, adding the simulated data to the existing data could be a solution for small treatment cohort size issue. Taking statin treatments as an example, Atorvastatin is the most prevalent one while some other outdated statin options (i.e., Provastatin) are not popular anymore and it's corresponding cohort size is smaller.

Applicability to Personalized Medicine: The ability to handle complex treatment regimens makes

two-stage models particularly suited for personalized medicine, where treatment plans are tailored to individual patient characteristics. This ”two-stage” framework can also extended to any types of treatment delivery methods or any types of treatment forms mentioned by Wang et al. [31]. The process may be designed as ”three-stage” models, or Weighting combinations from different treatment dimensionality (i.e., treatment agent, treatment dosages, treatment delivery, treatment forms). That gives us the largest possibility of doing personalized medicine in terms of those different treatment characteristics/aspects. This approach can lead to more effective and personalized treatment recommendations, ultimately improving patient outcomes [20].

2.0.8 Future Research Directions

To address these gaps, future research should focus on:

- Developing Robust Methods for High-Dimensional Data: Integrating advanced machine learning techniques with causal inference methods to handle high-dimensional datasets effectively.
- Addressing Unmeasured Confounders: Innovating new methods or improving existing ones to account for unmeasured confounders in observational data.
- Scalability and Efficiency: Designing scalable algorithms that maintain accuracy and computational feasibility for large-scale datasets.
- Rigorous Validation: Conducting extensive validation studies across diverse medical datasets to ensure the generalizability of counterfactual prediction models.
- Handling Complex Treatment Regimens: Expanding research on two-stage models and other approaches to better handle complex treatment regimens involving multiple agents and varying dosages, or treatment forms and treatment delieveries.

By addressing these areas, the application of counterfactual prediction in medical research can be significantly advanced, leading to more accurate treatment effect estimates and improved patient care.

2.0.9 Significance of the Thesis

This thesis makes significant contributions to the field of causal inference and personalized medicine by developing and evaluating a two-stage model for counterfactual prediction in the context of Statin medication. The primary significance of this research includes:

Enhanced Precision in Treatment Effect Estimation: By separating the treatment agent and treatment intensity into two stages, this model provides more accurate estimates of the effects of Statin medications at various dosages. This precision is crucial for developing tailored treatment plans that optimize patient outcomes.

Addressing Complexity in Treatment Regimens: The two-stage model effectively manages the complexity of Statin therapy, where patients may receive various types of Statins at different dosages. This approach ensures that the treatment effects are estimated more accurately, accounting for the nuances of real-world clinical practices.

Improving Model Robustness and Scalability: By breaking down the treatment assignment into two stages, the model mitigates issues related to small sample sizes for specific treatment combinations, enhancing the robustness and reliability of the estimates. This scalability is essential for applying the model to large-scale healthcare databases.

Contributing to Personalized Medicine: The findings from this research can directly inform clinical decision-making, enabling healthcare providers to prescribe Statin medications that are tailored to the individual characteristics of patients. This personalized approach has the potential to improve treatment adherence and effectiveness, ultimately leading to better health outcomes.

Methodological Innovation: This thesis advances the methodological framework for counterfactual prediction by integrating modern machine learning techniques with causal inference methods. This innovation is valuable for the broader application of these methods across various medical treatments beyond Statins.

2.0.10 Challenges and Future Directions

Despite the advancements, several challenges remain in the application of counterfactual prediction in medical research. High-dimensional data, the presence of unmeasured confounders, and the need for large sample sizes are persistent issues. Addressing these challenges requires ongoing methodological innovations and rigorous validation studies.

Future research directions include the integration of counterfactual prediction methods with real-time clinical decision support systems, the development of more robust algorithms to handle high-dimensional and sparse data, and the exploration of causal inference techniques in new medical domains such as genomics and personalized healthcare.

By addressing these areas, the application of counterfactual prediction in medical research can be significantly advanced, leading to more accurate treatment effect estimates and improved patient care.

Chapter 3

Methodology

3.1 Introduction

The counterfactual predictive analysis and proactive strategy for selecting the optimal statin treatment plan (PSTP) is a comprehensive framework designed to personalize statin therapy based on individual patient characteristics. This methodology is structured into several integrated steps: data extraction and preprocessing, model development using counterfactual prediction models with generalized propensity scores and generalized overlap weights, clinical trial simulation with multiple arms, and data simulation for model comparisons. Each component plays a critical role in ensuring that the PSTP accurately identifies optimal treatment recommendations tailored to individual patients, thereby improving clinical outcomes.

The process begins with rigorous data extraction and preprocessing to ensure that high-quality data is used for model development. Relevant clinical variables, patient demographics, and historical treatment data are extracted from electronic health records (EHRs), focusing on key variables such as age, sex, comorbidities, baseline lipid levels, and prior statin use. The data preparation involves cleaning steps, including the removal of duplicates, handling missing values, and normalization of variables to enhance comparability. Categorical variables are encoded, and continuous variables are scaled to optimize model performance. Feature engineering is employed to create interaction terms and composite variables that reflect patient-specific factors influencing treatment outcomes, ensuring data robustness for subsequent modeling.

The core of the methodology lies in modeling using counterfactual prediction with generalized propensity scores (GPS) and generalized overlap weights (GOW). The GPS model balances observed covariates across different treatment groups by estimating the probability of receiving each statin treatment based on patient characteristics. This flexible approach accommodates multiple treatment levels, enabling the estimation of treatment effects across various statin agents and dosages simultaneously. Generalized overlap weights are applied to enhance causal inference by focusing on covariate regions

of the clinical equipoise where treatment groups overlap, minimizing extrapolation and improving precision. The counterfactual model predicts outcomes such as Statin-Associated Symptoms (SAS), discontinuation rates, and LDL-C reduction under different treatment scenarios. By leveraging GPS and overlap weights, the model generates counterfactual prediction of each individual's treatment outcomes, by using both baseline one stage counterfactual prediction model and two stage counterfactual predictions model with first stage of Statin agents and second stage of Statin dosages/intensities, supporting personalized statin therapy decisions tailored to each patient's clinical profile.

To validate and compare the counterfactual predictive model, a clinical trial simulation is conducted with four distinct arms: Random Arm, Clinical Guideline Arm, Practical Arm, and Pareto Optimization Arm. Each arm represents a different decision-making strategy for statin treatment selection. The Random Arm serves as a control by randomly assigning patients to statin treatments, ignoring guidelines and personalized factors. The Clinical Guideline Arm assigns treatments according to established clinical guidelines, reflecting current standard care practices. The Practical Arm utilized real-world clinical data, incorporating physician preferences and patient adherence. The Pareto Optimization Arm employs the counterfactual prediction model and Pareto optimization techniques to identify optimal treatment strategies that balance multiple clinical outcomes, such as minimizing SAS and discontinuation while maximizing LDL-C reduction. The simulation evaluates the effectiveness of each arm in achieving desirable clinical outcomes, providing insights into the comparative performance of personalized strategies versus traditional approaches.

Further evaluation of the predictive modeling approach is conducted through data simulations to compare a baseline one-stage model and a two-stage model. The baseline one-stage model makes counterfactual outcome predictions without distinguishing between statin agents and dosages, potentially oversimplifying the treatment process. In contrast, the two-stage model consists of an initial stage where propensity scores are estimated for the choice among different statin agents, followed by a second stage estimating counterfactual propensity scores for various dosages. This approach allows a detailed assessment of dosage-specific effects. Outcome prediction is then performed using weighted regression models for key clinical outcomes, including SAS, discontinuation, and LDL-C reduction. The data simulations generate synthetic patient cohorts resembling real-world populations, enabling a robust comparison of model performance in predicting clinically relevant outcomes. This ensures that the PSTP framework can refine treatment personalization by recommending not only the best statin agent but also the optimal dosage for each patient.

In summary, this multi-step methodology integrates advanced data extraction techniques, counterfactual predictive modeling, clinical trial simulations, and data-driven model comparisons to develop a proactive strategy for selecting the optimal statin treatment plan. By leveraging generalized propensity scores, overlap weights, and clinical trial simulations, the PSTP framework provides a rigorous, personalized approach to statin therapy, enhancing clinical decision-making and improving patient outcomes

in managing hyperlipidemia and cardiovascular risk.

3.1.1 Data Extraction and Preprocessing

The first step focuses on data extraction and preprocessing, which involves identifying key variables and potential confounders that are relevant to statin adherence and persistence, discontinuation rates, and LDL-C reduction effectiveness. This phase integrates diverse data sources, including demographic data, claims data, laboratory results, and diagnostic data, to create a comprehensive dataset. The preprocessing stage involves cleaning the data, handling missing values, and transforming the variables to ensure consistency and suitability for the modeling process. The goal is to create a reliable dataset that captures the complexity of patient profiles and treatment responses, allowing for an in-depth analysis of how various factors influence the effectiveness of statin therapy. More details are in section 3.2 about the data.

3.1.2 Model Development and Prediction

In the second step, the methodology employs advanced statistical and machine learning techniques to generate counterfactual predictions of treatment outcomes. This phase leverages a both baseline models (”one stage”) and two-stage modeling approach that separately addresses the definition of treatment agents and dosages or intensities, using a Generalized Propensity Score (GPS) model. By applying counterfactual prediction methods, the model simulates the potential outcomes of different treatment strategies, accounting for both observed and unobserved confounders. This approach allows for a nuanced understanding of how variations in treatment regimes impact patient outcomes, such as changes in LDL-C levels and the likelihood of discontinuation or adherence. The model aims to predict individual responses to each statin treatment scenario, providing the foundation for personalized treatment recommendations.

3.1.3 Optimization and Clinical Trial Simulation

The third step involves implementing Multi-Objective Optimization (MOO) on the predicted outcomes to identify the optimal treatment strategy for each patient. MOO techniques balance multiple treatment objectives, such as maximizing LDL-C reduction while minimizing side effects and discontinuation risks. The optimization process ensures that treatment recommendations are not only clinically effective but also aligned with patient preferences and clinical guidelines. Following the optimization, the PSTP is evaluated using a robust comparative analysis within a clinical trial simulation (CTS) framework. The PSTP’s performance is benchmarked against other arms, including a random assignment arm, a clinical practice arm reflecting standard care, and an arm based on established statin clinical guidelines. This

comparative analysis provides critical insights into the efficacy and practical applicability of the PSTP in real-world clinical settings.

3.1.4 Data Simulation for Model Evaluation

The third step involves simulating data to assess and compare the performance of the baseline model and the two-stage model. The simulation process generates synthetic datasets that replicate the complexities observed in real-world clinical data, allowing for robust evaluation under controlled conditions. Different numbers of treatment scenarios (e.g., 4, 9, 12, 16, 20, 21 treatments) are simulated to evaluate how the models handle varying levels of treatment granularity and overlap in treatment effects. These simulations help to identify conditions under which the regular two-stage model, which prioritizes dosage/intensity as the second stage, outperforms the reversed two-stage model (TSRO) that uses statin agents as the second stage. Notably, this step has revealed that the regular two-stage model performs particularly well when the total number of treatments reaches 20 or 21, highlighting the importance of model structure in accurately capturing treatment effects. The insights gained from this simulation step are crucial for refining the modeling approach and understanding the trade-offs between different model configurations.

Overall, the methodology integrates rigorous data analysis, advanced predictive modeling, optimization techniques, and data simulation to provide a comprehensive, evidence-based approach to statin treatment planning. This strategy not only enhances individualized patient care but also addresses the broader challenge of optimizing chronic disease management in clinical practice.

3.2 Data

There are two major data sources used for the Statin patient cohort construction. The OptumLabs® Data Warehouse has over 126.6 million of the claims data and 108 millions Electronic Health Records (EHR) data records. The data extraction contains adults' records who took Statin treatment in their medication table. Other related data tables contain the basic demographic information, patients' insurance claims data, electronic health records including diagnosis and comorbidities information, and laboratory results.

The first statin data cohort utilized in this thesis were obtained from the OptumLabs Data Warehouse, a large de-identified database that integrates claims data with electronic health records (EHR). Our study cohort was initially drawn from a population of 42,835,911 patients who had a statin prescription recorded between October 1, 2015, and April 1, 2022. The inclusion and exclusion criteria were meticulously applied to ensure the selection of a homogenous and clinically relevant sample for analysis.

We first excluded 38,933,796 patients who were prescribed any added or combined drug with statin therapy, leaving us with 3,902,115 patients. From this subset, we included patients who had their first

statin pharmacy fill within a clean period of at least one year and prior to April 1, 2022. Further refining the cohort, we excluded 25,706 patients whose statin prescriptions were primarily covered by Medicare or who lacked 365 days of continuous medical and pharmacy coverage prior to their statin index date. This step resulted in a cohort of 942,376 patients.

Next, we included only those patients who had at least one LDL measurement recorded within 365 days prior to their statin index date, narrowing the cohort down to 377,919 patients. From these, we further restricted our analysis to patients aged 20 to 75 years, resulting in a sample of 332,636 patients.

The cohort was then further refined by including only those patients diagnosed with high LDL, diabetes, and ASCVD, resulting in a final sample size of 179,413. To account for potential confounding, missing values for race, gender, geographical location, low-income status, and rent/housing status were imputed using an iterative imputer with an 'Unknown' indicator for missingness. After the imputation process, 176,233 patients remained.

Lastly, we excluded a small proportion of patients who took Lovastatin at a moderate intensity level (0.28%) or Pitavastatin at a low intensity level (0.1%), along with patients who were missing SAS outcomes. This yielded a final analytic sample of 176,233 patients, ensuring a robust dataset for our study's subsequent analyses. Details are listed in the consort diagram A.2

Another data source for the Statin cohort is the data repository from the University of Minnesota, called the Academic Health Center Information Exchange (AHC-IE). The AHC-IE contains EHRs from the Epic® system for more than 4.5 million patients seen at 8 hospitals and more than 40 Fairview Health Services clinics located mainly in State of Minnesota.

The data focused on patients aged 18 and above who had at least one recorded statin prescription. Initially, 193,396 patients were identified within the Fairview EHR system. The cohort was refined by excluding 45,567 patients whose initial statin prescription occurred either before 2010 or after 2020, narrowing the sample to 147,829 patients whose first statin prescription fell within the 2010 to 2020 period and who had at least one year of clean period of statin medication prior to the index date.

Further exclusion criteria were applied to ensure the availability of critical clinical data during the baseline period. Specifically, we excluded 62,020 patients who lacked any Fairview encounters, had no recorded blood pressure or weight measurements, or were missing lab data during this period. This resulted in 85,809 patients who were regular users of the Fairview system in the baseline period.

Next, we examined patient data in the follow-up period, excluding 68,920 patients who had no Fairview encounters, lacked blood pressure or weight measurements, or had missing lab data during this time, leaving a cohort of 16,889 patients who were regular Fairview system users in both the baseline and follow-up periods.

The final stage of cohort selection involved excluding statin cohorts with a sample size smaller than 100, which accounted for 4,003 patients, and any patients with missing values for BMI, systolic, or diastolic blood pressure. This refinement resulted in a final analytic cohort of 12,886 patients, providing

a robust dataset for evaluating the impact of statin therapy on the targeted clinical outcomes.

3.3 Model

3.3.1 Baseline Counterfactual Model: Counterfactual prediction using Overlap Weights

Overlap Weight for binary treatment is defined as the following as mentioned in several Li's paper: [15] [23] [33]. $w_i = 1 - \hat{e}_i$ for treated units and $w_i = \hat{e}_i$ for control group unit. In this case, the OW up-weights the patients who are equipoise in the study cohort. Those patients have a similar probability of being assigned to the treatment and control groups. On the contrary, OW tends to down-weight the patients' group with a relatively determined probability of being assigned to either group(i.e., patients with a 0.96 probability of being assigned to the treatment group will be down-weighted.)

Overlap Weight for multiple treatments (Generalized Overlap Weights) is defined as the following based on Li et al.'s article. [33]:

Assume that the propensity score is calculated using multinomial logistic regression and the conditional probability for each treatment group is defined as the following:

$$e_t(X) = Pr(T = t|X), t \in \mathbf{T} \tag{3.1}$$

In this case, the sum of the totally probability equals to 1 a.k.a $\sum_{t=1}^{T} e_t(X) = 1$. In multiple treatment settings, we want to maintain the standard assumptions in causal inference and propensity score analysis: 1): consistency, 2): positivity, 3): conditional exchangeability(conditional ignorability), and 4): no interference [34]. Inverse probability weighting is a type of balancing weights [23] and the target population is defined as $g(X) = f(x)h(x)$, where $h(X)$ can be prespecified function of covariates. It is called the tilting function. Then different types of balancing weights and target population for making the pairwise comparisons can be generated by switching different tilting functions, as specified in Table 1 in [33].

The real-world scenario of making the CF prediction for the statin cohort has the problem that it is always the multiple treatment setting. In order to adjust for confounding, weighting method is always a good tool to consider. There are many Statin options for patients to choose with different Statin type and dosage combination. Doing the observational study and its related prediction tasks using generalized OW for multiple treatments helped us aim for doing the treatment selection optimization for patients more robustly by controlling the confounding variables using generalized balancing weights [33]. By making the CF prediction for multiple clinical endpoints and outcomes, we can use the optimization tool to select the optimal Statin medications for patients based on those predicted outcomes.

Instead of using the weights to calculate the weighted average for the outcome to estimate the causal effect(contrast between treated and control), we incorporate the calculated OW into the ML model as sample weights for the outcome prediction. Every single patient (data point) will have their weight. The

application of the weighting procedure is specified as the following similar to the simulation study in Li et al.'s paper [33]:

First of all, work on building up the ML(DL) model(aka, Propensity Net) to predict probability of being assigned to a medication(i.e., Atorvastatin 10 mg, Atorvastatin 20 mg, Simvastatin 20mg, etc.) $:= e_t$ where $t \in 1,2,3,...,T$(There are t number of treatments options) and $\sum_1^T e_t(X) = 1$. Then calculate the OW using the estimated propensity score $\hat{e}_t(X)$:

$$OW := (\sum_{t=1}^{T} 1/\hat{e}_t(X))^{-1}/\hat{e}_t(X), t \in \mathbf{T} \tag{3.2}$$

with the tilting function as:

$$h(X) = (\sum_{t=1}^{T} 1/\hat{e}_t(X))^{-1} \tag{3.3}$$

After the OW is calculated, the way of using the OW for making the CF prediction is to incorporate the weights into the outcome models as the sample weight, for each treatment cohort. So there will be T number of models:

$$f_t(X_t; OW_t) := \mathbb{E}(Y_t^{obs}|X_t; OW_t), t \in \mathbf{T} \tag{3.4}$$

where f_t is trained with the observed Y label for each treatment cohort t. This modeling procedure is also very similar to the domain adaptation learner mentioned in [35]. For example, we can use Statin Atorvastatin 10 mg treatment cohort to build the model $f_{AT_{10mg}}$ where the model is weighted by $OW_{AT_{10mg}}$. Then this model will be used to make prediction for the whole cohort to generate the CF for treatment Atorvastatin 10 mg.

The model f_t here can be any supervised learning model we want to use. For example, it can be simple logistic regression so that it will become so-called weighted logistic regression. Alternatively, f can be a more complex model such as neural networks, in which case the OW will used as sample weights during the gradient descent training procedures for each treatment cohort.

Algorithm 1 CF prediction using Overlap Weights:

Require: $t \in \mathbf{T}$

Train $e_t(X) := Pr(T = t|X), t \in \mathbf{T}$

Get prediction $\hat{e}_t(X)$

Calculate Overlap Weights:

$OW := (\sum_{t=1}^{T} 1/\hat{e}_t(X))^{-1}/\hat{e}_t(X), t \in \mathbf{T}, OW \in \mathbb{R}^{N \times T}$

for $t = 1,2,3,...,T$ **do**

$Y_t \sim X_t, t \in \mathbf{T}$

$f_t := \mathbb{E}(Y_t^{obs}|X_t; OW_t), t \in \mathbf{T}$

CF prediction: $\hat{Y}^{(t)} = \hat{f}_t(X)$

end

Another method in causal inference is called Doubly Robuts estimator. It is considered "doubly robust" because it provides consistent estimates of treatment effects if either the propensity score model or the outcome model (but not necessarily both) is correctly specified. This dual reliance on two models enhances the estimator's resilience to misspecification, making it a widely used approach in observational studies to reduce bias. The Doubly Robust method is often considered under in practice and it is superior to other methods due to multiple aspects. First of all, it utilizes the combination of outcome prediction models and treatment assignment models (propensity score models). It leverages of each models while mitigating the individual weakness [36]. Secondly, it has the robustness against misspecification of either the outcome model or the treatment assignment model. Even if one of the models is not correctly specified, the estimator remains consistent as long as at least one of the models is correctly specified [24]. By combining information from both models, they can achieve lower variance in their estimates, leading to more precise and reliable predictions.

In our counterfactual prediction model using doubly robust mechanism, Overlap weights are calculated and will be included as one of the covariate to the outcome model, to achieve the doubly robustness for the counterfactual prediction procedure.

Algorithm 2 CF prediction using Doubly Robust method with OW:

Require: $t \in \mathbf{T}$

Train $e_t(X) := Pr(T = t|X), t \in \mathbf{T}$

Get prediction $\hat{e} * t(X)$

Calculate Overlap Weights:

$OW := (\sum_{t=1}^{T} 1/\hat{e}_t(X))^{-1}/\hat{e}_t(X), t \in \mathbf{T}, OW \in \mathbb{R}^{N \times T}$

for $t = 1, 2, 3, ..., T$ **do**

$Y_t \sim X_t, t \in \mathbf{T}$

$f_t := \mathbb{E}(Y_t^{obs}|[X_t, e_t(X_t)]; OW_t), t \in \mathbf{T}$

CF prediction: $\hat{Y}^{(t)} = \hat{f}_t(X)$

end

3.3.2 Two-stage Counterfactual Model: Propensity for Statin agent and counterfactual propensity for Statin intensity level

The implementation of the two stage model is specified as the following:

The first stage(same as the procedure in Aim1): build up the ML(DL) model to predict probability of being assigned to a medication type (i.e., Atorvastatin, Simvastatin, etc.) $= e_t$ where $t \in 1, 2, 3, 4, 5, 6, 7$ (There are 7 major type of Statin medication) and $\sum_1^7 e_t(X) = 1$. Then calculate the OW using the above probabilities and incorporate OW into the second stage PS model for making the intensity level CF probability(propensity) prediction again. The second PS model is also regarded as the CF outcome prediction model using the first stage propensity score model. Then the CF prediction is proceeded again

to make the prediction for the outcome target(i.e., SAS, Discontinuation, and LDL reduction) based on the second stage model predicted propensity score for dosage and its calculated OW. Details are as the following [4]:

Algorithm 3 Two Stage CF prediction using Overlap Weights:

Require: $t \in \mathbf{T}$ $d \in \mathbf{D}$

Stage 1 PS model:

$e_t(X) = \mathbb{E}(\mathbb{1}\{T = t\}|X) = \Pr(T = t|X), t \in \mathbf{T}$

$OW^{[1]} := (\sum_{t=1}^{T} 1/\hat{e}_t(X))^{-1}/\hat{e}_t(X), t \in \mathbf{T}, OW^{[1]} \in \mathbb{R}^{N\times T}$

for $t = 1, 2, 3, ..., T$ **do**

Stage 2 PS model:

$e_d(X_t; OW_t^{[1]})^{[t]} = \mathbb{E}(\mathbb{1}\{D = d\}|X_t; OW_t^{[1]}) = \Pr(D = d|X_t; OW_t^{[1]}), d \in \mathbf{D}, t \in \mathbf{T}$

$OW^{[2][t]} := (\sum_{t=1}^{T} 1/\hat{e}_d^{[t]})^{-1}/\hat{e}_d^{[t]}, d \in \mathbf{D}, OW^{[2][t]} \in \mathbb{R}^{N\times D}$

for $d = 1, 2, ..., D$ **do**

$f_{d,t}(X_d; OW_d^{[2][t]}) = \mathbb{E}(Y_d^{obs}|X_d; OW_d^{[2][t]}), d \in \mathbf{D}, t \in \mathbf{T}$

$\hat{Y}^{(d,t)} = \hat{f}_{d,t}(X), d \in \mathbf{D}, t \in \mathbf{T}$

end

end

3.3.3 Reversed Order Two-stage (TSRO) counterfactual model: Reverse the order of Statin agents and Statin dosage/intensity levels

The reversed order of Two stage model is implemented in the simulated data (Chapter 4) for the comparisons between the number of Statin agents and number of Statin dosage/intensity levels. We suspect that the smaller number of the second stage levels (i.e., the number of Statin agents or the number of intensity levels)

Algorithm 4 Two Stage Reversed Order (TSRO) CF prediction using Overlap Weights:

Require: $t \in \mathbf{T}\ d \in \mathbf{D}$

Stage 1 PS model:

$e_d(X) = \mathbb{E}(\mathbb{1}\{D = d\}|X) = \Pr(D = d|X), d \in \mathbf{D}$

$OW^{[1]} := (\sum_{d=1}^{T} 1/\hat{e}_d(X))^{-1}/\hat{e}_d(X), d \in \mathbf{D}, OW^{[1]} \in \mathbb{R}^{N\times D}$

for $d = 1, 2, 3, ..., D$ **do**

Stage 2 PS model:

$e_t(X_d; OW_d^{[1]})^{[d]} = \mathbb{E}(\mathbb{1}\{T = t\}|X_d; OW_d^{[1]}) = \Pr(T = t|X_d; OW_d^{[1]}), t \in \mathbf{T}, d \in \mathbf{D}$

$OW^{[2][d]} := (\sum_{d=1}^{D} 1/\hat{e}_t^{[d]})^{-1}/\hat{e}_t^{[d]}, t \in \mathbf{T}, OW^{[2][d]} \in \mathbb{R}^{N\times T}$

for $t = 1, 2, ..., T$ **do**

$f_{t,d}(X_t; OW_t^{[2][d]}) = \mathbb{E}(Y_t^{obs}|X_t; OW_t^{[2][d]}), t \in \mathbf{T}, d \in \mathbf{D}$

$\hat{Y}^{(t,d)} = \hat{f}_{t,d}(X), t \in \mathbf{T}, d \in \mathbf{D}$

end

end

3.4 Clinical Trial Simulation

In this section, we simulate the Statin prescription arms for each patient and use the counterfactual (CF) prediction results to compare the simulated arms with each other [37]. This approach provides a framework for evaluating CF predictions using Clinical Trial Simulation (CTS), which consists of four arms: Random Arm, Actual Arm, Optimization Arm, and Clinical Guidelines Arm.

The CTS comprises four arms simulating different Statin treatment strategies. Arm 1, the Random Arm, assigns each patient a randomly selected treatment, serving as the control arm for comparison purposes. The Clinical Guidelines Arm (Arm 2) follows the 2018 updated guidelines for primary prevention of cardiovascular disease [38]. This arm selects Statin types and dosages based on specific guideline criteria; for example, if the patient is over 50 years old, a lower dosage is selected according to the guidelines.

The Practice Arm (Arm 3) represents the treatments patients received in real-world practice, based on our data. Finally, Arm 4, the Optimization (a.k.a, Personalized Statin Treatment Plan (PSTP) Arm) Arm, simulates Statin treatments optimized using the CF prediction model. The CF model incorporates target variables such as Statin-Associated Symptoms (SAS) risk (adverse event risk), discontinuation risk, and LDL-C reduction. Based on the predicted probabilities of these outcomes, the optimization method selects the optimal Statin treatment for each patient. For instance, the system aims to choose the Statin medication with the lowest SAS risk, the lowest discontinuation probability, and the highest LDL-C reduction, making this treatment the optimal choice for Arm 4.

Comparing these four arms provides insights into the effectiveness of CF predictions. The focus is particularly on comparing the Optimization (PSTP) Arm (Arm 4) with other arms, such as the Clinical

Guidelines Arm and the Actual Arm. The comparison between the Actual Arm (Arm 3) and the Optimization Arm (Arm 4) reveals whether the Statin treatments patients received in reality were the most beneficial based on the CF predictions.

Additionally, comparisons between CF predictions and factual outcomes in the Actual Arm (Arm 3) assess the accuracy of the CF model in reflecting real-world results, helping to determine whether the predictions align closely with actual patient outcomes.

3.5 Simulation

Using causal data simulation to evaluate the performance of counterfactual prediction models provides a robust framework for understanding how different models behave under various treatment regimens. This approach involves simulating datasets with known causal structures and varying the number of treatments to observe how well different models, such as the two-stage models, predict outcomes in controlled scenarios. As these models are exposed to increasing amounts of training data, particularly in settings where multiple treatments are simulated, they are hypothesized to perform more effectively. This improvement is anticipated because two-stage model gives each treatment additional data for training process and this provides a richer basis for the models to learn complex patterns and treatment interactions, enhancing their ability to predict counterfactual outcomes accurately. By systematically comparing these models across varied simulated treatment scenarios, researchers can gain deeper insights into model capabilities and limitations, guiding the optimization of model design for real-world applications in healthcare and policy-making.

For counterfactual prediction, we never observe the ground truth for each of the treatment options for patients. Only one factual outcome is observed in observational data. Therefore, data simulation based on our domain knowledge and causal graph would generate the ground truth for each of the treatment options, by defining the causal effect among all treatment options beforehand. With the simulated data with ground truth under each treatment options, we can apply our models and get CF predictions that can be validated using Machine Learning metrics, such as Area Under Receiver Operating Characteristic Curve (AUROC), Mean Squared Error(MSE), etc.

More details will be on Simulation chapter.

Chapter 4

Simulation

4.1 Introduction

In this chapter, we delve into the simulation framework designed to evaluate counterfactual predictions for Statin medication outcomes. The primary objective of this simulation is to rigorously compare the performance of a baseline counterfactual model against a more sophisticated two-stage model. These models are instrumental in estimating the effects of different Statin treatments and their dosages on patient outcomes, specifically focusing on outcomes such as Statin-Associated Symptoms (SAS), discontinuation rates, and LDL-C reduction. We know that in real-world data, we don't observe all potential outcomes under each treatment strategies. This is why data simulation can mend this critical flaw where we can use simulation to generate synthetic potential outcomes for each treatment strategies. This simulation also is also comprised as one of the evaluation method for our counterfactual prediction models.

As mentioned in the Chapter 3, the baseline model, also known as the one-stage model, directly estimates the treatment effects on outcomes without differentiating between the type of Statin agent and its dosage. In this simpler approach, a single generalized propensity score model and generalized overlap weights are used to predict treatment assignment, followed by an outcome model that uses these weights to adjust for confounding. While this model is straightforward and computationally efficient, it may overlook the complexities involved in Statin treatment decisions, where the type of medication and dosage can have distinct impacts on patient outcomes. This limitation may result in suboptimal performance and less precise estimates, especially in scenarios where different Statin agents and dosages exhibit varied effectiveness and safety profiles.

In contrast, the two-stage model introduces a more nuanced approach by breaking down the treatment assignment process into two distinct stages. The first stage involves estimating generalized propensity scores for the selection of the Statin agent, accounting for patient characteristics that influence the choice among different Statin types (e.g., atorvastatin, simvastatin). The second stage involves estimating counterfactual propensity scores for the selection of specific dosages within each Statin type. This

two-stage approach allows for a more granular assessment of dosage-specific effects, capturing the layered decision-making process clinicians face when prescribing Statins. By incorporating generalized propensity scores and generalized overlap weights, the two-stage model enhances the ability to balance confounders across multiple treatment levels, ensuring that the comparisons between different Statin agents and dosages are more accurate and reliable.

The simulation framework involves generating synthetic patient data using normal distribution along with confounders, ensuring that the simulated scenarios closely reflect the complexities of clinical practice. Data simulation is critical as it allows for the controlled evaluation of model performance under various conditions, such as varying patient characteristics, treatment patterns, and outcome distributions. Through these simulations, we assess how well each model can recover the true treatment potential outcomes, quantify biases, and evaluate precision in outcome prediction.

The comparison between the baseline and two-stage models is centered on key performance metrics, including bias reduction, estimation accuracy, and predictive validity. By comparing these models, the simulation aims to demonstrate the potential advantages of the two-stage approach in capturing the true counterfactual outcomes of Statin treatments more effectively than the one-stage model, especially when the number of treatment options increases expansively. Additionally, the use of generalized overlap weights in the two-stage model improves the robustness of the estimates by focusing the analysis on patients where treatment groups overlap, and treatment number is larger than 2, reducing extrapolation and enhancing the validity of the counterfactual predictions.

Overall, this chapter provides a detailed examination of how different modeling strategies influence the accuracy of counterfactual predictions in Statin treatment scenarios. The findings from these simulations will inform the broader application of these models in personalized medicine, where precise estimation of treatment effects is crucial for optimizing patient outcomes. By systematically evaluating the strengths and limitations of each modeling approach, this chapter aims to highlight the critical importance of the two-stage counterfactual prediction model with GPS and generalized overlap weights, in improving clinical decision-making for Statin therapy.

4.1.1 Baseline Model Overview

The baseline model leverages neural networks to estimate multi-class propensity scores for various Statin treatment options. These propensity scores are pivotal in balancing the covariates among different treatment groups, thus enabling a robust counterfactual analysis. By employing Overlap Weights, which are derived from the predicted propensity scores, the model adjusts for potential confounders and facilitates the prediction of counterfactual outcomes. This approach ensures that the counterfactual predictions are unbiased and reflective of the true causal relationships.

4.1.2 Two-Stage Model Overview

The two-stage model introduces a more nuanced methodology by decoupling the treatment agent and treatment dosage into separate dimensions. In the first stage, the model estimates the propensity score for the Statin treatment agent using neural networks. This initial propensity score serves as a foundational input for Overlap Weights for the second stage, where the intensity or dosage of the treatment is considered. The second stage utilizes the first-stage propensity scores to predict counterfactual propensity scores for different treatment intensities. These refined propensity scores are then used to compute second stage Overlap Weights, which facilitate the final counterfactual outcome predictions.

Besides the regular Two-stage model with first stage modeling treatment agents and second stage modeling treatment dosages, the Two-stage model with Reversed order (TSRO) (first stage modeling treatment dosages, and second-stage modeling treatment agents) is also performed to check the performance. Since the more number of treatment options(i.e., Treatment space cardinality), the worse performance of the model because each treatment cohort size is small for training the model that can perform very well. The goal of the two-stage model is to mitigate this issue, and TSRO can be another options for us to consider. If we put the dosage/intensity level as the second stage because its cardinality is small, we should expect to have better performance of the TSRO model.

4.1.3 Simulation Goals and Evaluation

The simulation aims to benchmark the baseline model against the two-stage model, assessing their respective abilities to predict counterfactual outcomes accurately. To achieve this, we simulate patient outcomes under various treatment scenarios, utilizing ground truth outcomes available for each treatment option. The performance of the models is evaluated using machine learning metrics such as Brier Score and Area Under the Curve (AUC). These metrics provide a quantitative measure of the models' predictive accuracy and their ability to generalize beyond the observed data.

By conducting this simulation, we seek to highlight the strengths and limitations of each modeling approach, ultimately contributing to the development of more effective methodologies for counterfactual prediction in medical research. The insights gained from this comparative analysis will inform future applications and enhancements in personalized medicine, particularly in optimizing Statin therapy for improved patient outcomes.

4.2 Methods

4.2.1 Simulation Design

The simulation study is designed to closely mimic statin medication data. The key steps and parameters of the simulation are as follows:

- **Population Size**: We simulate a cohort of 60,000 patients, reflecting the size of the real dataset.
- **Ground Truth Treatment Effects**: The true treatment effects are set as differences in SAS (Survival After Statin) probabilities, with a mean difference of approximately 0.015. This value is chosen to represent the typical differences observed between treatment options in the real data.
- **Covariates and Confounders**: A total of 186 covariates are simulated, with 100 identified as confounders that influence treatment selection. These 100 confounders are used to generate treatment assignment probabilities and are also included in the outcome simulation to reflect their impact on treatment outcomes.
- **Overlapped and Additive Treatment effects**: This is one of the important component in data simulation where we want to ensure that there are two types of scenarios for simulating the outcomes/treatment effects for different treatment option. So cases are splited into two: one scenario is that there are overlapped treatment effects, which means that some of the treatments options have the similar treatment outcomes. Second scenario is that: there is no overlapped treatment effects. Treatment effect contrasts are generated by additive treatment effects with a fixed, predefined treatment effect.
- **Non-linear components**: Cases are also split for Non-linear component into two: Scenario 1 is with Non-linear component added including polynomial terms and interaction terms for independent variables. Scenario 2 is that there are no non-linear terms for the simulated data.
- **Treatment numbers** In order to see the trend of performance change, simulations are done for different treatment numbers for Statin agent/dosage combinations. The baseline case is that we have 4 treatment options: 2 "Statin" agents (i.e., A vs. B), and 2 dosages (i.e., High vs. Low), in which case it gives us total of 4 combinations.

4.2.2 Simulation steps

1. Set fixed parameters for the simulation:
2.
   - Set the number of patients (data points) $N_p = 60000$
   - Set the number of features $N_f = 186$
   - Set the number of treatments (cardinality) $N_t = [4, 9, 12, 16, 20, 21]$
   - Set the number of confounders $N_c = 100$ to be concatenated to the matrix of patient features.
3. Simulate patient features using standardized normal: $X_c \sim \mathcal{N}(0,1) \in \mathbb{R}^{N_p \times N_f}$
4. Brainstorm some random numbers as the ground truth vector β for confounder matrix: $\beta_c = (\beta_0, \beta_1, ..., \beta_{n_c}) \in \mathbb{R}^{N_C \times |T|}$

5. Generate logits for treatment selection where treatment variable is categorical (i.e., a_low, a_mod, etc.): $logits = X_c\beta_c \in \mathbb{R}^{N_p \times |T|}$, and then use `softmax` function to generate probability distribution for each treatment options: The softmax function is defined as follows [39]:

$$\text{softmax}(\mathbf{z})_i = \frac{\exp(z_i)}{\sum_{j=1}^{n} \exp(z_j)} \quad \text{for } i = 1, \ldots, n$$

where:

- $\mathbf{z} = (z_1, z_2, \ldots, z_n)$ is the input vector of scores or logits.
- $\exp(z_i)$ represents the exponential function applied to the i-th element of the vector.
- $\sum_{j=1}^{n} \exp(z_j)$ is the sum of the exponentials of all elements in the input vector.

The softmax function converts a vector of real-valued scores into a probability distribution over n possible outcomes. $\mathbf{P} = softmax(logits), \mathbf{P} \in \mathbb{R}^{N_p \times |T|}$, where sum of each row of $\mathbf{P} = 1$

6. Simulate (sample) the treatment labels for each patient i, using the multinomial probability distribution generated in last step:

$$T_i \sim \text{Categorical}(\mathbf{P}_i), \quad \text{for } i = 1, 2, \ldots, n_p$$

7. Set treatment effect to be fixed number: $\delta = 0.015$ between each consecutive pairs of those treatment options. For example, if we define the the mean of outcome of the treatment a_low (i.e., SAS prevalence of treatment a_low) as $\bar{Y}^{a\_low}$, then $\bar{Y}^{a\_mod} - \bar{Y}^{a\_low} = \delta = 0.015$, and $\bar{Y}^{a\_high} - \bar{Y}^{a\_mod} = \delta = 0.015$, and etc. The array of the treatment contrasts (i.e., treatment effects = [0.015, 0.03, 0.045, ...] where each number increments by 0.015 for each treatment option) is constructed for later usage in the simulation process Incorporated into the probability threshold for the outcome simulation to ensure the prevalence of the outcomes are following the treatment effects we assumed.

8. For each different number of treatment options, we use alphabetical letter to represent treatment agents (i.e., statin medication type) and categorization of the dosage levels to represent the dosages: High dosage/intensity is named as “high”, moderate dosage/intensity is named as “mod”, and low/intensity dosage is named as “low”. For example, for 4 treatment options, there would be these treatment options named as: “a_high”, “a_low”, “b_high”, “b_low”. For 9 treatment options, it will be all combinations of the treatment type “a”, “b”, “c” and different dosage/intensity level: ”high”, ”mod”, and ”low”. For 12, 16, 20 treatment options, we add one more dosage/intensity level to make the dosage levels more diversified. 21 treatment options goes back to 3 dosage levels to be a comparison for 20 number of treatment options. Following is the python dictionary used for treatment names and labels for the simulation and modeling:

```
treatment_labels_dic = {
    4: {"a_low": 0, "a_high": 1,
        "b_low": 2, "b_high": 3},
    9: {"a_low": 0, "a_mod": 1, "a_high": 2,
        "b_low": 3, "b_mod": 4, "b_high": 5,
        "c_low": 6, "c_mod": 7, "c_high": 8},
    12: {"a_low": 0, "a_mod": 1, "a_high": 2, "a_high2": 3,
         "b_low": 4, "b_mod": 5, "b_high": 6, "a_high2": 7,
         "c_low": 8, "c_mod": 9, "c_high": 10, "c_high2": 11},
    16: {"a_low": 0, "a_mod": 1, "a_high": 2, "a_high2": 3,
         "b_low": 4, "b_mod": 5, "b_high": 6, "a_high2": 7,
         "c_low": 8, "c_mod": 9, "c_high": 10, "c_high2": 11,
         "d_low": 12, "d_mod": 13, "d_high": 14, "d_high2": 15},
    20: {"a_low": 0, "a_mod": 1, "a_high": 2, "a_high2": 3,
         "b_low": 4, "b_mod": 5, "b_high": 6, "a_high2": 7,
         "c_low": 8, "c_mod": 9, "c_high": 10, "c_high2": 11,
         "d_low": 12, "d_mod": 13, "d_high": 14, "d_high2": 15,
         "e_low": 16, "e_mod": 17, "e_high": 18, "e_high2": 19, },
    21: {"a_low": 0, "a_mod": 1, "a_high": 2,
         "b_low": 3, "b_mod": 4, "b_high": 5,
         "c_low": 6, "c_mod": 7, "c_high": 8,
         "d_low": 9, "d_mod": 10, "d_high": 11,
         "e_low": 12, "e_mod": 13, "e_high": 14,
         "f_low": 15, "f_mod": 16, "f_high": 17,
         "g_low": 18, "g_mod": 19, "g_high": 20}
}
```

9. Brainstorm another $\beta_{outcome} \in \mathbb{R}^{(N_f+N_c+1)\times 1}$ vector for simulating the outcome variable. The length of this vector is the sum of number of feature variables, the number of confounders, and 1, where the "1" is the variable of treatment label. So we horizontally stack 2 matrices and 1 vector ($X_f \in \mathbb{R}^{N_p \times N_f}$, $X_c \in \mathbb{R}^{N_p \times N_c}$, $T \in \mathbb{R}^{N_p \times 1}$) to form the final matrix $X_{outcome} \in \mathbb{R}^{N_p \times (N_f+N_c+1)}$ for outcome simulation

10. Generate another logits for binary outcome (i.e., SAS) simulation $baseline\_logits = X_{outcome}\beta_{outcome}$

11. Calculate the first potential outcome probability for treatment "a_low", because we use treatment "a_low" as the baseline treatment option:$p_{\text{a\_low}} = \text{sigmoid}(\text{baseline\_logits} + \varepsilon), \varepsilon \sim \mathcal{N}(0,1)$, where ε is just some random noise.

12. Scale the $p_{\text{a\_low}} < 0.7$ to be within the range of $(0, 0.3)$. Probabilities larger than 0.7 remains the same, while probabilities less than 0.7 are re-scaled to range of $(0, 0.3)$. This ensures the probabilities less than 0.7 will have an even smaller probability, in which case the chance of having outcome "1" is smaller, while those probabilities greater than 0.7 maintains the larger chance of getting outcome "1" The processing of re-scaling is to force the simulated outcome to have mean value 0.3, and it can help us to involving the treatment effects in terms of the outcome prevalence difference between different treatment options.

13. Generate the binary outcomes using the Bernoulli simulation based on the probability vector we generated $p_{\text{a\_low}} \in \mathbb{R}^{N_p \times 1}$

```
from scipy.stats import bernoulli
outcome_a_low = bernoulli.rvs(p_a_low)
```

    where *bernolli.rvs* method is similar to *rbinom* function in R. It generates actually samples based using the given probability distribution.

14. Loop over each treatment options, and generate the scaled probabilities using the following code:

```
        np.where(outcome_probabilities > 0.7 - treatment_effects[i],
↪ outcome_probabilities, scale_probabilities(outcome_probabilities,
↪ 0.000001, 1 - (0.7 - treatment_effects[i])))
```

    where the threshold changes every time for different treatment options because we assumed that the treatment effects are linearly incremental between the current treatment options and baseline treatment option (i.e., a_low). The probability threshold for baseline treatment (i.e., a_low) is 0.7, while for each treatment options later on, the threshold of the probability will be 0.7 - treatment effects to make sure that we are following the treatment contrasts constructed earlier.

15. Simulate the binary outcome (i.e., SAS) based on the new generated probability from last step, for each treatment outcomes. These outcomes are the counterfactual outcomes for each treatment options, that remains unseen during the modeling. The treatment effects shows up here to tell each treatment differences in terms of probabilities that would be used to simulate the outcome.

```
    for i, trt in enumerate([a_mod, a_high, b_low, b_mod, ..., ]):
        trt_treatment_outcome = f"outcome_{trt}"
        trt_outcome_probabilities = sigmoid(baseline_logits + np.random.
↪ randn(num_patients))
        trt_outcome_probabilities_bernoulli_simulation = np.where(
↪ outcome_probabilities > 0.7 - treatment_effects[i],
↪ trt_outcome_probabilities,scale_probabilities(
↪ trt_outcome_probabilities,0.000001, 1 - (0.7 - treatment_effects[i
↪ ])))
        data[trt_treatment_outcome] = bernoulli.rvs(
↪ trt_outcome_probabilities_bernoulli_simulation)
```

16. The last step is to generate the observed outcome using consistency assumption under causal inference context. We select the corresponding counterfactual outcome of the treament label T_i for each patient i.

```
Y_treatments = []
Y_treatments.append(np.array(data['outcome_a_low']))
for trt in treatments:
    column_name = f"outcome_{trt}"
    Y_treatments.append(np.array(data[column_name]))
SAS_observed = np.array([Y_treatments[z][i] for i, z in enumerate(
↪ labels)])  # CONSISTENCY: IMPORTANT!
data['SAS_observed'] = SAS_observed
```

Following is the pseudo code for the data simulation process.

Algorithm 5 Data Simulation with Additive Treatment Effects, with polynomial and interaction terms

Input: Number of patients N, Number of features p, Number of treatments T, Number of confounders C

Output: Simulated dataset with additive treatment effects

Set random seed: seed $\leftarrow 2$

Initialize simulation parameters:

$p \leftarrow 186$ ▷ Number of features for patient data

$p_{poly} \leftarrow 50$ ▷ Number of features for patient data $C \leftarrow 100$ ▷ Number of features for confounders

$C_{poly} \leftarrow 30$ ▷ Number of polynomial features for confounders

Step 1: Generate Patient Features

Generate patient feature matrix:

$$\mathbf{X} \sim \mathcal{N}(0,1) \in \mathbb{R}^{N \times p}$$

Select random indices for polynomial features:

$$I_{poly} \leftarrow \text{Randomly select } p_{poly} \text{ indices from } \{1,\ldots,p\}$$

Extract subset of features:

$$\mathbf{X}_{poly} \leftarrow \mathbf{X}_{[:,I_{poly}]}$$

Generate polynomial features:

$$\mathbf{X}_{poly} \leftarrow \text{PolynomialFeatures}(\mathbf{X}_{poly}, \text{degree} = 2)$$

Combine original and polynomial features:

$$\mathbf{X}_{final} \leftarrow \left[\mathbf{X}_{[:,\text{remaining indices}]}, \mathbf{X}_{poly}\right]$$

Step 2: Generate Confounders and Polynomial Features

Generate confounders:

$$\mathbf{Z} \sim \mathcal{N}(0,1) \in \mathbb{R}^{N \times C}$$

Select random indices:

$$I_{poly}^{conf} \leftarrow \text{Randomly select } C_{poly} \text{ indices from } \{1,\ldots,C\}$$

Extract subset:

$$\mathbf{Z}_{poly} \leftarrow \mathbf{Z}_{[:,I_{poly}^{conf}]}$$

Generate polynomial features:

$$\mathbf{Z}_{poly} \leftarrow \text{PolynomialFeatures}(\mathbf{Z}_{poly}, \text{degree} = 2)$$

Combine confounder features:

$$\mathbf{Z}_{final} \leftarrow \left[\mathbf{Z}_{[:,\text{remaining indices}]}, \mathbf{Z}_{poly}\right] \in \mathbb{R}^{N \times C_{final}}$$

Step 3: Assign Treatment Labels

Brainstorm some coefficients:

$$\boldsymbol{\beta} \in \mathbb{R}^{C_{final} \times T}$$

Calculate logits:

$$\text{logits} \leftarrow \mathbf{Z}_{final} \cdot \boldsymbol{\beta}$$

Calculate probabilities using the softmax function:

$$\mathbf{P} \leftarrow \frac{\exp(\text{logits})}{\sum \exp(\text{logits})}$$

Assign treatments based on probabilities:

$$\text{Labels} \leftarrow \text{Sample based on } \mathbf{P}$$

Step 4: Define Treatment Effects

Set base treatment effect:

$$\tau \leftarrow 0.015$$

Define treatment effects:

$$\boldsymbol{\tau} \leftarrow \tau \times (1,2,\ldots,T)$$

Step 5: Calculate Baseline Outcomes

Generate outcome coefficients:

$$\boldsymbol{\gamma} \in \mathbb{R}^{p_{final}+C_{final}+1}$$

concatenate features and labels:

$$\mathbf{XZL} \leftarrow \left[\mathbf{X}_{final}, \mathbf{Z}_{final}, \text{Labels}\right]$$

Calculate baseline logits:

$$\text{Baseline\_logits} \leftarrow \mathbf{XZL} \cdot \boldsymbol{\gamma}$$

Step 6: Simulate Potential Outcomes

for *each treatment t* **do**

Calculate outcome probabilities:

$$p^{t}_{outcome} \leftarrow \sigma\left(\text{Baseline\_logits} + \mathcal{N}(0, 0.1)\right)$$

Scale the probabilities:

$$p^{t'}_{outcome} \leftarrow \text{ifelse}(p_{outcome} > 0.7 - \boldsymbol{\tau}[t], p_{outcome}, \text{scale\_proba}(p_{outcome}, \min = 0.00001, \max = 1 - (0.7 - \boldsymbol{\tau}[t])))$$

Sample the outcome variable $Y^{[t]}$

$$Y^{[t]} \leftarrow \text{bernoulli.rvs}(p^{t'}_{outcome})$$

Step 7: Compute Observed Outcomes

Map observed outcomes to treatments assigned to ensure **consistency**

Step 8: Split Data into Training and Testing Sets

Split data into training and testing: 80% training, 20% testing

The other three simulation algorithms for additive treatment effects with NO interaction terms, overlapped treatment effects with interactions terms, and overlapped treatment effects with NO interaction terms are presented in appendix 6 8 7

4.2.3 Simulation Scenarios

Multiple scenarios are set up to evaluate model performance and sensitivity:

- **Additive Treatment effects, No interactions of independent variables (patient features)**: This is the baseline scenario of simulation decribed in above section.
- **Addtive Treatment effects, with interactions of independent variables (patient features), and confounders** Among all of the patient features simulated from above, randomly select 50 number of features and 30 numbers of confounders to add polynomial features (i.e., $X_i^2, i \in 1, 2, ..., 50$) and interaction features (i.e., $X_j * X_i, i, j \in 1, 2, ..., 50$) for treatment label T simulation and outcome Y simulation
- **Overlapped Treatment effects, No interactions of the independent variables (patient features)** For all treatment options in the earlier defined dictionary "treatment_labels_dic", randomly select half of the treatment options among all choices and form pairs. For each pair, we force them to have equal treatment effect.
- **Overlapped Treatment effects, with interactions of independent variables (patient features) and confounders** The last scenario combines the overlapped treatment effects and interaction terms together.

4.2.4 Goals and Evaluation Metrics

The primary goals of the simulation are:

1. **Model Performance and Comparison**: To compare the baseline counterfactual prediction model, and the two-stage model in terms of predictive accuracy, Metrics such as Mean Squared Error (MSE) and Area Under the Curve (AUC) are used to evaluate model performance.

2. **Sensitivity Analysis**: To assess the two-stage model's effectiveness in handling numerous treatment options and mitigating the small sample size problem for each treatment cohort.

By simulating these scenarios, we aim to identify the most effective and suitable model for counterfactual prediction in the context of Statin medication, thereby improving our understanding of treatment effects and informing clinical decision-making.

In order to check on the performance changes over different numbers of treatment options, the data is simulated under the following treatment space cardinality: 4, 9, 12, 16, 20, 21. When treatment cardinality $|\mathscr{T}| = 4$, it stands for the most basic scenario where we have two treatment agents and two treatment dosages.

Results are presented in Results chapter and Discussions are in Discussion chapter

Chapter 5

Results

5.1 Result from Optum Labs data

5.1.1 Data

Table A.1 in appendix presents partial list of the variables used for model. It contains patient baseline characteristics across different statin intensity levels and agents. The columns stands for the used statin agent name across our database and rows are the related clinical variables used for modeling. In Table A.1, moderate intensity accounted for the larger proportion of the total statin users. The pre-index LDL-C level showed a higher value for High (i.e., Atorvastatin and Rosuvastatin) and Moderate cohorts, while the low-intensity (i.e., Lovastatin and Pravastatin) cohort's pre-index LDL-C level was relatively lower than the other two cohorts. This is intuitive that doctors might make the prescription based on patients' pre-LDL-C lab value. Table A.1 also shows Atorvastatin, Pravastatin, Rosuvastatin, and Simvastatin are more frequently prescribed comparing to the other two types of statins. Several covariates showed differences across different statin groups. Interestingly, Pitavastatin group tends to have a larger proportion of the patients who had or were having lipid regulating types of the prescriptions. The data for the modeling also included some other demographics variables like Race, Income status. We also included bunch of comorbidity variables into the counterfactual modeling process.

5.1.2 Consort diagram

Figure A.2 shows the consort diagram of the data preprocessing with inclusion and exclusion criteria.

5.1.3 Balancing check

Figure A.4(a) and Figure A.4(b) showed the PSD and ASD calculation for each covariate in our data. PSD calculates the standardized differences between each treatment cohort covariate mean and the population means across all treatment arms. The red lines show the weighted result and the blue line shows

the unweighted result. ASD calculates a pairwised standard difference between covariates within every single exsiting pairs of covariates. Similarly, Red lines give us the weighted result, and the blue lines stand for unweighted result. Figure 2 showed how OW performed to balance out the confounding covariates in our statin data. The closer to 0 of the red lines, the better balancing result we are achieving using GOW among all statin treatment arms. Even though GOW did not make the PSD and ASD to be 0, it shows a considerable improvement in balancing out the variables in the dataset.

5.1.4 Overlap check

In Figure A.5(b) and Figure A.5(a), we presented one of important covariate age overlap check for those 10 Statin treatment plans. Here we use a typical confounding variable age as an example. The weighted distribution (Figure A.5(a)) has better and closer overlap for statin PI_moderate treatment option for the age variable comparing to unweighted distribution (FigureA.5(b)). Another two overlap check for Charlson Quan comorbidity score and pre-LDL-C values are shown in supplementary table.

5.1.5 Counterfactual prediction

Figures A.6(a) and Figure A.6(b) show examples of ten individualized counterfactual survival plots for SAS and discontinuation for four treatment plans (AT_High, AT_Moderate, SI_Low, and RO_Moderate). These plots were produced by the SAS and discontinuation counterfactual survival models. We use the trained model to predict survival function and it will give us the survival probability $P(Y_t > y)$, and then we use $1 - P(Y_t > y) = P(Y_t < y)$ to retrieve the corresponding cumulative SAS event risks and discontinuation event risks within the time period between statin index date and specific time point y. More surivival curves plots for all statin treatments are provided in supplement.

5.1.6 Optimization showcase

Here, we demonstrated our pipeline to select the personalized statin treatment plan (PSTP) for 10 simulated patients using multi-objective optimization when minimizing SAS risks, minimizing discontinuation risks, and maximizing LDL-C reduction at 30 to 360 days. In the illustration shown in Figure A.7(a) - A.7(f), Atoravatstatin with High intensity level, Atorvastatin with Moderate intensity level, Pitavastatin with Moderate intensity level, Rosuvastatin with High intensity level, and Simivastatin with Moderate intensity level (Figure A.7(a) in purple) were selected at 30 days as the Pareto front optimal treatment after thorough consideration of SAS risks, discontinuation risks, and LDL-C reduction for patient, as shown in purple color. Among all Pareto front optimal treatment options, TOPSIS selected one of the best from those treatment plans in purple. At different time points, the selected optimal treatments can be different. There are fewer pareto optimal treatments when increase the number of days since statin index date.

5.1.7 Clinical Trial Simulation

Figure A.7 represents the summary of the counts each CTS arm for all statin treatment options. We chose 180 days, 210 days, and 360 days survival risks as examples when selecting out the optimal treatment. Pitavastatin with moderate intensity (Figure 8) was the most recommended by the optimization based on 180 days survival risks. For the survival risks within 1 year, Atorvastatin with high intensity, Atorvastatin with moderate intensity, and Rosuvastatin with moderate intensity were the most recommended when considering all 3 prediction targets. The Rosuvastatin with high intensity was the most recommended by optimization for 210 days of the survival risks for both SAS and Discontinuation risks.

Figures A.8 depicts the mean value of predicted survival risks over time for each treatment arms for our cohort data. The PSTP Arm calculated using multi-objective optimization performed the best in SAS. The elbow in figure A.8(a) in 180 days and 210 days showed that SAS prediction increases rapidly after 180 days of the study period. For the discontinuation risks, the PSTP showed similar risks compared to other arms. When it comes to LDL-C reduction (Figure A.8(c)), the value is higher when patients can stay on the treatment for a longer time. The reason why Figure A.8(c) only has the horizontal line for Random, Guideline, and Practice arms is that the outcome variable LDL-C reduction is calculated as the subtraction of the pre-index average LDL-C value and post-index average LDL-C value, which is static over time. LDL-C are recorded in laboratory data and the calculation of the LDL-C reduction is static. Our PSTP is determined by MOO considering SAS, Discontinuation and LDL-C reduction, so the dynamic SAS and discontinuation survival risks gives us dynamic PSTP arm across different time points, which is corresponding to different LDL-C reduction in Figure A.8(c). In general, we see PSTP has a lower SAS risk and similar discontinuation risk compared to all other arms. The PSTP selected in later period has a better LDL-C reduction as shown in Figure Figure A.8(c).

5.2 Result from AHC-IE data

5.2.1 Data

The Table A.2 summarizes the characteristics of patients across different statin generic agents, including atorvastatin, lovastatin, pravastatin, rosuvastatin, and simvastatin. A total of 95,555 patients were included for the predition model. And we can see that most of patients are using atorvastatin and simvastatin. Rosuvastatin cohort has larger percentages of quitting the statin medication (with one year discontinuation rate of 46.6%), while Simvastatin cohort has smaller percentage of discontinue the medication (with one year discontinuation rate of 34.7%). They all have pretty close rate of SAS event, where rosuvastatin and pravastatin has largest SAS even rate of approximately 33.6% and 33.8%, and simvastatin has the smallest SAS event rate of 31.8%.

The Table A.3 presents patient characteristics across different statin intensity categories: high, moderate, and low. A total of 95,555 patients were included, with 14,855 receiving high-intensity statins, 66,208 moderate-intensity, and 14,492 low-intensity. The overall statin discontinuation rate within one year was 38.1%, with slightly lower discontinuation rates observed for moderate-intensity statins (36.0%) and higher rates for high-intensity statins (44.2%). High intensity group has larger SAS rate (36.6%), while moderate intensity group has the smallest SAS rate (30.5%). It also included the basic descriptions of other variables used for the model, including comorbidities, vital signs, hospital encounter information, and so on.

5.2.2 Clinical Trial Simulation

The AHC-IE data does not have longitudinal information for the patients, in which case the counterfactual prediction was not applied with survival model for different time points. The baseline model and two-stage model shows slightly different outcomes based on the clinical trial simulation plots Figure A.9 and FigureA.10. But the Atorvastatin-High and Simvastatin-Moderate are two major statin options recommened by PSTP. Rosuvastatin-moderate was recommended quite often in baseline model while it is very less likely to be recommended in based on two-stage model. Both models do not recommend Lovastatin. Simvastatin-Moderate is the only one that shows similar amount of patients between clinical practice and PSTP arms, in both models. There are certain amount of patients who were prescribed with Pravastatin-Low, while Pravastatin-Low was not quite recommended by our PSTP.

The example of individual multi-objective optimization (i.e., pareto optimization) (A.11) also shows different result between baseline model and two-stage model. It is very consistent with the CTS result that both models recommend Rosuvastatin-High and Atorvastatin-High, while baseline model takes Pravastatin-Moderate as another optimal option for this patient.

Interestingly, the CTS result from two different data source (OptumLabs and AHC-IE) are showing some similar pattern. Both shows that the Atorvatstatin-High are the quite recommended drugs given 1 year follow-up period. AHC-IE data does not have Pitavastatin patients

5.3 Result from Simulations

Figure A.12 - Figure A.15 shows the aggregated results for the simulations process for comparing the effecacy and effectiveness of the baseline model vs. Two stage model. The first impression is that two of the Two-stage model shows better performance in terms of ROAUC comparing baseline model. Under the scenario of Overlapped treatment effects with interactions, the Two-stage models also shows significant better performance in terms of the Brier score for both training and testing data. As the treatment number increases, the worse of the performance of baseline model is, while Two-stage model maintains the horizontal curve of the performance metric (i.e., ROAUC) even though it also has little

downward trend. The downward trend is quite significant from treatment number = 4 to treatment number = 9, and the marginal decrease is smaller as the treatment number increases.

The regular two-stage model result is in orange while Two-stage model with Reversed Order (TSRO) is in green line. For all scenarios, TSRO has pretty similar performance to regular Two-stage model except the case when treatment number is close to 21. In our simulation, when treatment number = 21, it only has 3 dosage/intensity levels and 7 treatment agents, while for treatment number = 20, it has 4 dosage/intensity levels and 5 treatment agents. When treatment number = 21, the second stage of the model utilizes 7 treatment agents cohort separately for training the counterfactual prediction model, in which case it has smaller training data size comparing to regular Two-stage model because the regular one takes dosage/intensity levels (3 levels) as the second stages. The regular Two-stage model has larger training data size for the second stage, which is why the result has a very obvious contrast for regular Two-stage model and TSRO when treatment number = 21. The design of the treatment space for treatment number = 21 caused the little re-bounce of the performance for both baseline model and Two-stage model.

All p-values are generated using mixed-effect ANOVA [40] by regarding treatment numbers as the repeated measures of each model. Regular two-stage model and TSRO are combined as group A and baseline model is regarded as group B, to test the significant difference between group A and group B. Two-stage model has universal significant superior performance across all scenarios in terms of AUROC while it only shows better performance under Overlapped Treatment effects, with interaction terms, in terms of Brier score. For example, in figure A.12, when treatment number reaches to 20, baseline model's testing AUROC gives us the number around 0.76 while both two-stage model gives the performance around 0.83. It shows the largest differences in terms of the model performance between baseline CF model and two-stage model/TSRO. It also infers that both two-stage model would perform better than the baseline model as the treatment number increases.

Both Training AUC and testing AUC and training Brier and testing Brier score didn't show much differences under all 4 scenarios.

Figure A.16 - Figure A.19 shows the shows the individual treatment results for the simulations process for comparing the efficacy and effectiveness of the baseline model vs. Two stage model, for treatment a_mod, a_high, b_mod, b_high. Since there are only 4 treatment options: a_low, a_high, b_low, b_high, in 4 treatment scenario, those 4 treatments would be the most comprehensive examples out of all other treatments because they exist in all treatment number scenarios. In general, the baseline model tends to fluctuates across different treatment numbers, while the two of the two-stage model has very stable trend with treatment number increases. We still don't see much differences between training and testing metric results. For data scenarios with interaction terms, two-stage model performs better than baseline model, in terms of both ROAUC and Brier score.

5.3.1 Comparison between ground Truth and predicted treatment effects

The causal differences from the prediction data is also calculated to see if they are consistent with our original δ value set during simulation. The data simulations conducted without interaction terms produced causal differences that were highly consistent with the expectations set during the data simulation process. Specifically, the predicted outcomes from the model closely aligned with the predefined causal effects embedded within the simulated datasets, suggesting that interaction terms has very significant impact on prediction if we didn't consider interactions and non-linearity during the prediction process. This consistency for scenarios of not having interactions underscores the robustness of the model's ability to accurately capture the direct causal relationships between treatment and outcomes as originally specified, using our counterfactual prediction model. These findings highlight the reliability of the counterfactual model in scenarios where interactions are not present, providing confidence in the straightforward interpretation of causal differences under the simulated conditions.

Chapter 6

Discussion

6.1 Model: Baseline model

In this thesis, the baseline model is a counterfactual prediction model using a double machine learning approach that was employed to analyze patient data from two distinct sources: OptumLabs and the AHC-IE database from the University of Minnesota, Twin Cities. The approach involves first applying a propensity score model to estimate the probability of treatment assignment, followed by outcome models that utilize calculated overlapping weights from the propensity score model. This methodology offers several advantages over traditional machine learning models that simply use all covariates and treatment variables to predict outcomes.

1. Causal Inference vs. Predictive Modeling: The primary distinction between the double machine learning approach and a standard predictive model lies in their objectives. Traditional machine learning models are designed to optimize predictive accuracy, often without considering the causal relationships between variables. They may inadvertently capture spurious correlations, leading to biased estimates when the goal is to infer causal effects. In contrast, the double machine learning approach explicitly accounts for the causal structure by modeling the treatment assignment through the propensity score. This helps to isolate the effect of the treatment on the outcome from confounding variables, leading to more reliable causal predictions. One of the main challenges in observational studies and observational data is the presence of confounding bias, where covariates affect both the treatment and the outcome. The double machine learning approach mitigates this by using the Generalized Propensity score and Generalized Overlap Weights to adjust for confounders, thereby balancing the treatment groups on observed covariates. The calculated overlapping weights further refine this adjustment by focusing on the region of overlap between treatment groups, ensuring that the comparison among all treatment groups is more meaningful. In contrast, a single machine learning model might not effectively handle confounding bias, especially when the covariates have complex relationships with the treatment and outcome.

2. Handling Multiple Treatments scenarios: Generalized Overlap Weights (GOW) play a vital role in

handling multiple treatments comparisons and emphasizing the patients who are at the clinical equipoise, where patients who are having larger uncertainty and even probabilities being assigned to any possible treatment options are up-weighted, while patients who have very certain probabilities to be assigned to a specific treatment will be de-emphasized by nature of the Overlap Weights. [33] GOW has a very superior characteristics that it can incorporate multiple treatment options. Li et al. proposed this method in 2019 for doing multiple treatment comparisons (more than 3 treatments)

3. Robustness to Model Misspecification: Double machine learning models are less sensitive to model misspecification compared to traditional approaches. By separating the treatment assignment process from the outcome modeling, the double machine learning framework allows for more flexible model specifications. Even if one of the models (propensity score or outcome model) is misspecified, the bias in the treatment effect estimation can be reduced or corrected by the other model. In contrast, a single machine learning model that simultaneously considers all variables is more prone to overfitting and misspecification, particularly when dealing with high-dimensional data or complex relationships between variables.

4. Generalizability Across Datasets: The application of the double machine learning approach to both OptumLabs and AHC-IE data demonstrates its robustness and generalizability. By relying on the causal structure of the data, rather than purely predictive performance, this counterfactual prediction framework is better suited to produce consistent and interpretable results across different datasets. In contrast, a traditional machine learning model may perform well on one dataset but poorly on another due to differences in variable distributions or treatment assignment mechanisms, leading to less generalizable findings.

5. Interpretation and Policy Implications: Finally, the counterfactual prediction framework using GPS and GOW provides estimates of causal effects that are more interpretable and actionable, particularly in the context of policy decisions or clinical guidelines. Understanding the causal impact and the counterfactual predictions of treatments on outcomes is critical for informing treatment decisions, developing guidelines, and shaping healthcare policy. A traditional machine learning model, while potentially accurate in prediction, does not directly offer insights into causality, limiting its utility in these contexts.

In conclusion, while traditional machine learning models excel in predictive tasks, the double machine learning approach employed in this thesis offers a more nuanced and robust framework for causal inference. By effectively addressing confounding bias, improving model robustness, and providing generalizable and interpretable results, this approach is better suited for understanding the causal relationships in observational healthcare data, making it a valuable tool for informing clinical and policy decisions.

6.2 Model: Two-stage model

The idea of innovating this two-stage model implementation for counterfactual prediction originates from larger number of treatment options where some treatment arm/cohort has very small sample size. When we simply use a trained model from small sample size and generalize the model to the whole dataset, it naturally has several problems. Firstly, the model might be overfitted. A model trained on small data might become overly complex and it might learn the noise from the data resulting in poor performance for prediction. Small data size might also cause insufficient representation problem. It often fails to capture the full diversity and variability of the underlying pattern of the specific treatment population. Another problem is the high variance. Trained model from small dataset can lead to high variance in model performance, meaning the model's prediction can change dramatically with small changes in the data. Therefore, the counterfactual prediction from a specific treatment cohort can be less trustworthy.

The two-stage model delineates the components of treatment definition—specifically, treatment agents and treatment dosages/intensities—into two distinct parts. This approach helps to capture common characteristics across treatments while mitigating data loss in the counterfactual prediction model. In the first stage, we train a Generalized Propensity Score (GPS) model for all treatment agents, calculating Overlap Weights (OW) for each. Next, we use the cohort from a specific treatment agent (e.g., the Atorvastatin cohort) to train a dosage propensity score model. This "dosage propensity score model" predicts the "counterfactual propensity score" for the entire population under that specific treatment agent.

The counterfactual propensity score answers the following question: What would be the propensity of different dosages/intensities (e.g., High, Moderate, Low) if all patients were to receive this specific treatment agent (e.g., Atorvastatin)? Using this counterfactual propensity score and the corresponding calculated GOW, we then train our outcome model for this treatment branch (e.g., Atorvastatin-High) using the whole high-intensity cohort as the training data and the second-stage "counterfactual propensity score" calculated GOW as sample weights. Users can subsequently make counterfactual outcome predictions for the entire population.

This counterfactual GOW allows us to create a pseudo-population representing the high-intensity cohort under Atorvastatin, even though the training data comprises all high-intensity cohorts. It also enables cross-borrowing of data between different treatment dimensions, thereby enhancing the validity of counterfactual predictions while ensuring a larger training dataset for the outcome model.

The separation of treatment agents from their dosages or intensities within the two-stage model is instrumental in capturing shared characteristics across different treatment regimens. This methodological distinction enables a more nuanced understanding of how varying agents interact with dosage levels,

thereby enhancing the accuracy of the calculated propensity scores and improving the reliability of subsequent counterfactual predictions. By acknowledging that different treatments may share underlying mechanisms or patterns, this approach strengthens the model's ability to discern these commonalities, which are critical for producing robust and generalizable predictions.

A another key advantage of the two-stage model is its ability to mitigate data loss, a common challenge in counterfactual prediction models, especially when dealing with multiple treatments and varying dosages. The model addresses this issue by maintaining a comprehensive view of the treatment landscape in the first stage, where treatment agents are analyzed holistically. In the second stage, the model delves into specific dosage levels, allowing for more detailed predictions without compromising data integrity and quantity. This dual-stage approach ensures that the maximum amount of available data is utilized, thereby enhancing the model's capacity to generalize across diverse treatment scenarios.

The introduction of counterfactual propensity scores in the model's second stage is particularly significant for exploring hypothetical scenarios, such as those in clinical settings where direct experimentation may not be feasible or ethical. By predicting the outcomes of a situation where all patients receive a particular treatment agent at varying dosage levels, the model provides valuable insights into potential effects and side effects associated with different treatment strategies. This capability is crucial for informing clinical decisions and optimizing treatment protocols, as it allows for a thorough exploration of various treatment pathways.

Another critical advantage of the two-stage model is the use of Overlap Weights (OW) to create a pseudo-population that mirrors the high-intensity cohort under a specific treatment agent. This approach ensures that the model remains rooted in real-world data while exploring counterfactual scenarios, bridging the gap between observed and hypothetical data. The pseudo-population serves as a robust foundation for drawing reliable conclusions about treatment effects, while also allowing for data borrowing across different treatment dimensions. This cross-borrowing enhances the model's robustness, reducing the risk of overfitting and increasing the reliability of its predictions.

The implications of the two-stage model extend beyond the immediate context of statin treatment analysis. This framework is adaptable to other medical treatments where dosage and intensity play a crucial role in patient outcomes. Moreover, the model's structure could be applied in other fields, such as economics or social sciences, where treatment effects are complex and multifaceted. The ability to accurately explore counterfactual scenarios with this model opens up new avenues for research and decision-making, providing a robust, data-driven foundation for policy-making and clinical practice.

6.3 Optum Labs data

The updated PSTP framework in this article fully addressed the three major issues in our previous study. Counterfactual prediction and the potential confounding issue were addressed through incorporating

GPS and GOW into our modeling procedure. In addition, we used LDL-C percentage reduction as an example of a benefit and demonstrates how we maximize benefits and minimize risks in the 3D Multi-Objective Optimization process. Finally, we used clinical trial simulations to evaluate the proactive strategy, PSTP.

Instead of doing intention-to-treat analysis, the individualized survival model A.6(a)A.6(b) provided us the flexibility of retrieving risks on the per-protocol basis, where the survival functions give us dynamic predictions for an individual's SAS and discontinuations risks on one year follow-up period. Rather than using SAS and discontinuation as dichotomous labels for model training, doctors or patients have the flexibility to choose which time point event risks they care about. The PSTP along all time points showed better SAS results while it didn't show a better result for discontinuation. The predicted SAS result has larger influence on deciding the PSTP optimal treatment options (Figure 8). We could have different weights for those three factors(SAS, Discontinuation, LDL-C reduction) for deciding the PSTP in the future work. In addition, LDL-C reduction data is not abundant enough to do dynamic analysis across all time points. We can fit LDL-C reduction model for each time point to generate the dynamic plot like SAS and Discontinuation in the future work if data allows.

In our study, the Random survival forest was used for SAS and Discontinuation risk prediction, and the linear regression with L1 regularization was used for LDL-C reduction prediction. The extrapolation using linear models is critical to treatment cohorts especially when the sample size is small. Due to lack of data such as Lovastatin or Pitavastatin, the model is not able to fully learn the characteristics of patients who were prescribed for these two treatments. Linear models with extrapolation ensured that we could make individualized predictions. [41]

By adding the LDL-C percentage reduction as another new objective for the modeling procedures, the updated PSTP fully addressed the balance between benefits and risks when making data-driven prescriptions. The individual 3D optimization shows an example of deciding statin treatment at different time points. In general, the longer time after Statin index date, the fewer optimal Statin options. This is because patients have higher probabilities of experiencing SAS and/or discontinuation when we have a sufficiently long time of using statin therapies. By taking advantage of EHR data and claims records, our counterfactual survival model framework, therefore, is able to identify the optimal treatment options at different time periods.

There might be some other data and model limitations. We dropped certain amount of data who don't have LDL-C value records in our data. Patients might also purchased statins in a way that cannot be tracked by the pharmacy claims. Multiple causal related model assumptions including no unmeasured confounding, positivity, generalizability, and clinical equipoise also restricted the practical use of our PSTP.

6.4 AHC-IE data

The AHC-IE data does not have the timeline information for survival model. Figure A.9 and A.10 shows the comparison of Statin Treatment Assignment Strategies in Baseline and Two-Stage Models. The two figures illustrate the distribution of statin treatments across four different assignment strategies: Random, Clinical Practice, Clinical Guideline, and PSTP. These strategies are compared across various statin agents and their respective intensity levels, under both the baseline model and the two-stage model.

1. Random Assignment Arm In the Random assignment strategy, the distribution of patients across different statin treatments is relatively even, as expected. This lack of bias results in a more uniform distribution across the different statin-intensity combinations, with no particular preference towards any specific treatment. Random arm mainly serves as a reference arms for other arms to compare, and we can see how much patients there would be if we simply randomly assign each patient to each statin treatment options.

2. Clinical Practice Arm The Clinical Practice arm reflects real-world prescribing patterns, where certain statins and intensity levels are more commonly prescribed based on historical trends and clinician preferences. In the baseline model, the Clinical Practice arm shows a higher number of patients assigned to atorvastatin and simvastatin, particularly at moderate to high intensities, reflecting their widespread use in clinical settings. The two-stage model, it is using the same data where clinical practice arms are exactly the same as the baseline model's result. Atorvastatin and simvastatin are mostly prescribed two statins in reality. It is highly possible that there would be some selection bias with higher prevalence of those two statin types, while weighting methods using GPS and GOW are aimed to mitigate those issues.

3. Clinical Guideline Arm The Clinical Guideline arm represents treatment assignments based on established guidelines, which typically prioritize efficacy and safety over using 2013 statin prescription guidelines. This arm is generated using patients' age, gender, comorbidities, and other possible factors (i.e., LDL-C level). In both models, the guideline-based assignments lead to a noticeable concentration of patients in treatments that are considered most effective according to guidelines, such as atorvastatin and rosuvastatin at high intensities.

4. PSTP (Personalized Statin Therapy Plan) Arm The PSTP arm aims to tailor treatment assignments based on 3 important personalized patient characteristics (predicted SAS risk, discontinuation risk, LDL-C reduction percentage), which might deviate from standard practice or guidelines. In the baseline model, PSTP shows a more diverse distribution compared to the Clinical practice arm, reflecting individualized treatment decisions. The two-stage model enhances counterfactual prediction capacity by further separating treatment spaces into 2 dimensions(i.e., treatment agents

and dosages level). This results in a more nuanced distribution, where certain more commonly prescribed statins, such as lovastatin at lower intensities, receive no patients compared to the other strategies by PSTP strategy. The pravastatin-moderate was less likely prescribed for patients in clinical practice, while it was quite often recommended and selected as the optimal treatment option for large number of patients. This indicates that the two-stage model might be better at capturing and implementing personalized treatment strategies, leading to a broader range of statin-intensity combinations being utilized.

5. Summary Overall, the CTS gives us a general sense of what CF prediction and optimization tell us which statin to prescribe would be the best for patient. The two-stage model provides a more refined and tailored approach to treatment assignment across all strategies compared to the baseline model. While it does not necessarily tell us that the two-stage is superior than the baseline model. This actually triggers the problem of the model evaluation for counterfactual prediction that how we can confidently tell which model performs better and which model performs worse? Data simulation would be a very direct methodology for us to tell which of these two models performs better than the other one. By using the simulation, we simulate the ground truth of each potential outcomes under every single treatment options, and then use direct ML evaluation metric to show and compare the model performance, such as ROAUC, Brier score, etc. to tell which model performs better.

6.5 Simulation

In the chapter of simulation, we evaluate the performance of counterfactual prediction models using a novel two-stage approach compared to a traditional one-stage baseline model. The evaluation is carried out across four different data simulation scenarios to investigate the models' robustness as the complexity of treatment options and interactions increases. Specifically, we assess model performance using AUC and Brier scores as the primary metrics.

Scenario 1: Additive Treatment Effects

The first scenario considers a simple case where the treatment effects are additive. As the number of treatment options increases, the performance of the baseline model (one-stage) demonstrates a significant decline, as evidenced by the sharp decrease in both AUC and Brier scores (see Figure A.12 in appendix). This decline highlights the model's limitation in handling a growing number of treatment options effectively. In contrast, the two-stage model, which separates the treatment definition into distinct stages for treatment agents and dosages/intensities, maintains relatively stable performance. This stability suggests that the two-stage model is better equipped to manage the complexity associated with an increasing number of treatments.

Scenario 2: Additive treatment effects with Interaction and Polynomial Terms

In this scenario, the data simulation includes additive treatment effects along with interaction and polynomial terms, while the modeling only considers regular terms. The one-stage model struggles even more with the introduction of interactions and non-linearities, as reflected by a more pronounced decline in performance metrics compared to the two-stage model. The two-stage model, although not perfect, still manages to retain better predictive power, indicating its superior ability to generalize in the presence of unaccounted interaction effects A.13.

Scenario 3: Overlapped Treatment Effects

When the treatment effects overlap, where some treatments produce identical effects, the baseline model's performance degradation is even more evident. The overlap creates ambiguity in the treatment-effect relationship, making it harder for the one-stage model to distinguish between different treatments. As shown in A.14, this results in a significant drop in AUC and an increase in Brier scores. The two-stage model, however, shows resilience, maintaining its performance despite the overlap, demonstrating its robustness in scenarios where treatment effects are not distinct.

Scenario 4: Overlapped Treatment Effects with Interaction and Polynomial Terms

The fourth scenario further complicates the simulation by adding interaction terms to the overlapped treatment effects. This represents the most challenging scenario, where both overlap and interactions are present. The one-stage model's performance deteriorates rapidly, with the lowest AUC scores and highest Brier scores observed across all scenarios (see A.15). On the other hand, the two-stage model continues to outperform the baseline model, albeit with some decline in performance. This scenario underscores the two-stage model's capacity to handle the complexities of real-world treatment scenarios where effects are neither independent nor distinct.

Across all four scenarios, the two-stage model consistently outperforms the traditional one-stage baseline model as the number of treatment options increases. The two-stage model's ability to maintain higher AUC and lower Brier scores, even in the presence of interactions, non-linearities, and overlapping treatment effects, indicates its robustness and potential for application in complex counterfactual prediction tasks, such as statin treatment evaluation. The findings suggest that as the number of treatments and their interactions grow, the two-stage model provides a more reliable framework for counterfactual predictions, making it a valuable tool in personalized medicine and treatment optimization.

Beyond the comparison between the baseline and two-stage models, the differences between the regular two-stage model and the two-stage reversed order (TSRO) model reveal some intriguing insights. Notably, when the total number of treatments reaches 20 or 21, the regular two-stage model, which treats dosage/intensity levels as the second stage, consistently outperforms the TSRO model, where statin agents are the second stage. This distinction becomes particularly apparent when the number of treatments is 21. In this scenario, there are 7 treatment agents (a, b, c, d, e, f, g) combined with 3 dosage/intensity levels (Low, Moderate, High). When dosage/intensity levels are used as the second stage, the model benefits from larger datasets, allowing for greater cross-borrowing of patient outcomes

for counterfactual prediction. Consequently, the model demonstrates superior performance. Conversely, when statin agents are assigned as the second stage in the TSRO model, the data is split into 7 separate cohorts for the second stage, each requiring independent training and prediction. This fragmentation leads to poorer model performance, as reflected in both AUC and Brier scores.

6.6 Significance

This thesis presents a rigorous evaluation of counterfactual prediction models in the context of statin treatment, providing valuable insights into the comparative effectiveness of different modeling strategies. The findings underscore the importance of model structure in handling complex treatment scenarios in terms of statin types and dosages/intensities, particularly as the number of treatment options increases in the data simulation. The OptumLabs and AUCH-IE real world data also presents how we can optimize the treatment plans using the counterfactual prediction model and the optimization algorithm.

By demonstrating the superior performance of the two-stage model over the traditional one-stage baseline model, especially in scenarios with additive treatment effects, interactions, and overlapping effects, this research contributes to the growing body of evidence supporting more nuanced approaches to treatment modeling. The two-stage model's ability to maintain robust performance across a range of complex scenarios suggests that it is better suited for real-world applications where treatment effects are multifaceted and interdependent.

Moreover, the comparison between the regular two-stage model and the two-stage reversed order (TSRO) model reveals critical insights into the optimal sequencing of model stages. The significant performance gains observed when dosage/intensity levels are used as the second stage, particularly when the number of treatments is high, highlight the importance of model design in maximizing predictive accuracy. These findings suggest that careful consideration must be given to the order of model stages to enhance the reliability of counterfactual predictions.

The implications of this research extend beyond the specific application of statin treatment to broader contexts in personalized medicine. As healthcare increasingly moves towards individualized treatment plans, the ability to accurately predict counterfactual outcomes based on complex treatment regimens becomes essential. The methodologies and insights developed in this thesis provide a foundation for future research and practical applications in diverse medical and clinical settings.

In conclusion, this thesis not only advances the understanding of counterfactual prediction models but also provides actionable guidance for their implementation in clinical practice and counterfactual model evaluation. The demonstrated superiority of the two-stage model and the importance of stage sequencing offer a pathway to more accurate and effective treatment predictions, ultimately contributing to better patient outcomes and more personalized healthcare.

6.7 Limitation

The whole thesis working on counterfactual prediction and its evaluation is not without its limitation. One key limitation is the heavy reliance on the assumptions underlying the GPS model, such as no unmeasured confounding, correct model specification, and the assumption of overlap. If any of these assumptions are violated, the resulting treatment effect estimates could be biased, which undermines the reliability and validity of the findings. This is particularly critical given the complexity of real-world data, where unobserved factors can influence both treatment assignment and outcomes.

Another limitation is related to the handling of high dimensionality and complex treatment scenarios. The performance of the model can be sensitive when dealing with numerous treatments, as observed in scenarios with 20 or 21 treatment groups. Differences in performance between the regular two-stage model and the TSRO model indicate that model complexity and computational burden may affect the robustness of the results. Such variability highlights potential challenges in applying these methods in practice, particularly in settings with many treatment options or where computational resources are constrained.

The generalizability of your findings may also be limited by the specific datasets used, namely OptumLabs patient data and AHC-IE data from the University of Minnesota, Twin Cities EHR database. While these datasets provide valuable insights, they may not represent other populations, healthcare settings, or medication types beyond statins. Differences in demographics, clinical practices, or healthcare systems can influence the external validity of your conclusions, making it crucial to be cautious when extrapolating the results to broader contexts.

Additionally, the separation of treatment definitions into agents and dosages might not fully capture the complexities of clinical practice, such as variations in patient adherence, off-label use, or potential misclassification of treatment intensity. These factors can introduce bias, affecting the model's ability to accurately predict treatment effects. Similarly, the overlapping weights calculated from the propensity score model are essential for balancing treatment groups, but extreme weights or sparse data can lead to convergence issues and unstable estimates.

Residual confounding and bias from unmeasured variables, such as genetic factors, lifestyle behaviors, or patient preferences, remain concerns even with advanced machine learning techniques. These unmeasured factors could skew results, complicating causal interpretations. Additionally, the reliance on simulated data for different treatment numbers may not fully replicate the variability seen in real-world patient responses, which could lead to overly optimistic assessments of model performance that might not hold in practical applications.

Another challenge lies in the sensitivity of model outcomes to the specific algorithms, hyperparameters, and performance metrics chosen. Small changes in model specifications could result in different conclusions, underscoring the importance of conducting sensitivity analyses to assess the robustness of

your findings. Furthermore, missing data on treatment, outcomes, or covariates can introduce bias, particularly if the data are not missing at random. While imputation methods can mitigate this issue, they may not fully eliminate the impact of missingness on the study's results.

Finally, the computational complexity and scalability of the two-stage approach, particularly with high-dimensional data and double machine learning techniques, pose practical challenges. The significant computational resources required may limit the feasibility of applying these methods in some research or clinical settings, especially those with limited access to advanced computing infrastructure. Addressing these limitations thoughtfully in your thesis will demonstrate a critical understanding of the constraints of your methodology and help provide a balanced interpretation of your findings, highlighting both the strengths and areas for future refinement.

Chapter 7

Conclusion

This thesis has focused on advancing the understanding of multilevel treatment options for statin therapy and their impact on patient outcomes, particularly through the lens of counterfactual prediction using a two-stage model. The research presented herein has demonstrated that personalized medicine, driven by machine learning and causal inference methods, can significantly improve treatment efficacy and minimize adverse side effects.

1. Summary of Contributions

The primary contribution of this work lies in the development and application of baseline counterfactual model and two-stage counterfactual prediction model. By separating the treatment process into two distinct stages — statin agents and dosage levels — two-stage model has shown superior performance in predicting clinical outcomes compared to baseline models. Specifically, these approaches enable a more nuanced understanding of the complex interactions between treatment choices and patient-specific factors, thus providing more accurate recommendations for personalized statin treatment plans (PSTPs).

Through the use of real-world data from the OptumLabs and AHC-IE datasets, we demonstrated the practical applicability of this model in a clinical context. The results from these datasets showed that our two-stage model could effectively predict key outcomes such as Statin-Associated Symptoms (SAS), discontinuation rates, and LDL-C reduction. Furthermore, the integration of overlap weights and doubly robust estimation enhanced the model's ability to control for confounding factors and produce reliable predictions.

2. Implications for Personalized Medicine

The findings from this research have important implications for the field of personalized medicine. The use of counterfactual predictions allows healthcare providers to simulate various treatment scenarios and choose the optimal strategy for individual patients. This is particularly relevant in the case of statin therapy, where multiple drug options and dosages are available, and patient responses can vary widely. By leveraging predictive models based on real-world data, clinicians can make more informed decisions, reduce trial-and-error in medication management, and ultimately improve patient adherence and health

outcomes.

Additionally, the use of machine learning techniques to predict counterfactual outcomes highlights the potential of data-driven approaches in optimizing treatment strategies. As the healthcare industry continues to evolve, integrating such methods into routine clinical practice could lead to more effective and efficient patient care, reducing the burden of adverse events and improving the overall quality of life for patients undergoing long-term statin therapy.

3. Limitations and Future Research

While this thesis has made significant advancements in the field of counterfactual prediction for statin therapy, there are still several limitations to address. One key challenge is the reliance on high-dimensional data, which can complicate model development and increase computational complexity. Future research should focus on refining algorithms to handle large-scale datasets more efficiently and explore methods to mitigate the impact of unmeasured confounders.

Moreover, while this thesis has focused on statin treatment, the methodology developed could be extended to other areas of personalized medicine, such as genomics or treatment optimization for chronic diseases. Further research should explore the generalizability of the two-stage model to other medical domains and investigate how real-time clinical decision support systems can incorporate counterfactual predictions to enhance patient care.

4. Conclusion

In conclusion, this thesis has contributed valuable insights into the use of counterfactual prediction and two-stage modeling for optimizing statin therapy. By combining advanced machine learning techniques with causal inference methods, we have developed a robust framework for personalizing treatment plans based on individual patient characteristics. These findings pave the way for future research in personalized medicine, where data-driven approaches will continue to play a critical role in improving patient outcomes and advancing healthcare as a whole.

References

[1] Omeed Sizar, Swapnil Khare, Radia T Jamil, and Raja Talati. Statin medications. 2017.

[2] Paul D Thompson, Gregory Panza, Amanda Zaleski, and Beth Taylor. Statin-associated side effects. *Journal of the American College of Cardiology*, 67(20):2395–2410, 2016.

[3] Boguang Sun, Pui Ying Yew, Chih-Lin Chi, Meijia Song, Matt Loth, Rui Zhang, and Robert J Straka. Development and application of pharmacological statin-associated muscle symptoms phenotyping algorithms using structured and unstructured electronic health records data. *medRxiv*, pages 2023–05, 2023.

[4] Jerome D Cohen, Eliot A Brinton, Matthew K Ito, and Terry A Jacobson. Understanding statin use in america and gaps in patient education (usage): an internet-based survey of 10,138 current and former statin users. *Journal of clinical lipidology*, 6(3):208–215, 2012.

[5] Chih-Lin Chi, Jin Wang, Pui Ying Yew, Tatiana Lenskaia, Matt Loth, Prajwal Mani Pradhan, Yue Liang, Prashanth Kurella, Rishabh Mehta, Jennifer G Robinson, et al. Producing personalized statin treatment plans to optimize clinical outcomes using big data and machine learning. *Journal of biomedical informatics*, 128:104029, 2022.

[6] Hao Zou, Peng Cui, Bo Li, Zheyan Shen, Jianxin Ma, Hongxia Yang, and Yue He. Counterfactual prediction for bundle treatment. *Advances in Neural Information Processing Systems*, 33:19705–19715, 2020.

[7] Pierre Gutierrez and Jean-Yves Gérardy. Causal inference and uplift modelling: A review of the literature. In Claire Hardgrove, Louis Dorard, Keiran Thompson, and Florian Douetteau, editors, *Proceedings of The 3rd International Conference on Predictive Applications and APIs*, volume 67 of *Proceedings of Machine Learning Research*, pages 1–13. PMLR, 11–12 Oct 2017.

[8] Michael J Wilkinson, Luke J Laffin, and Michael H Davidson. Overcoming toxicity and side-effects of lipid-lowering therapies. *Best Practice & Research Clinical Endocrinology & Metabolism*, 28(3):439–452, 2014.

[9] Jonathan Tyler, Sung Won Choi, and Muneesh Tewari. Real-time, personalized medicine through wearable sensors and dynamic predictive modeling: a new paradigm for clinical medicine. *Current opinion in systems biology*, 20:17–25, 2020.

[10] Lijing Lin, Matthew Sperrin, David A Jenkins, Glen P Martin, and Niels Peek. A scoping review of causal methods enabling predictions under hypothetical interventions. *Diagnostic and prognostic research*, 5:1–16, 2021.

[11] Joseph E Lucas, Taylor C Bazemore, Celan Alo, Patrick B Monahan, and Deepak Voora. An electronic health record based model predicts statin adherence, ldl cholesterol, and cardiovascular disease in the united states military health system. *PLoS One*, 12(11):e0187809, 2017.

[12] Eonji Na, Sunyoung Cho, Dae Jung Kim, Junjeong Choi, and Euna Han. Time-varying and dose-dependent effect of long-term statin use on risk of type 2 diabetes: a retrospective cohort study. *Cardiovascular diabetology*, 19:1–11, 2020.

[13] Ankur Pandya, Stephen Sy, Sylvia Cho, Milton C Weinstein, and Thomas A Gaziano. Cost-effectiveness of 10-year risk thresholds for initiation of statin therapy for primary prevention of cardiovascular disease. *Jama*, 314(2):142–150, 2015.

[14] MA Hernan and J Robins. Causal inference: What if. boca raton: Chapman & hill/crc. 2020.

[15] Fan Li, Laine E Thomas, and Fan Li. Addressing extreme propensity scores via the overlap weights. *American journal of epidemiology*, 188(1):250–257, 2019.

[16] Victor Chernozhukov, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins. Double/debiased machine learning for treatment and causal parameters. *arXiv preprint arXiv:1608.00060*, 2016.

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

[18] Barbra A Dickerman and Miguel A Hernán. Counterfactual prediction is not only for causal inference. *European journal of epidemiology*, 35:615–617, 2020.

[19] Marc Höfler. Causal inference based on counterfactuals. *BMC medical research methodology*, 5:1–12, 2005.

[20] Miguel A Hernán and James M Robins. Using big data to emulate a target trial when a randomized trial is not available. *American journal of epidemiology*, 183(8):758–764, 2016.

[21] Christine M Anderson-Cook. Experimental and quasi-experimental designs for generalized causal inference, 2005.

[22] Zachary K Collier, Walter L Leite, and Allison Karpyn. Neural networks to estimate generalized propensity scores for continuous treatment doses. *Evaluation review*, page 0193841X21992199, 2021.

[23] Fan Li, Kari Lock Morgan, and Alan M Zaslavsky. Balancing covariates via propensity score weighting. *Journal of the American Statistical Association*, 113(521):390–400, 2018.

[24] Heejung Bang and James M Robins. Doubly robust estimation in missing data and causal inference models. *Biometrics*, 61(4):962–973, 2005.

[25] Fabio Pellegrini, Konstantina Chalkou, Carl de Moor, and Shibeshih Belachew. A two-stage prediction model for heterogeneous effects of many treatment options: Application to drugs for multiple sclerosis (1709). *Neurology*, 96(15_supplement):1709, 2021.

[26] Qingyuan Zhao and Trevor Hastie. Causal interpretations of black-box models. *Journal of Business & Economic Statistics*, 39(1):272–281, 2021.

[27] Jianqing Fan and Runze Li. Statistical challenges with high dimensionality: Feature selection in knowledge discovery. *arXiv preprint math/0602133*, 2006.

[28] Thomas D Cook and Donald T Campbell. *Experimental and quasi-experimental designs for generalized causal inference*. Figures, 2007.

[29] Andrew Ying. Causality for complex continuous-time functional longitudinal studies with dynamic treatment regimes. *arXiv preprint arXiv:2406.06868*, 2024.

[30] Kosmas I Paraskevas, Dimitri P Mikhailidis, and Frank J Veith. Optimal statin type and dosage for vascular patients. *Journal of vascular surgery*, 53(3):837–844, 2011.

[31] Yingrong Wang, Haoxuan Li, Minqin Zhu, Anpeng Wu, Ruoxuan Xiong, Fei Wu, and Kun Kuang. Causal inference with complex treatments: A survey. *arXiv preprint arXiv:2407.14022*, 2024.

[32] Annette Kjær Ersbøll and Bjarne Kjær Ersbøll. Epidemiological studies based on small sample sizes–a statistician's point of view. *Acta Veterinaria Scandinavica*, 44(Suppl 1):S127, 2003.

[33] Fan Li. Propensity score weighting for causal inference with multiple treatments. *The Annals of Applied Statistics*, 13(4):2389–2415, 2019.

[34] Miguel A Hernán and James M Robins. Causal inference, 2010.

[35] Sören R Künzel, Jasjeet S Sekhon, Peter J Bickel, and Bin Yu. Metalearners for estimating heterogeneous treatment effects using machine learning. *Proceedings of the national academy of sciences*, 116(10):4156–4165, 2019.

[36] James M Robins and Andrea Rotnitzky. Semiparametric efficiency in multivariate regression models with missing data. *Journal of the American Statistical Association*, 90(429):122–129, 1995.

[37] Kourosh Ravvaz, John A Weissert, Christian T Ruff, Chih-Lin Chi, and Peter J Tonellato. Personalized anticoagulation: optimizing warfarin management using genetics and simulated clinical trials. *Circulation: Cardiovascular Genetics*, 10(6):e001804, 2017.

[38] Donna K Arnett, Roger S Blumenthal, Michelle A Albert, Andrew B Buroker, Zachary D Goldberger, Ellen J Hahn, Cheryl Dennison Himmelfarb, Amit Khera, Donald Lloyd-Jones, J William McEvoy, et al. 2019 acc/aha guideline on the primary prevention of cardiovascular disease: a report of the american college of cardiology/american heart association task force on clinical practice guidelines. *Circulation*, 140(11):e596–e646, 2019.

[39] Ian Goodfellow. Deep learning, 2016.

[40] Andy Field. Discovering statistics using spss (third edit). *London and New York: Sage*, 2009.

[41] Rachel C Nethery, Fabrizia Mealli, and Francesca Dominici. Estimating population average causal effects in the presence of non-overlap: The effect of natural gas compressor station exposure on cancer mortality. *The annals of applied statistics*, 13(2):1242, 2019.

Appendix A

Supplementary Figures and Tables

A.1 Tables

Table A.1: OptumLabs data cohort: Subset of Baseline covariates across different Statin types(each cell is mean(SD) or proportion of each column;Total N is defined in first row)

	Atorvastatin	Lovastatin	Pitavastatin	Pravastatin	Rosuvastatin	Simvastatin	Overall
N	104551	2569	736	15240	36505	16632	176233
Intensity Level High	34567	0	0	0	10281	0	44848
Intensity Level Moderate	69984	0	736	4872	26224	11101	112917
Intensity Level Low	0	2569	0	10368	0	5531	18468
Baseline Characteristics:							
Age at index	59.8 (10.2)	60.3 (9.6)	64.9 (8.0)	61.3 (9.7)	59.9 (10.2)	60.1 (10.0)	60.0 (10.1)
Average LDL value prior index	132.7 (42.4)	128.8 (39.1)	147.4 (44.3)	130.4 (39.1)	144.1 (46.3)	130.2 (41.2)	134.6 (43.1)
Age adjusted Charlson score	4.0 (2.7)	3.9 (2.4)	4.8 (2.8)	4.2 (2.6)	3.7 (2.6)	3.9 (2.6)	4.0 (2.7)
Charlson Quan score	2.4 (2.3)	2.3 (2.0)	2.7 (2.5)	2.5 (2.3)	2.1 (2.2)	2.3 (2.2)	2.4 (2.3)
Charlson Quan updated score	1.3 (1.8)	1.1 (1.5)	1.5 (1.8)	1.3 (1.7)	1.1 (1.6)	1.2 (1.6)	1.3 (1.7)
Elixhauser ahrq score	1.4 (9.0)	-0.2 (7.5)	1.3 (8.1)	0.7 (8.3)	0.5 (7.7)	0.1 (7.8)	1.0 (8.6)
Elixhauser van score	3.7 (7.9)	2.2 (6.3)	3.8 (7.3)	3.1 (7.2)	2.8 (6.8)	2.5 (6.9)	3.3 (7.5)
Smoking never used	7.9%	7.3%	7.3%	8.0%	9.3%	7.5%	8.2%
Smoking previously used	25.8%	19.8%	20.6%	21.8%	20.7%	20.2%	23.8%
Smoking currently using	1.7%	1.3%	<1.2%	1.3%	1.4%	1.5%	1.6%
Medication at index:							
Rx at index lipid regulating	3.6%	4.0%	16.0%	5.3%	5.7%	4.0%	4.3%
Rx at index immunosuppressant	1.5%	1.6%	2.0%	2.2%	1.7%	1.3%	1.6%
Rx at index blood pressure med	63.9%	64.0%	68.8%	66.1%	57.1%	61.2%	62.5%
Medications prior index date:							
Rx prior year lipid regulating	5.7%	6.8%	27.8%	8.4%	8.7%	6.2%	6.7%
Rx prior year immunosuppressant	2.8%	2.9%	4.4%	3.7%	3.1%	2.4%	2.9%
Rx prior year blood pressure med	70.3%	71.5%	75.1%	73.5%	64.2%	68.3%	69.2%
Race:							
Asian	3.7%	3.0%	2.6%	2.9%	3.5%	4.4%	3.7%
Black	13.2%	14.2%	12.2%	15.8%	11.0%	13.9%	13.0%
Hispanic	13.7%	18.0%	9.0%	12.4%	12.5%	18.2%	13.8%
White	50.0%	43.5%	55.3%	49.0%	54.5%	43.3%	50.1%
Gender:							
Male	47.5%	54.9%	63.5%	57.0%	50.6%	51.9%	49.6%
Gender unknown	1.3%	1.2%	<1.2%	0.9%	1.7%	1.3%	1.3%

Geographical location:							
Midwest	20.4%	16.7%	12.1%	14.9%	12.1%	14.9%	17.6%
Northeast	11.2%	4.7%	10.5%	7.6%	13.3%	11.4%	11.2%
Other	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
South	53.7%	65.2%	67.0%	67.5%	61.8%	60.2%	57.4%
West	13.2%	11.8%	9.5%	8.9%	10.9%	11.9%	12.2%
Income status:							
Level 1.0	7.6%	9.1%	9.8%	8.9%	5.6%	8.9%	7.4%
Level 2.0	3.3%	4.0%	3.2%	3.8%	2.0%	3.8%	3.2%
Level 3.0	0.9%	0.8%	<1.2%	0.8%	0.4%	0.9%	0.8%
Level 4.0	0.6%	0.9%	<1.2%	0.8%	0.5%	0.7%	0.6%
Level unknown	56.1%	53.8%	34.3%	48.4%	58.8%	53.8%	55.6%

Table A.2: AHC-IE data cohort: Baseline covariates across different Statin types

Grouped by statin_generic_name							
	Label	Overall	atorvastatin	lovastatin	pravastatin	rosuvastatin	simvastatin
n		95555	40951	2342	8217	6843	37202
age_at_index, mean (SD)		59.1 (13.1)	59.1 (12.9)	61.3 (13.1)	60.2 (13.0)	61.6 (12.4)	58.3 (13.4)
Gender							
Female, n (%)	1	45095 (47.2)	18090 (44.2)	1299 (55.5)	4319 (52.6)	3308 (48.3)	18079 (48.6)
Male, n (%)	1	50452 (52.8)	22858 (55.8)	1043 (44.5)	3896 (47.4)	3533 (51.6)	19122 (51.4)
Unknown, n (%)	1	8 (0.0)	3 (0.0)		2 (0.0)	2 (0.0)	1 (0.0)
Race							
white, n (%)	1	82635 (86.5)	34911 (85.3)	2111 (90.1)	7140 (86.9)	6055 (88.5)	32418 (87.1)
black, n (%)	1	5051 (5.3)	2387 (5.8)	85 (3.6)	398 (4.8)	306 (4.5)	1875 (5.0)
asian, n (%)	1	4214 (4.4)	2010 (4.9)	70 (3.0)	376 (4.6)	238 (3.5)	1520 (4.1)
indian_alaska_native, n (%)	1	652 (0.7)	299 (0.7)	14 (0.6)	56 (0.7)	46 (0.7)	237 (0.6)
native_hawaiian_pacific, n (%)	1	151 (0.2)	81 (0.2)	3 (0.1)	12 (0.1)	10 (0.1)	45 (0.1)
hispanic_latino, n (%)	1	1125 (1.2)	511 (1.2)	19 (0.8)	102 (1.2)	70 (1.0)	423 (1.1)
Statin Intensity							
High, n (%)	1	14855 (15.5)	12003 (29.3)			2852 (41.7)	
Moderate, n (%)	1	66208 (69.3)	28948 (70.7)	754 (32.2)	2927 (35.6)	3991 (58.3)	29588 (79.5)

Low, n (%)	1	14492 (15.2)		1588 (67.8)	5290 (64.4)		7614 (20.5)
statin_1y_discontinued, n (%)	1	36421 (38.1)	16166 (39.5)	846 (36.1)	3322 (40.4)	3188 (46.6)	12899 (34.7)
sas_flag_1y_postindex, n (%)	1	30596 (32.0)	13343 (32.6)	764 (32.6)	2779 (33.8)	2297 (33.6)	11413 (30.7)
ldl_value_avg_1y_preindex, mean (SD)		127.8 (44.5)	125.6 (45.2)	129.1 (40.1)	130.2 (40.4)	121.7 (50.3)	130.7 (43.5)
ldl_value_avg_1y_postindex, mean (SD)		94.7 (34.3)	88.1 (34.3)	106.9 (29.4)	107.0 (32.8)	88.8 (37.1)	98.8 (32.8)
any_services_count_1y_preindex, mean (SD)		18.4 (20.8)	19.1 (21.9)	14.4 (14.4)	20.8 (25.4)	22.0 (23.8)	16.7 (17.8)
any_services_recent_days_preindex, mean (SD)		-0.8 (8.0)	-1.0 (9.0)	-0.9 (8.0)	-0.8 (8.3)	-1.0 (7.5)	-0.7 (6.9)
hospital_at_index, n (%)	1	5757 (6.0)	3184 (7.8)	25 (1.1)	369 (4.5)	527 (7.7)	1652 (4.4)
hospital_count_30d_preindex, n (%)	1	1514 (1.6)	862 (2.1)	6 (0.3)	110 (1.3)	115 (1.7)	421 (1.1)
outpatient_at_index, n (%)	1	77565 (81.2)	32376 (79.1)	1946 (83.1)	6834 (83.2)	5491 (80.2)	30918 (83.1)
outpatient_count_30d_preindex, mean (SD)		3.0 (4.3)	3.0 (4.2)	2.6 (2.4)	3.3 (8.6)	3.2 (3.3)	2.9 (3.0)
outpatient_count_0_1y_preindex, mean (SD)		17.9 (20.2)	18.5 (21.2)	14.3 (14.2)	20.5 (24.8)	21.4 (23.2)	16.4 (17.4)
outpatient_count_1_2y_preindex, mean (SD)		11.0 (17.1)	11.6 (18.1)	8.0 (12.4)	13.1 (21.4)	13.1 (19.8)	9.5 (14.2)
outpatient_count_2_3y_preindex, mean (SD)		8.7 (15.0)	9.4 (16.3)	6.2 (11.5)	10.5 (17.8)	10.3 (17.2)	7.4 (12.3)
outpatient_recent_days_preindex, mean (SD)		-0.8 (8.3)	-1.0 (9.4)	-1.0 (9.2)	-0.8 (8.3)	-1.0 (7.6)	-0.7 (7.1)
weight_preindex, mean (SD)		196.4 (49.6)	197.7 (50.3)	191.9 (47.7)	192.9 (49.5)	195.7 (47.9)	196.3 (49.2)
bmi_days_preindex, mean (SD)		-48.6 (197.9)	-53.5 (222.8)	-45.6 (201.6)	-39.0 (153.2)	-59.3 (232.7)	-43.5 (167.9)
bmi_collected_preindex, n (%)	1	95555 (100.0)	40951 (100.0)	2342 (100.0)	8217 (100.0)	6843 (100.0)	37202 (100.0)
bmi_preindex, mean (SD)		30.7 (6.8)	30.8 (6.8)	30.6 (6.7)	30.5 (6.8)	30.7 (6.4)	30.8 (6.8)
bp_days_preindex, mean (SD)		-36.9 (190.4)	-42.9 (216.6)	-25.7 (160.6)	-29.8 (141.2)	-51.7 (237.1)	-30.0 (158.1)
bp_collected_preindex, n (%)	1	95555 (100.0)	40951 (100.0)	2342 (100.0)	8217 (100.0)	6843 (100.0)	37202 (100.0)
bp_treated_preindex, n (%)	1	51578 (54.0)	22400 (54.7)	1301 (55.6)	4599 (56.0)	3744 (54.7)	19534 (52.5)
bp_systolic_preindex, mean (SD)		127.8 (16.5)	128.9 (16.8)	126.7 (15.9)	127.5 (16.3)	127.8 (16.8)	126.7 (16.2)
bp_diastolic_preindex, mean (SD)		76.8 (10.6)	77.7 (10.7)	75.0 (9.9)	76.1 (10.2)	76.0 (10.5)	76.3 (10.4)
alcohol_use_preindex, n (%)	1	42789 (44.8)	18106 (44.2)	621 (26.5)	3909 (47.6)	3265 (47.7)	16888 (45.4)

drug_use_preindex, n (%)	1	1428 (1.5)	790 (1.9)	12 (0.5)	94 (1.1)	101 (1.5)	431 (1.2)
tobacco_use_preindex, n (%)	1	24922 (26.1)	10969 (26.8)	359 (15.3)	2253 (27.4)	1808 (26.4)	9533 (25.6)
tobacco_previous_use_preindex, n (%)	1	13835 (14.5)	5466 (13.3)	205 (8.8)	1328 (16.2)	1155 (16.9)	5681 (15.3)
smokeless_tobacco_use_preindex, n (%)	1	1270 (1.3)	649 (1.6)	15 (0.6)	99 (1.2)	83 (1.2)	424 (1.1)
smokeless_tobacco_previous_use_preindex, n (%)	1	2416 (2.5)	1260 (3.1)	19 (0.8)	212 (2.6)	187 (2.7)	738 (2.0)
chew_tobacco_use_preindex, n (%)	1	1403 (1.5)	707 (1.7)	13 (0.6)	106 (1.3)	95 (1.4)	482 (1.3)
snuff_use_preindex, n (%)	1	198 (0.2)	100 (0.2)	2 (0.1)	20 (0.2)	13 (0.2)	63 (0.2)
smoking_tobacco_use_preindex, n (%)	1	11780 (12.3)	5114 (12.5)	168 (7.2)	966 (11.8)	727 (10.6)	4805 (12.9)
smoking_tobacco_previous_use_preindex, n (%)	1	25766 (27.0)	10757 (26.3)	385 (16.4)	2528 (30.8)	2145 (31.3)	9951 (26.7)
cigars_use_preindex, n (%)	1	1264 (1.3)	543 (1.3)	20 (0.9)	104 (1.3)	97 (1.4)	500 (1.3)
pipes_use_preindex, n (%)	1	399 (0.4)	162 (0.4)	4 (0.2)	37 (0.5)	28 (0.4)	168 (0.5)
cigarettes_use_preindex, n (%)	1	19112 (20.0)	8358 (20.4)	274 (11.7)	1820 (22.1)	1410 (20.6)	7250 (19.5)
any_med_count_preindex, mean (SD)		30.6 (59.1)	34.7 (68.4)	18.0 (30.1)	36.2 (78.1)	33.7 (64.6)	25.1 (39.6)
immunosuppressant_med_use_preindex, n (%)	1	2595 (2.7)	1189 (2.9)	32 (1.4)	443 (5.4)	238 (3.5)	693 (1.9)
immunosuppressant_med_use_6mo_preindex, n (%)	1	2372 (2.5)	1076 (2.6)	28 (1.2)	409 (5.0)	213 (3.1)	646 (1.7)
immunosuppressant_med_use_days_preindex, mean (SD)		25.4 (228.8)	25.9 (198.2)	13.8 (142.4)	51.5 (266.3)	29.7 (198.8)	19.1 (258.4)
other_lipid_regulating_med_use_preindex, n (%)	1	6982 (7.3)	2245 (5.5)	238 (10.2)	764 (9.3)	781 (11.4)	2954 (7.9)
other_lipid_regulating_med_use_6mo_preindex, n (%)	1	6503 (6.8)	2062 (5.0)	228 (9.7)	709 (8.6)	718 (10.5)	2786 (7.5)
other_lipid_regulating_med_use_days_preindex, mean (SD)		56.8 (265.7)	40.8 (225.1)	85.5 (328.7)	76.6 (309.2)	80.1 (298.6)	64.0 (284.0)
ezetimibe_med_use_preindex, n (%)	1	1972 (2.1)	360 (0.9)	41 (1.8)	179 (2.2)	256 (3.7)	1136 (3.1)
ezetimibe_med_use_6mo_preindex, n (%)	1	1895 (2.0)	346 (0.8)	39 (1.7)	169 (2.1)	237 (3.5)	1104 (3.0)
ezetimibe_med_use_days_preindex, mean (SD)		14.2 (127.5)	5.8 (80.1)	10.5 (107.6)	16.2 (140.1)	23.5 (154.9)	21.6 (158.2)
pcsk9_med_use_preindex, n (%)	1	29 (0.0)	10 (0.0)	1 (0.0)	4 (0.0)	14 (0.2)	

pcsk9_med_use_6mo_preindex, n (%)	1	24 (0.0)	9 (0.0)	1 (0.0)	2 (0.0)	12 (0.2)	
pcsk9_med_use_days_preindex, mean (SD)		0.1 (8.3)	0.1 (8.9)	0.2 (11.5)	0.1 (6.5)	0.6 (19.8)	0.0 (0.0)
fibrate_med_use_preindex, n (%)	1	3502 (3.7)	1294 (3.2)	98 (4.2)	402 (4.9)	386 (5.6)	1322 (3.6)
fibrate_med_use_6mo_preindex, n (%)	1	3299 (3.5)	1221 (3.0)	94 (4.0)	379 (4.6)	359 (5.2)	1246 (3.3)
fibrate_med_use_days_preindex, mean (SD)		30.1 (191.3)	24.4 (167.1)	38.8 (226.7)	43.2 (233.2)	42.6 (221.9)	30.7 (197.0)
niacin_med_use_preindex, n (%)	1	1735 (1.8)	542 (1.3)	116 (5.0)	232 (2.8)	182 (2.7)	663 (1.8)
niacin_med_use_6mo_preindex, n (%)	1	1495 (1.6)	461 (1.1)	113 (4.8)	194 (2.4)	158 (2.3)	569 (1.5)
niacin_med_use_days_preindex, mean (SD)		12.0 (131.4)	9.1 (122.3)	35.0 (207.9)	18.0 (149.7)	15.6 (143.5)	11.8 (127.9)
nitrate_med_use_preindex, n (%)	1	1346 (1.4)	643 (1.6)	20 (0.9)	124 (1.5)	194 (2.8)	365 (1.0)
nitrate_med_use_6mo_preindex, n (%)	1	1267 (1.3)	612 (1.5)	20 (0.9)	106 (1.3)	183 (2.7)	346 (0.9)
nitrate_med_use_days_preindex, mean (SD)		7.7 (97.5)	8.0 (101.1)	6.9 (87.4)	7.5 (89.8)	14.7 (127.2)	6.3 (89.0)
blood_pressure_med_use_preindex, n (%)	1	59238 (62.0)	26019 (63.5)	1445 (61.7)	5205 (63.3)	4411 (64.5)	22158 (59.6)
blood_pressure_med_use_6mo_preindex, n (%)	1	58321 (61.0)	25614 (62.5)	1437 (61.4)	5107 (62.2)	4335 (63.3)	21828 (58.7)
blood_pressure_med_use_days_preindex, mean (SD)		710.9 (808.4)	696.5 (775.3)	787.0 (808.7)	758.0 (784.2)	675.3 (718.6)	718.2 (862.4)
ace_inhibitor_med_use_preindex, n (%)	1	31223 (32.7)	13342 (32.6)	792 (33.8)	2752 (33.5)	1973 (28.8)	12364 (33.2)
ace_inhibitor_med_use_6mo_preindex, n (%)	1	29548 (30.9)	12602 (30.8)	762 (32.5)	2558 (31.1)	1865 (27.3)	11761 (31.6)
ace_inhibitor_med_use_days_preindex, mean (SD)		300.4 (578.4)	283.2 (542.2)	345.8 (612.3)	315.6 (586.5)	242.5 (498.4)	323.7 (623.6)
beta_blocker_med_use_preindex, n (%)	1	24978 (26.1)	10962 (26.8)	611 (26.1)	2100 (25.6)	2048 (29.9)	9257 (24.9)
beta_blocker_med_use_6mo_preindex, n (%)	1	23515 (24.6)	10339 (25.2)	589 (25.1)	1933 (23.5)	1932 (28.2)	8722 (23.4)
beta_blocker_med_use_days_preindex, mean (SD)		227.8 (525.5)	218.5 (523.5)	276.7 (557.5)	233.5 (523.8)	229.5 (494.8)	233.3 (531.0)

calcium_channel_blocker_med_use_preindex, n (%)	1	14208 (14.9)	6777 (16.5)	306 (13.1)	1435 (17.5)	1192 (17.4)	4498 (12.1)
calcium_channel_blocker_med_use_6mo_preindex, n (%)	1	13225 (13.8)	6308 (15.4)	287 (12.3)	1324 (16.1)	1107 (16.2)	4199 (11.3)
calcium_channel_blocker_med_use_days_preindex, mean (SD)		124.7 (386.6)	131.1 (390.0)	126.8 (388.8)	153.1 (429.8)	136.4 (394.2)	109.1 (370.4)
diuretic_med_use_preindex, n (%)	1	27401 (28.7)	11473 (28.0)	776 (33.1)	2583 (31.4)	1963 (28.7)	10606 (28.5)
diuretic_med_use_6mo_preindex, n (%)	1	25883 (27.1)	10753 (26.3)	753 (32.2)	2435 (29.6)	1856 (27.1)	10086 (27.1)
diuretic_med_use_days_preindex, mean (SD)		273.2 (583.4)	250.5 (531.5)	358.8 (635.6)	303.0 (578.7)	245.7 (505.1)	291.3 (644.5)
colchicine_med_use_days_preindex, mean (SD)		4.3 (72.6)	4.8 (74.3)	1.5 (39.1)	4.3 (75.0)	5.0 (70.0)	3.8 (72.3)
amiodarone_med_use_days_preindex, mean (SD)		2.2 (48.2)	2.1 (48.2)	1.8 (50.7)	2.1 (41.3)	3.7 (54.4)	1.9 (48.2)
diltiazem_med_use_days_preindex, mean (SD)		15.1 (140.6)	14.5 (136.8)	17.6 (149.9)	24.0 (178.6)	15.5 (138.6)	13.7 (134.7)
verapamil_med_use_days_preindex, mean (SD)		6.0 (88.9)	5.1 (79.3)	7.5 (89.5)	9.3 (111.6)	8.1 (107.2)	5.7 (89.5)
amlodipine_med_use_days_preindex, mean (SD)		96.2 (340.0)	104.9 (348.5)	88.7 (325.7)	113.5 (373.0)	105.5 (348.1)	81.5 (321.2)
antifungal_med_use_days_preindex, mean (SD)		14.9 (121.6)	14.2 (121.3)	19.1 (145.0)	22.0 (149.7)	16.9 (129.2)	13.5 (111.4)
CYP2C9_inhibitors_med_use_days_preindex, mean (SD)		38.9 (204.8)	38.6 (200.4)	43.9 (232.2)	53.2 (252.2)	40.1 (194.1)	35.6 (197.8)
CYP2C9_inducers_med_use_days_preindex, mean (SD)		3.7 (137.1)	3.1 (62.0)	7.6 (139.1)	3.5 (77.9)	8.6 (445.7)	3.4 (70.1)
CYP3A4_inhibitors_med_use_days_preindex, mean (SD)		216.9 (520.4)	223.8 (525.3)	204.2 (477.0)	265.9 (550.5)	228.1 (498.2)	197.1 (513.7)
CYP3A4_inducers_med_use_days_preindex, mean (SD)		16.5 (187.1)	14.9 (138.1)	25.8 (201.8)	16.2 (150.4)	22.6 (465.1)	16.7 (146.8)
pgp_inhibitors_med_use_days_preindex, mean (SD)		237.2 (538.0)	224.4 (503.7)	253.3 (921.6)	293.8 (585.5)	241.8 (518.1)	236.9 (533.7)
pgp_inducers_med_use_days_preindex, mean (SD)		3.0 (132.4)	2.5 (55.9)	5.2 (100.8)	1.7 (45.7)	8.1 (444.9)	2.8 (64.1)
OATP1B1_inhibitors_med_use_days_preindex, mean (SD)		58.6 (305.9)	56.1 (275.1)	74.1 (336.7)	71.1 (314.3)	61.3 (520.5)	57.2 (279.7)
OATP1B3_inhibitors_med_use_days_preindex, mean (SD)		6.4 (92.7)	6.4 (96.7)	4.0 (63.8)	12.7 (127.1)	8.9 (103.3)	4.7 (77.4)
aids_flag_preindex, n (%)	1	144 (0.2)	81 (0.2)		27 (0.3)	17 (0.2)	19 (0.1)
aids_flag_1y_preindex, n (%)	1	135 (0.1)	78 (0.2)		27 (0.3)	14 (0.2)	16 (0.0)
afib_flag_preindex, n (%)	1	6143 (6.4)	2904 (7.1)	97 (4.1)	621 (7.6)	598 (8.7)	1923 (5.2)
afib_flag_1y_preindex, n (%)	1	5317 (5.6)	2500 (6.1)	82 (3.5)	529 (6.4)	512 (7.5)	1694 (4.6)
aki_flag_preindex, n (%)	1	3443 (3.6)	1875 (4.6)	38 (1.6)	383 (4.7)	318 (4.6)	829 (2.2)

aki_flag_1y_preindex, n (%)	1	2043 (2.1)	1099 (2.7)	25 (1.1)	218 (2.7)	185 (2.7)	516 (1.4)
aud_flag_preindex, n (%)	1	12988 (13.6)	5506 (13.4)	211 (9.0)	1156 (14.1)	724 (10.6)	5391 (14.5)
aud_flag_1y_preindex, n (%)	1	7829 (8.2)	2928 (7.2)	158 (6.7)	724 (8.8)	443 (6.5)	3576 (9.6)
ane_flag_preindex, n (%)	1	4620 (4.8)	2076 (5.1)	122 (5.2)	521 (6.3)	378 (5.5)	1523 (4.1)
ane_flag_1y_preindex, n (%)	1	2928 (3.1)	1285 (3.1)	92 (3.9)	298 (3.6)	239 (3.5)	1014 (2.7)
ang_flag_preindex, n (%)	1	3123 (3.3)	1751 (4.3)	28 (1.2)	226 (2.8)	453 (6.6)	665 (1.8)
ang_flag_1y_preindex, n (%)	1	2381 (2.5)	1400 (3.4)	22 (0.9)	147 (1.8)	337 (4.9)	475 (1.3)
ang_cad_inp_flag_preindex, n (%)	1	265 (0.3)	174 (0.4)	1 (0.0)	35 (0.4)	35 (0.5)	20 (0.1)
ang_cad_inp_flag_1y_preindex, n (%)	1	249 (0.3)	168 (0.4)	1 (0.0)	29 (0.4)	32 (0.5)	19 (0.1)
arr_flag_preindex, n (%)	1	20412 (21.4)	9464 (23.1)	330 (14.1)	1940 (23.6)	1848 (27.0)	6830 (18.4)
arr_flag_1y_preindex, n (%)	1	14090 (14.7)	6680 (16.3)	199 (8.5)	1297 (15.8)	1322 (19.3)	4592 (12.3)
ashd_flag_preindex, n (%)	1	15143 (15.8)	7473 (18.2)	218 (9.3)	1190 (14.5)	2060 (30.1)	4202 (11.3)
ashd_flag_1y_preindex, n (%)	1	13438 (14.1)	6653 (16.2)	193 (8.2)	1010 (12.3)	1871 (27.3)	3711 (10.0)
ashd_ap_flag_preindex, n (%)	1	4871 (5.1)	2332 (5.7)	58 (2.5)	400 (4.9)	751 (11.0)	1330 (3.6)
ashd_ap_flag_1y_preindex, n (%)	1	3530 (3.7)	1759 (4.3)	42 (1.8)	266 (3.2)	491 (7.2)	972 (2.6)
cabg_flag_preindex, n (%)	1	250 (0.3)	101 (0.2)	2 (0.1)	17 (0.2)	33 (0.5)	97 (0.3)
cabg_flag_1y_preindex, n (%)	1	194 (0.2)	81 (0.2)	2 (0.1)	11 (0.1)	20 (0.3)	80 (0.2)
cabg_inp_flag_preindex, n (%)	1	53 (0.1)	31 (0.1)	2 (0.1)	2 (0.0)	6 (0.1)	12 (0.0)
cabg_inp_flag_1y_preindex, n (%)	1	48 (0.1)	28 (0.1)	2 (0.1)	2 (0.0)	4 (0.1)	12 (0.0)
cad_flag_preindex, n (%)	1	15818 (16.6)	7862 (19.2)	222 (9.5)	1227 (14.9)	2117 (30.9)	4390 (11.8)
cad_flag_1y_preindex, n (%)	1	14124 (14.8)	7042 (17.2)	194 (8.3)	1044 (12.7)	1936 (28.3)	3908 (10.5)
cancer_flag_preindex, n (%)	1	7315 (7.7)	3228 (7.9)	137 (5.8)	699 (8.5)	658 (9.6)	2593 (7.0)
cancer_flag_1y_preindex, n (%)	1	5489 (5.7)	2412 (5.9)	109 (4.7)	532 (6.5)	510 (7.5)	1926 (5.2)
cere_flag_preindex, n (%)	1	9284 (9.7)	4625 (11.3)	134 (5.7)	725 (8.8)	811 (11.9)	2989 (8.0)
cere_flag_1y_preindex, n (%)	1	7548 (7.9)	3851 (9.4)	106 (4.5)	535 (6.5)	624 (9.1)	2432 (6.5)
chf_flag_preindex, n (%)	1	6208 (6.5)	2993 (7.3)	89 (3.8)	605 (7.4)	679 (9.9)	1842 (5.0)
chf_flag_1y_preindex, n (%)	1	5242 (5.5)	2576 (6.3)	67 (2.9)	489 (6.0)	572 (8.4)	1538 (4.1)
ckd_flag_preindex, n (%)	1	1603 (1.7)	943 (2.3)	7 (0.3)	174 (2.1)	168 (2.5)	311 (0.8)
ckd_flag_1y_preindex, n (%)	1	1188 (1.2)	682 (1.7)	7 (0.3)	130 (1.6)	123 (1.8)	246 (0.7)
coag_flag_preindex, n (%)	1	3316 (3.5)	1645 (4.0)	40 (1.7)	357 (4.3)	296 (4.3)	978 (2.6)
coag_flag_1y_preindex, n (%)	1	2151 (2.3)	1055 (2.6)	26 (1.1)	225 (2.7)	190 (2.8)	655 (1.8)
cpd_flag_preindex, n (%)	1	16500 (17.3)	7153 (17.5)	323 (13.8)	1706 (20.8)	1310 (19.1)	6008 (16.1)
cpd_flag_1y_preindex, n (%)	1	12113 (12.7)	5141 (12.6)	250 (10.7)	1233 (15.0)	978 (14.3)	4511 (12.1)

cv_death_flag_preindex, n (%)	1	292 (0.3)	169 (0.4)		29 (0.4)	38 (0.6)	56 (0.2)
cv_death_flag_1y_preindex, n (%)	1	215 (0.2)	129 (0.3)		19 (0.2)	25 (0.4)	42 (0.1)
cva_flag_preindex, n (%)	1	501 (0.5)	335 (0.8)	7 (0.3)	28 (0.3)	47 (0.7)	84 (0.2)
cva_flag_1y_preindex, n (%)	1	411 (0.4)	272 (0.7)	5 (0.2)	22 (0.3)	38 (0.6)	74 (0.2)
cva_inp_flag_preindex, n (%)	1	185 (0.2)	132 (0.3)	1 (0.0)	10 (0.1)	18 (0.3)	24 (0.1)
cva_inp_flag_1y_preindex, n (%)	1	160 (0.2)	111 (0.3)	1 (0.0)	10 (0.1)	14 (0.2)	24 (0.1)
cva_init_flag_preindex, n (%)	1	4898 (5.1)	2647 (6.5)	65 (2.8)	314 (3.8)	325 (4.7)	1547 (4.2)
cva_init_flag_1y_preindex, n (%)	1	4193 (4.4)	2321 (5.7)	44 (1.9)	240 (2.9)	258 (3.8)	1330 (3.6)
cva_init_inp_flag_preindex, n (%)	1	1586 (1.7)	1018 (2.5)	3 (0.1)	76 (0.9)	83 (1.2)	406 (1.1)
cva_init_inp_flag_1y_preindex, n (%)	1	1507 (1.6)	966 (2.4)	2 (0.1)	67 (0.8)	78 (1.1)	394 (1.1)
dementia_flag_preindex, n (%)	1	995 (1.0)	393 (1.0)	24 (1.0)	109 (1.3)	74 (1.1)	395 (1.1)
dementia_flag_1y_preindex, n (%)	1	804 (0.8)	314 (0.8)	21 (0.9)	90 (1.1)	56 (0.8)	323 (0.9)
dep_flag_preindex, n (%)	1	21916 (22.9)	9503 (23.2)	485 (20.7)	1975 (24.0)	1641 (24.0)	8312 (22.3)
dep_flag_1y_preindex, n (%)	1	17274 (18.1)	7334 (17.9)	389 (16.6)	1567 (19.1)	1287 (18.8)	6697 (18.0)
dm_flag_preindex, n (%)	1	28281 (29.6)	12169 (29.7)	750 (32.0)	2614 (31.8)	2059 (30.1)	10689 (28.7)
dm_flag_1y_preindex, n (%)	1	27318 (28.6)	11726 (28.6)	733 (31.3)	2514 (30.6)	1966 (28.7)	10379 (27.9)
dm1_flag_preindex, n (%)	1	2659 (2.8)	1136 (2.8)	62 (2.6)	273 (3.3)	192 (2.8)	996 (2.7)
dm1_flag_1y_preindex, n (%)	1	2064 (2.2)	897 (2.2)	48 (2.0)	203 (2.5)	138 (2.0)	778 (2.1)
dm2_flag_preindex, n (%)	1	27782 (29.1)	11976 (29.2)	737 (31.5)	2556 (31.1)	2020 (29.5)	10493 (28.2)
dm2_flag_1y_preindex, n (%)	1	26508 (27.7)	11379 (27.8)	714 (30.5)	2427 (29.5)	1907 (27.9)	10081 (27.1)
dm_c_flag_preindex, n (%)	1	11328 (11.9)	5659 (13.8)	190 (8.1)	1101 (13.4)	961 (14.0)	3417 (9.2)
dm_c_flag_1y_preindex, n (%)	1	9897 (10.4)	5037 (12.3)	157 (6.7)	956 (11.6)	843 (12.3)	2904 (7.8)
dm_uc_flag_preindex, n (%)	1	26811 (28.1)	11303 (27.6)	731 (31.2)	2510 (30.5)	1933 (28.2)	10334 (27.8)
dm_uc_flag_1y_preindex, n (%)	1	25374 (26.6)	10572 (25.8)	708 (30.2)	2371 (28.9)	1778 (26.0)	9945 (26.7)
dm_cc_flag_preindex, n (%)	1	9969 (10.4)	5038 (12.3)	165 (7.0)	966 (11.8)	866 (12.7)	2934 (7.9)
dm_cc_flag_1y_preindex, n (%)	1	8727 (9.1)	4476 (10.9)	131 (5.6)	840 (10.2)	762 (11.1)	2518 (6.8)
dm_ucc_flag_preindex, n (%)	1	27762 (29.1)	11909 (29.1)	740 (31.6)	2563 (31.2)	2006 (29.3)	10544 (28.3)
dm_ucc_flag_1y_preindex, n (%)	1	26748 (28.0)	11435 (27.9)	724 (30.9)	2457 (29.9)	1899 (27.8)	10233 (27.5)
dud_flag_preindex, n (%)	1	2644 (2.8)	1406 (3.4)	18 (0.8)	229 (2.8)	210 (3.1)	781 (2.1)
dud_flag_1y_preindex, n (%)	1	1558 (1.6)	793 (1.9)	12 (0.5)	133 (1.6)	117 (1.7)	503 (1.4)
fluid_flag_preindex, n (%)	1	10269 (10.7)	5072 (12.4)	144 (6.1)	1032 (12.6)	852 (12.5)	3169 (8.5)
fluid_flag_1y_preindex, n (%)	1	6275 (6.6)	3057 (7.5)	93 (4.0)	612 (7.4)	543 (7.9)	1970 (5.3)
hcabg_flag_preindex, n (%)	1	1689 (1.8)	830 (2.0)	21 (0.9)	112 (1.4)	287 (4.2)	439 (1.2)

hemiplegia_flag_preindex, n (%)	1	768 (0.8)	405 (1.0)	9 (0.4)	58 (0.7)	50 (0.7)	246 (0.7)
hemiplegia_flag_1y_preindex, n (%)	1	565 (0.6)	291 (0.7)	7 (0.3)	42 (0.5)	36 (0.5)	189 (0.5)
hmi_flag_preindex, n (%)	1	1980 (2.1)	978 (2.4)	20 (0.9)	157 (1.9)	322 (4.7)	503 (1.4)
hpci_flag_preindex, n (%)	1	2665 (2.8)	1443 (3.5)	19 (0.8)	190 (2.3)	484 (7.1)	529 (1.4)
htn_c_flag_preindex, n (%)	1	5796 (6.1)	2801 (6.8)	58 (2.5)	707 (8.6)	566 (8.3)	1664 (4.5)
htn_c_flag_1y_preindex, n (%)	1	4760 (5.0)	2287 (5.6)	49 (2.1)	593 (7.2)	437 (6.4)	1394 (3.7)
htn_uc_flag_preindex, n (%)	1	57971 (60.7)	25012 (61.1)	1483 (63.3)	5232 (63.7)	4467 (65.3)	21777 (58.5)
htn_uc_flag_1y_preindex, n (%)	1	55243 (57.8)	23747 (58.0)	1439 (61.4)	4983 (60.6)	4214 (61.6)	20860 (56.1)
hypo_flag_preindex, n (%)	1	12644 (13.2)	5082 (12.4)	365 (15.6)	1249 (15.2)	1035 (15.1)	4913 (13.2)
hypo_flag_1y_preindex, n (%)	1	11609 (12.1)	4608 (11.3)	344 (14.7)	1143 (13.9)	937 (13.7)	4577 (12.3)
hypo_bg_flag_preindex, n (%)	1	12325 (12.9)	4970 (12.1)	350 (14.9)	1228 (14.9)	1001 (14.6)	4776 (12.8)
hypo_bg_flag_1y_preindex, n (%)	1	11210 (11.7)	4425 (10.8)	336 (14.3)	1104 (13.4)	905 (13.2)	4440 (11.9)
ld_flag_preindex, n (%)	1	2885 (3.0)	1385 (3.4)	40 (1.7)	316 (3.8)	255 (3.7)	889 (2.4)
ld_flag_1y_preindex, n (%)	1	1135 (1.2)	480 (1.2)	18 (0.8)	133 (1.6)	100 (1.5)	404 (1.1)
ld_mild_flag_preindex, n (%)	1	4994 (5.2)	2438 (6.0)	60 (2.6)	540 (6.6)	415 (6.1)	1541 (4.1)
ld_mild_flag_1y_preindex, n (%)	1	2975 (3.1)	1359 (3.3)	40 (1.7)	339 (4.1)	256 (3.7)	981 (2.6)
ld_ms_flag_preindex, n (%)	1	416 (0.4)	204 (0.5)	3 (0.1)	64 (0.8)	32 (0.5)	113 (0.3)
ld_ms_flag_1y_preindex, n (%)	1	249 (0.3)	109 (0.3)	3 (0.1)	38 (0.5)	18 (0.3)	81 (0.2)
lymphoma_flag_preindex, n (%)	1	751 (0.8)	327 (0.8)	9 (0.4)	100 (1.2)	63 (0.9)	252 (0.7)
lymphoma_flag_1y_preindex, n (%)	1	546 (0.6)	239 (0.6)	8 (0.3)	75 (0.9)	49 (0.7)	175 (0.5)
met_flag_preindex, n (%)	1	986 (1.0)	494 (1.2)	15 (0.6)	88 (1.1)	91 (1.3)	298 (0.8)
met_flag_1y_preindex, n (%)	1	653 (0.7)	307 (0.7)	9 (0.4)	57 (0.7)	70 (1.0)	210 (0.6)
mi_flag_preindex, n (%)	1	1026 (1.1)	497 (1.2)	10 (0.4)	83 (1.0)	95 (1.4)	341 (0.9)
mi_flag_1y_preindex, n (%)	1	737 (0.8)	359 (0.9)	6 (0.3)	47 (0.6)	61 (0.9)	264 (0.7)
mi_inp_flag_preindex, n (%)	1	268 (0.3)	161 (0.4)	1 (0.0)	21 (0.3)	17 (0.2)	68 (0.2)
mi_inp_flag_1y_preindex, n (%)	1	238 (0.2)	142 (0.3)	1 (0.0)	15 (0.2)	16 (0.2)	64 (0.2)
mi_init_flag_preindex, n (%)	1	3495 (3.7)	2154 (5.3)	18 (0.8)	194 (2.4)	405 (5.9)	724 (1.9)
mi_init_flag_1y_preindex, n (%)	1	2903 (3.0)	1835 (4.5)	9 (0.4)	152 (1.8)	303 (4.4)	604 (1.6)
mi_init_inp_flag_preindex, n (%)	1	1097 (1.1)	745 (1.8)	3 (0.1)	61 (0.7)	137 (2.0)	151 (0.4)
mi_init_inp_flag_1y_preindex, n (%)	1	1022 (1.1)	698 (1.7)	3 (0.1)	55 (0.7)	124 (1.8)	142 (0.4)
mus_oth_flag_preindex, n (%)	1	11148 (11.7)	5350 (13.1)	156 (6.7)	1067 (13.0)	884 (12.9)	3691 (9.9)
mus_oth_flag_1y_preindex, n (%)	1	4140 (4.3)	1915 (4.7)	70 (3.0)	403 (4.9)	326 (4.8)	1426 (3.8)
myal_flag_preindex, n (%)	1	5823 (6.1)	2449 (6.0)	113 (4.8)	673 (8.2)	465 (6.8)	2123 (5.7)
myal_flag_1y_preindex, n (%)	1	2236 (2.3)	780 (1.9)	58 (2.5)	291 (3.5)	182 (2.7)	925 (2.5)

myo_flag_preindex, n (%)	1	174 (0.2)	76 (0.2)	5 (0.2)	27 (0.3)	28 (0.4)	38 (0.1)
myo_flag_1y_preindex, n (%)	1	76 (0.1)	28 (0.1)	3 (0.1)	9 (0.1)	18 (0.3)	18 (0.0)
myosi_flag_preindex, n (%)	1	265 (0.3)	142 (0.3)	4 (0.2)	31 (0.4)	28 (0.4)	60 (0.2)
myosi_flag_1y_preindex, n (%)	1	79 (0.1)	42 (0.1)	1 (0.0)	10 (0.1)	11 (0.2)	15 (0.0)
neuro_flag_preindex, n (%)	1	4791 (5.0)	2430 (5.9)	66 (2.8)	412 (5.0)	337 (4.9)	1546 (4.2)
neuro_flag_1y_preindex, n (%)	1	3599 (3.8)	1797 (4.4)	53 (2.3)	293 (3.6)	244 (3.6)	1212 (3.3)
obesity_flag_preindex, n (%)	1	21211 (22.2)	9963 (24.3)	379 (16.2)	1925 (23.4)	1670 (24.4)	7274 (19.6)
obesity_flag_1y_preindex, n (%)	1	15092 (15.8)	7058 (17.2)	269 (11.5)	1331 (16.2)	1199 (17.5)	5235 (14.1)
paralysis_flag_preindex, n (%)	1	766 (0.8)	404 (1.0)	9 (0.4)	58 (0.7)	50 (0.7)	245 (0.7)
paralysis_flag_1y_preindex, n (%)	1	563 (0.6)	291 (0.7)	7 (0.3)	42 (0.5)	36 (0.5)	187 (0.5)
pci_flag_preindex, n (%)	1	1644 (1.7)	1066 (2.6)	7 (0.3)	80 (1.0)	210 (3.1)	281 (0.8)
pci_flag_1y_preindex, n (%)	1	1457 (1.5)	993 (2.4)	6 (0.3)	51 (0.6)	175 (2.6)	232 (0.6)
pci_inp_flag_preindex, n (%)	1	384 (0.4)	258 (0.6)		20 (0.2)	49 (0.7)	57 (0.2)
pci_inp_flag_1y_preindex, n (%)	1	382 (0.4)	258 (0.6)		18 (0.2)	49 (0.7)	57 (0.2)
pg_flag_preindex, n (%)	1	1672 (1.7)	791 (1.9)	25 (1.1)	116 (1.4)	112 (1.6)	628 (1.7)
pg_flag_1y_preindex, n (%)	1	337 (0.4)	154 (0.4)	9 (0.4)	26 (0.3)	20 (0.3)	128 (0.3)
poi_flag_preindex, n (%)	1	57 (0.1)	13 (0.0)	1 (0.0)	11 (0.1)	16 (0.2)	16 (0.0)
poi_flag_1y_preindex, n (%)	1	29 (0.0)	5 (0.0)		8 (0.1)	7 (0.1)	9 (0.0)
psy_flag_preindex, n (%)	1	1672 (1.7)	735 (1.8)	25 (1.1)	158 (1.9)	96 (1.4)	658 (1.8)
psy_flag_1y_preindex, n (%)	1	1083 (1.1)	451 (1.1)	20 (0.9)	95 (1.2)	54 (0.8)	463 (1.2)
pud_flag_preindex, n (%)	1	1392 (1.5)	632 (1.5)	32 (1.4)	167 (2.0)	117 (1.7)	444 (1.2)
pud_flag_1y_preindex, n (%)	1	741 (0.8)	340 (0.8)	20 (0.9)	85 (1.0)	55 (0.8)	241 (0.6)
pud_eb_flag_preindex, n (%)	1	1208 (1.3)	544 (1.3)	28 (1.2)	147 (1.8)	102 (1.5)	387 (1.0)
pud_eb_flag_1y_preindex, n (%)	1	636 (0.7)	282 (0.7)	17 (0.7)	74 (0.9)	50 (0.7)	213 (0.6)
pul_flag_preindex, n (%)	1	3252 (3.4)	1550 (3.8)	37 (1.6)	389 (4.7)	304 (4.4)	972 (2.6)
pul_flag_1y_preindex, n (%)	1	2362 (2.5)	1135 (2.8)	17 (0.7)	259 (3.2)	232 (3.4)	719 (1.9)
pvd_flag_preindex, n (%)	1	7925 (8.3)	3738 (9.1)	127 (5.4)	742 (9.0)	920 (13.4)	2398 (6.4)
pvd_flag_1y_preindex, n (%)	1	5811 (6.1)	2728 (6.7)	89 (3.8)	528 (6.4)	728 (10.6)	1738 (4.7)
renal_flag_preindex, n (%)	1	8557 (9.0)	3762 (9.2)	148 (6.3)	1053 (12.8)	787 (11.5)	2807 (7.5)
renal_flag_1y_preindex, n (%)	1	7498 (7.8)	3257 (8.0)	134 (5.7)	946 (11.5)	693 (10.1)	2468 (6.6)
rhabdo_flag_preindex, n (%)	1	163 (0.2)	91 (0.2)	2 (0.1)	21 (0.3)	15 (0.2)	34 (0.1)
rhabdo_flag_1y_preindex, n (%)	1	71 (0.1)	41 (0.1)	1 (0.0)	12 (0.1)	6 (0.1)	11 (0.0)
rheu_flag_preindex, n (%)	1	6570 (6.9)	2696 (6.6)	171 (7.3)	737 (9.0)	469 (6.9)	2497 (6.7)
rheu_flag_1y_preindex, n (%)	1	3491 (3.7)	1364 (3.3)	100 (4.3)	415 (5.1)	270 (3.9)	1342 (3.6)

tia_flag_preindex, n (%)	1	2639 (2.8)	1255 (3.1)	40 (1.7)	201 (2.4)	222 (3.2)	921 (2.5)
tia_flag_1y_preindex, n (%)	1	1836 (1.9)	880 (2.1)	25 (1.1)	125 (1.5)	129 (1.9)	677 (1.8)
tia_inp_flag_preindex, n (%)	1	338 (0.4)	198 (0.5)	1 (0.0)	19 (0.2)	22 (0.3)	98 (0.3)
tia_inp_flag_1y_preindex, n (%)	1	296 (0.3)	171 (0.4)	1 (0.0)	16 (0.2)	19 (0.3)	89 (0.2)
ua_flag_preindex, n (%)	1	1637 (1.7)	907 (2.2)	8 (0.3)	116 (1.4)	219 (3.2)	387 (1.0)
ua_flag_1y_preindex, n (%)	1	1247 (1.3)	734 (1.8)	6 (0.3)	76 (0.9)	153 (2.2)	278 (0.7)
ua_inp_flag_preindex, n (%)	1	324 (0.3)	200 (0.5)	3 (0.1)	33 (0.4)	38 (0.6)	50 (0.1)
ua_inp_flag_1y_preindex, n (%)	1	299 (0.3)	189 (0.5)	3 (0.1)	28 (0.3)	33 (0.5)	46 (0.1)
valvular_flag_preindex, n (%)	1	7664 (8.0)	3874 (9.5)	94 (4.0)	787 (9.6)	832 (12.2)	2077 (5.6)
valvular_flag_1y_preindex, n (%)	1	5620 (5.9)	2838 (6.9)	69 (2.9)	575 (7.0)	591 (8.6)	1547 (4.2)
vitd_def_flag_preindex, n (%)	1	419 (0.4)	270 (0.7)	2 (0.1)	38 (0.5)	27 (0.4)	82 (0.2)
vitd_def_flag_1y_preindex, n (%)	1	196 (0.2)	107 (0.3)	2 (0.1)	20 (0.2)	19 (0.3)	48 (0.1)
weight_loss_flag_preindex, n (%)	1	3376 (3.5)	1639 (4.0)	49 (2.1)	374 (4.6)	253 (3.7)	1061 (2.9)
weight_loss_flag_1y_preindex, n (%)	1	1716 (1.8)	801 (2.0)	30 (1.3)	191 (2.3)	123 (1.8)	571 (1.5)
ascvd_flag_preindex, n (%)	1	2858 (3.0)	1856 (4.5)	8 (0.3)	161 (2.0)	248 (3.6)	585 (1.6)
ascvd_flag_1y_preindex, n (%)	1	2658 (2.8)	1726 (4.2)	7 (0.3)	139 (1.7)	228 (3.3)	558 (1.5)
mace_4p_flag_preindex, n (%)	1	2957 (3.1)	1909 (4.7)	9 (0.4)	169 (2.1)	267 (3.9)	603 (1.6)
mace_4p_flag_1y_preindex, n (%)	1	2712 (2.8)	1757 (4.3)	8 (0.3)	142 (1.7)	238 (3.5)	567 (1.5)
sas_flag_preindex, n (%)	1	31797 (33.3)	13955 (34.1)	787 (33.6)	2969 (36.1)	2398 (35.0)	11688 (31.4)
sas_flag_1y_preindex, n (%)	1	29088 (30.4)	12609 (30.8)	752 (32.1)	2688 (32.7)	2135 (31.2)	10904 (29.3)
sams_icd_flag_preindex, n (%)	1	15367 (16.1)	7012 (17.1)	256 (10.9)	1560 (19.0)	1182 (17.3)	5357 (14.4)
sams_icd_flag_1y_preindex, n (%)	1	6160 (6.4)	2599 (6.3)	125 (5.3)	665 (8.1)	494 (7.2)	2277 (6.1)
age_adjusted_charlson_deyo_score_preindex, mean (SD)		7.2 (17.5)	7.8 (19.1)	5.1 (9.5)	9.1 (23.6)	9.2 (21.4)	5.7 (13.1)
charlson_quan_score_preindex, mean (SD)		3.4 (15.7)	3.9 (16.9)	1.5 (7.1)	4.9 (20.9)	4.9 (19.6)	2.4 (12.1)
elixhauser_quan_score_preindex, mean (SD)		6.4 (62.2)	7.6 (64.6)	0.1 (24.9)	11.7 (89.8)	12.9 (91.3)	3.1 (44.4)
any_lab_recent_days_preindex, mean (SD)		-80.2 (165.3)	-80.8 (169.9)	-89.9 (171.1)	-76.0 (155.0)	-81.4 (161.4)	-79.7 (162.7)
any_lab_collected_preindex, n (%)	1	95555 (100.0)	40951 (100.0)	2342 (100.0)	8217 (100.0)	6843 (100.0)	37202 (100.0)
any_lab_collected_6mo_preindex, n (%)	1	81814 (85.6)	35037 (85.6)	1962 (83.8)	7097 (86.4)	5848 (85.5)	31870 (85.7)

any_lab_collected_1yr_preindex, n (%)	1	90334 (94.5)	38628 (94.3)	2213 (94.5)	7803 (95.0)	6522 (95.3)	35168 (94.5)
any_lab_count_preindex, mean (SD)		173.6 (432.2)	177.6 (431.3)	109.9 (244.0)	252.2 (731.0)	213.7 (501.9)	148.4 (323.2)
ldl_collected_1y_preindex, n (%)	1	95555 (100.0)	40951 (100.0)	2342 (100.0)	8217 (100.0)	6843 (100.0)	37202 (100.0)
hdl_collected_1y_preindex, n (%)	1	94033 (98.4)	40383 (98.6)	2315 (98.8)	8069 (98.2)	6757 (98.7)	36509 (98.1)
tchol_collected_1y_preindex, n (%)	1	93953 (98.3)	40354 (98.5)	2313 (98.8)	8057 (98.1)	6757 (98.7)	36472 (98.0)
tg_collected_1y_preindex, n (%)	1	93722 (98.1)	40286 (98.4)	2310 (98.6)	8010 (97.5)	6753 (98.7)	36363 (97.7)
ck_collected_1y_preindex, n (%)	1	4627 (4.8)	1763 (4.3)	109 (4.7)	544 (6.6)	511 (7.5)	1700 (4.6)
allergy_flag_1y_preindex, n (%)	1	414 (0.4)	69 (0.2)	19 (0.8)	112 (1.4)	145 (2.1)	69 (0.2)

Table A.3: AHC-IE data cohort: Baseline covariates across different Statin intensity

		Grouped by statin_intensity			
	Label	Overall	High	Low	Moderate
n		95555	14855	14492	66208
age_at_index, mean (SD)		59.1 (13.1)	61.3 (12.6)	59.0 (13.7)	58.6 (13.0)
Gender					
Female, n (%)	1	45095 (47.2)	5669 (38.2)	8107 (55.9)	31319 (47.3)
Male, n (%)	1	50452 (52.8)	9184 (61.8)	6384 (44.1)	34884 (52.7)
Unknown, n (%)	1	8 (0.0)	2 (0.0)	1 (0.0)	5 (0.0)
Race					
white, n (%)	1	82635 (86.5)	12594 (84.8)	12587 (86.9)	57454 (86.8)
black, n (%)	1	5051 (5.3)	931 (6.3)	724 (5.0)	3396 (5.1)
asian, n (%)	1	4214 (4.4)	715 (4.8)	609 (4.2)	2890 (4.4)
indian_alaska_native, n (%)	1	652 (0.7)	120 (0.8)	119 (0.8)	413 (0.6)
native_hawaiian_pacific, n (%)	1	151 (0.2)	21 (0.1)	21 (0.1)	109 (0.2)
hispanic_latino, n (%)	1	1125 (1.2)	182 (1.2)	175 (1.2)	768 (1.2)
statin_generic_name, n (%)	atorvastatin	40951 (42.9)	12003 (80.8)		28948 (43.7)
statin_generic_name, n (%)	rosuvastatin	6843 (7.2)	2852 (19.2)		3991 (6.0)
statin_generic_name, n (%)	lovastatin	2342 (2.5)		1588 (11.0)	754 (1.1)
statin_generic_name, n (%)	pravastatin	8217 (8.6)		5290 (36.5)	2927 (4.4)
statin_generic_name, n (%)	simvastatin	37202 (38.9)		7614 (52.5)	29588 (44.7)
statin_1y_discontinued, n (%)	1	36421 (38.1)	6570 (44.2)	5993 (41.4)	23858 (36.0)
sas_flag_1y_postindex, n (%)	1	30596 (32.0)	5443 (36.6)	4944 (34.1)	20209 (30.5)
ldl_value_avg_1y_preindex, mean (SD)		127.8 (44.5)	114.6 (48.5)	130.5 (40.2)	130.2 (44.0)
ldl_value_avg_1y_postindex, mean (SD)		94.7 (34.3)	80.6 (35.7)	104.8 (32.4)	95.3 (33.5)
any_services_count_1y_preindex, mean (SD)		18.4 (20.8)	20.7 (22.7)	19.5 (22.2)	17.6 (20.0)
any_services_recent_days_preindex, mean (SD)		-0.8 (8.0)	-1.1 (10.4)	-0.7 (6.5)	-0.8 (7.7)
hospital_at_index, n (%)	1	5757 (6.0)	2124 (14.3)	740 (5.1)	2893 (4.4)
edv_at_index, n (%)	1	355 (0.4)	136 (0.9)	38 (0.3)	181 (0.3)
edv_count_1y_preindex, mean (SD)		0.2 (0.6)	0.3 (0.8)	0.1 (0.6)	0.1 (0.6)
edv_days_1y_preindex, n (%)	1	1940 (2.0)	558 (3.8)	260 (1.8)	1122 (1.7)
outpatient_at_index, n (%)	1	77565 (81.2)	11027 (74.2)	11991 (82.7)	54547 (82.4)
outpatient_count_30d_preindex, mean (SD)		3.0 (4.3)	3.1 (3.8)	3.1 (6.2)	2.9 (3.9)
outpatient_count_0_1y_preindex, mean (SD)		17.9 (20.2)	19.9 (22.1)	19.1 (21.6)	17.2 (19.4)
outpatient_count_1_2y_preindex, mean (SD)		11.0 (17.1)	11.8 (18.8)	12.2 (18.9)	10.5 (16.2)
outpatient_count_2_3y_preindex, mean (SD)		8.7 (15.0)	9.6 (16.9)	9.5 (15.8)	8.3 (14.4)
outpatient_recent_days_preindex, mean (SD)		-0.8 (8.3)	-1.2 (10.7)	-0.7 (6.6)	-0.8 (8.1)
weight_preindex, mean (SD)		196.4 (49.6)	199.3 (49.8)	192.1 (49.8)	196.8 (49.4)
bmi_days_preindex, mean (SD)		-48.6 (197.9)	-71.9 (279.3)	-39.3 (169.0)	-45.4 (180.7)
bmi_collected_preindex, n (%)	1	95555 (100.0)	14855 (100.0)	14492 (100.0)	66208 (100.0)
bmi_preindex, mean (SD)		30.7 (6.8)	30.8 (6.8)	30.5 (6.9)	30.8 (6.7)
bp_days_preindex, mean (SD)		-36.9 (190.4)	-70.6 (293.9)	-27.6 (145.1)	-31.4 (167.5)
bp_collected_preindex, n (%)	1	95555 (100.0)	14855 (100.0)	14492 (100.0)	66208 (100.0)
bp_treated_preindex, n (%)	1	51578 (54.0)	8437 (56.8)	7946 (54.8)	35195 (53.2)
bp_systolic_preindex, mean (SD)		127.8 (16.5)	129.9 (17.9)	126.8 (16.1)	127.6 (16.2)
bp_diastolic_preindex, mean (SD)		76.8 (10.6)	77.4 (11.6)	75.9 (10.2)	76.9 (10.4)
alcohol_use_preindex, n (%)	1	42789 (44.8)	6875 (46.3)	6772 (46.7)	29142 (44.0)

drug_use_preindex, n (%)	1	1428 (1.5)	355 (2.4)	168 (1.2)	905 (1.4)
tobacco_use_preindex, n (%)	1	24922 (26.1)	4646 (31.3)	3719 (25.7)	16557 (25.0)
tobacco_previous_use_preindex, n (%)	1	13835 (14.5)	2426 (16.3)	2207 (15.2)	9202 (13.9)
smokeless_tobacco_use_preindex, n (%)	1	1270 (1.3)	260 (1.8)	154 (1.1)	856 (1.3)
smokeless_tobacco_previous_use_preindex, n (%)	1	2416 (2.5)	510 (3.4)	330 (2.3)	1576 (2.4)
chew_tobacco_use_preindex, n (%)	1	1403 (1.5)	263 (1.8)	176 (1.2)	964 (1.5)
snuff_use_preindex, n (%)	1	198 (0.2)	52 (0.4)	23 (0.2)	123 (0.2)
smoking_tobacco_use_preindex, n (%)	1	11780 (12.3)	2257 (15.2)	1655 (11.4)	7868 (11.9)
smoking_tobacco_previous_use_preindex, n (%)	1	25766 (27.0)	4602 (31.0)	4119 (28.4)	17045 (25.7)
cigars_use_preindex, n (%)	1	1264 (1.3)	244 (1.6)	173 (1.2)	847 (1.3)
pipes_use_preindex, n (%)	1	399 (0.4)	79 (0.5)	57 (0.4)	263 (0.4)
cigarettes_use_preindex, n (%)	1	19112 (20.0)	3477 (23.4)	2979 (20.6)	12656 (19.1)
any_med_use_preindex, n (%)	1	95555 (100.0)	14855 (100.0)	14492 (100.0)	66208 (100.0)
any_med_use_6mo_preindex, n (%)	1	95555 (100.0)	14855 (100.0)	14492 (100.0)	66208 (100.0)
any_med_use_1_preindex, n (%)	1	95555 (100.0)	14855 (100.0)	14492 (100.0)	66208 (100.0)
any_med_count_preindex, mean (SD)		30.6 (59.1)	34.9 (66.9)	33.7 (65.4)	28.9 (55.6)
immunosuppressant_med_use_preindex, n (%)	1	2595 (2.7)	382 (2.6)	556 (3.8)	1657 (2.5)
immunosuppressant_med_use_6mo_preindex, n (%)	1	2372 (2.5)	356 (2.4)	510 (3.5)	1506 (2.3)
immunosuppressant_med_use_days_preindex, mean (SD)		25.4 (228.8)	20.5 (166.0)	37.9 (236.0)	23.8 (239.0)
other_lipid_regulating_med_use_preindex, n (%)	1	6982 (7.3)	1080 (7.3)	1212 (8.4)	4690 (7.1)
other_lipid_regulating_med_use_6mo_preindex, n (%)	1	6503 (6.8)	1010 (6.8)	1112 (7.7)	4381 (6.6)
other_lipid_regulating_med_use_days_preindex, mean (SD)		56.8 (265.7)	51.4 (250.9)	66.9 (288.2)	55.8 (263.7)
ezetimibe_med_use_preindex, n (%)	1	1972 (2.1)	325 (2.2)	319 (2.2)	1328 (2.0)
ezetimibe_med_use_6mo_preindex, n (%)	1	1895 (2.0)	315 (2.1)	303 (2.1)	1277 (1.9)
ezetimibe_med_use_days_preindex, mean (SD)		14.2 (127.5)	14.6 (126.4)	15.2 (134.3)	13.9 (126.1)
pcsk9_med_use_preindex, n (%)	1	29 (0.0)	12 (0.1)	4 (0.0)	13 (0.0)
pcsk9_med_use_6mo_preindex, n (%)	1	24 (0.0)	11 (0.1)	3 (0.0)	10 (0.0)
pcsk9_med_use_days_preindex, mean (SD)		0.1 (8.3)	0.4 (16.9)	0.1 (6.1)	0.0 (5.3)
fibrate_med_use_preindex, n (%)	1	3502 (3.7)	526 (3.5)	570 (3.9)	2406 (3.6)
fibrate_med_use_6mo_preindex, n (%)	1	3299 (3.5)	499 (3.4)	537 (3.7)	2263 (3.4)
fibrate_med_use_days_preindex, mean (SD)		30.1 (191.3)	26.5 (172.9)	35.2 (211.5)	29.9 (190.5)
niacin_med_use_preindex, n (%)	1	1735 (1.8)	250 (1.7)	376 (2.6)	1109 (1.7)
niacin_med_use_6mo_preindex, n (%)	1	1495 (1.6)	220 (1.5)	316 (2.2)	959 (1.4)
niacin_med_use_days_preindex, mean (SD)		12.0 (131.4)	11.0 (138.5)	16.5 (143.7)	11.3 (126.8)
nitrate_med_use_preindex, n (%)	1	1346 (1.4)	462 (3.1)	149 (1.0)	735 (1.1)
nitrate_med_use_6mo_preindex, n (%)	1	1267 (1.3)	450 (3.0)	133 (0.9)	684 (1.0)
nitrate_med_use_days_preindex, mean (SD)		7.7 (97.5)	15.0 (135.0)	5.8 (84.0)	6.5 (89.8)
blood_pressure_med_use_preindex, n (%)	1	59238 (62.0)	10544 (71.0)	8994 (62.1)	39700 (60.0)
blood_pressure_med_use_6mo_preindex, n (%)	1	58321 (61.0)	10430 (70.2)	8825 (60.9)	39066 (59.0)
blood_pressure_med_use_days_preindex, mean (SD)		710.9 (808.4)	690.1 (704.3)	756.9 (909.2)	705.5 (806.5)

ace_inhibitor_med_use_preindex, n (%)	1	31223 (32.7)	5180 (34.9)	4876 (33.6)	21167 (32.0)
ace_inhibitor_med_use_6mo_preindex, n (%)	1	29548 (30.9)	4964 (33.4)	4541 (31.3)	20043 (30.3)
ace_inhibitor_med_use_days_preindex, mean (SD)		300.4 (578.4)	266.7 (507.9)	326.6 (671.4)	302.2 (570.9)
beta_blocker_med_use_preindex, n (%)	1	24978 (26.1)	5178 (34.9)	3568 (24.6)	16232 (24.5)
beta_blocker_med_use_6mo_preindex, n (%)	1	23515 (24.6)	4973 (33.5)	3287 (22.7)	15255 (23.0)
beta_blocker_med_use_days_preindex, mean (SD)		227.8 (525.5)	243.8 (482.6)	232.0 (532.0)	223.3 (533.1)
calcium_channel_blocker_med_use_preindex, n (%)	1	14208 (14.9)	2706 (18.2)	2141 (14.8)	9361 (14.1)
calcium_channel_blocker_med_use_6mo_preindex, n (%)	1	13225 (13.8)	2536 (17.1)	1966 (13.6)	8723 (13.2)
calcium_channel_blocker_med_use_days_preindex, mean (SD)		124.7 (386.6)	125.7 (365.2)	130.1 (401.5)	123.3 (388.0)
diuretic_med_use_preindex, n (%)	1	27401 (28.7)	4150 (27.9)	4388 (30.3)	18863 (28.5)
diuretic_med_use_6mo_preindex, n (%)	1	25883 (27.1)	3907 (26.3)	4130 (28.5)	17846 (27.0)
diuretic_med_use_days_preindex, mean (SD)		273.2 (583.4)	216.8 (473.8)	311.0 (672.0)	277.6 (584.0)
colchicine_med_use_days_preindex, mean (SD)		4.3 (72.6)	5.1 (72.7)	3.7 (64.6)	4.2 (74.2)
amiodarone_med_use_days_preindex, mean (SD)		2.2 (48.2)	3.1 (61.6)	1.7 (40.4)	2.0 (46.3)
diltiazem_med_use_days_preindex, mean (SD)		15.1 (140.6)	12.7 (120.0)	17.6 (149.3)	15.1 (142.8)
verapamil_med_use_days_preindex, mean (SD)		6.0 (88.9)	4.1 (69.3)	7.5 (101.0)	6.1 (90.0)
amlodipine_med_use_days_preindex, mean (SD)		96.2 (340.0)	103.2 (333.9)	98.0 (352.6)	94.2 (338.5)
antifungal_med_use_days_preindex, mean (SD)		14.9 (121.6)	12.4 (111.0)	19.6 (138.2)	14.4 (119.9)
CYP2C9_inhibitors_med_use_days_preindex, mean (SD)		38.9 (204.8)	30.4 (169.3)	50.8 (239.1)	38.2 (203.8)
CYP2C9_inducers_med_use_days_preindex, mean (SD)		3.7 (137.1)	2.9 (56.1)	3.2 (75.5)	4.0 (158.6)
CYP3A4_inhibitors_med_use_days_preindex, mean (SD)		216.9 (520.4)	208.1 (451.6)	237.2 (525.7)	214.4 (533.3)
CYP3A4_inducers_med_use_days_preindex, mean (SD)		16.5 (187.1)	13.2 (121.4)	15.9 (144.6)	17.4 (206.5)
pgp_inhibitors_med_use_days_preindex, mean (SD)		237.2 (538.0)	208.5 (468.4)	274.4 (643.3)	235.5 (526.8)
pgp_inducers_med_use_days_preindex, mean (SD)		3.0 (132.4)	2.7 (54.9)	2.4 (62.5)	3.2 (154.2)
OATP1B1_inhibitors_med_use_days_preindex, mean (SD)		58.6 (305.9)	49.2 (250.3)	61.1 (293.0)	60.2 (319.7)
OATP1B3_inhibitors_med_use_days_preindex, mean (SD)		6.4 (92.7)	5.2 (72.4)	9.9 (119.9)	5.9 (89.9)
aids_flag_preindex, n (%)	1	144 (0.2)	32 (0.2)	20 (0.1)	92 (0.1)
aids_flag_1y_preindex, n (%)	1	135 (0.1)	29 (0.2)	19 (0.1)	87 (0.1)
afib_flag_preindex, n (%)	1	6143 (6.4)	1436 (9.7)	908 (6.3)	3799 (5.7)
afib_flag_1y_preindex, n (%)	1	5317 (5.6)	1238 (8.3)	772 (5.3)	3307 (5.0)
aki_flag_preindex, n (%)	1	3443 (3.6)	905 (6.1)	525 (3.6)	2013 (3.0)
aki_flag_1y_preindex, n (%)	1	2043 (2.1)	557 (3.7)	302 (2.1)	1184 (1.8)
aud_flag_preindex, n (%)	1	12988 (13.6)	2094 (14.1)	1996 (13.8)	8898 (13.4)
aud_flag_1y_preindex, n (%)	1	7829 (8.2)	1182 (8.0)	1253 (8.6)	5394 (8.1)
ane_flag_preindex, n (%)	1	4620 (4.8)	833 (5.6)	819 (5.7)	2968 (4.5)
ane_flag_1y_preindex, n (%)	1	2928 (3.1)	520 (3.5)	500 (3.5)	1908 (2.9)
ang_flag_preindex, n (%)	1	3123 (3.3)	1462 (9.8)	276 (1.9)	1385 (2.1)

ang_flag_1y_preindex, n (%)	1	2381 (2.5)	1205 (8.1)	182 (1.3)	994 (1.5)
ang_cad_inp_flag_preindex, n (%)	1	265 (0.3)	140 (0.9)	30 (0.2)	95 (0.1)
ang_cad_inp_flag_1y_preindex, n (%)	1	249 (0.3)	138 (0.9)	25 (0.2)	86 (0.1)
arr_flag_preindex, n (%)	1	20412 (21.4)	4395 (29.6)	3122 (21.5)	12895 (19.5)
arr_flag_1y_preindex, n (%)	1	14090 (14.7)	3343 (22.5)	2009 (13.9)	8738 (13.2)
ashd_flag_preindex, n (%)	1	15143 (15.8)	5489 (37.0)	1469 (10.1)	8185 (12.4)
ashd_flag_1y_preindex, n (%)	1	13438 (14.1)	5086 (34.2)	1229 (8.5)	7123 (10.8)
ashd_ap_flag_preindex, n (%)	1	4871 (5.1)	2016 (13.6)	482 (3.3)	2373 (3.6)
ashd_ap_flag_1y_preindex, n (%)	1	3530 (3.7)	1544 (10.4)	317 (2.2)	1669 (2.5)
cabg_flag_preindex, n (%)	1	250 (0.3)	93 (0.6)	30 (0.2)	127 (0.2)
cabg_flag_1y_preindex, n (%)	1	194 (0.2)	72 (0.5)	25 (0.2)	97 (0.1)
cabg_inp_flag_preindex, n (%)	1	53 (0.1)	22 (0.1)	6 (0.0)	25 (0.0)
cabg_inp_flag_1y_preindex, n (%)	1	48 (0.1)	20 (0.1)	6 (0.0)	22 (0.0)
cad_flag_preindex, n (%)	1	15818 (16.6)	5744 (38.7)	1512 (10.4)	8562 (12.9)
cad_flag_1y_preindex, n (%)	1	14124 (14.8)	5366 (36.1)	1266 (8.7)	7492 (11.3)
cancer_flag_preindex, n (%)	1	7315 (7.7)	1332 (9.0)	1121 (7.7)	4862 (7.3)
cancer_flag_1y_preindex, n (%)	1	5489 (5.7)	1032 (6.9)	836 (5.8)	3621 (5.5)
cere_flag_preindex, n (%)	1	9284 (9.7)	2452 (16.5)	1339 (9.2)	5493 (8.3)
cere_flag_1y_preindex, n (%)	1	7548 (7.9)	2092 (14.1)	1062 (7.3)	4394 (6.6)
chf_flag_preindex, n (%)	1	6208 (6.5)	1757 (11.8)	875 (6.0)	3576 (5.4)
chf_flag_1y_preindex, n (%)	1	5242 (5.5)	1527 (10.3)	686 (4.7)	3029 (4.6)
ckd_flag_preindex, n (%)	1	1603 (1.7)	385 (2.6)	234 (1.6)	984 (1.5)
ckd_flag_1y_preindex, n (%)	1	1188 (1.2)	302 (2.0)	175 (1.2)	711 (1.1)
coag_flag_preindex, n (%)	1	3316 (3.5)	667 (4.5)	548 (3.8)	2101 (3.2)
coag_flag_1y_preindex, n (%)	1	2151 (2.3)	431 (2.9)	347 (2.4)	1373 (2.1)
cpd_flag_preindex, n (%)	1	16500 (17.3)	2853 (19.2)	2699 (18.6)	10948 (16.5)
cpd_flag_1y_preindex, n (%)	1	12113 (12.7)	2094 (14.1)	1963 (13.5)	8056 (12.2)
cv_death_flag_preindex, n (%)	1	292 (0.3)	116 (0.8)	31 (0.2)	145 (0.2)
cv_death_flag_1y_preindex, n (%)	1	215 (0.2)	93 (0.6)	21 (0.1)	101 (0.2)
cva_flag_preindex, n (%)	1	501 (0.5)	174 (1.2)	59 (0.4)	268 (0.4)
cva_flag_1y_preindex, n (%)	1	411 (0.4)	145 (1.0)	53 (0.4)	213 (0.3)
cva_inp_flag_preindex, n (%)	1	185 (0.2)	81 (0.5)	20 (0.1)	84 (0.1)
cva_inp_flag_1y_preindex, n (%)	1	160 (0.2)	67 (0.5)	20 (0.1)	73 (0.1)
cva_init_flag_preindex, n (%)	1	4898 (5.1)	1423 (9.6)	709 (4.9)	2766 (4.2)
cva_init_flag_1y_preindex, n (%)	1	4193 (4.4)	1265 (8.5)	594 (4.1)	2334 (3.5)
cva_init_inp_flag_preindex, n (%)	1	1586 (1.7)	516 (3.5)	243 (1.7)	827 (1.2)
cva_init_inp_flag_1y_preindex, n (%)	1	1507 (1.6)	493 (3.3)	231 (1.6)	783 (1.2)
dementia_flag_preindex, n (%)	1	995 (1.0)	177 (1.2)	202 (1.4)	616 (0.9)
dementia_flag_1y_preindex, n (%)	1	804 (0.8)	137 (0.9)	166 (1.1)	501 (0.8)
dep_flag_preindex, n (%)	1	21916 (22.9)	3454 (23.3)	3494 (24.1)	14968 (22.6)
dep_flag_1y_preindex, n (%)	1	17274 (18.1)	2654 (17.9)	2780 (19.2)	11840 (17.9)
dm_flag_preindex, n (%)	1	28281 (29.6)	4763 (32.1)	4667 (32.2)	18851 (28.5)
dm_flag_1y_preindex, n (%)	1	27318 (28.6)	4534 (30.5)	4523 (31.2)	18261 (27.6)
dm1_flag_preindex, n (%)	1	2659 (2.8)	416 (2.8)	476 (3.3)	1767 (2.7)
dm1_flag_1y_preindex, n (%)	1	2064 (2.2)	320 (2.2)	360 (2.5)	1384 (2.1)
dm2_flag_preindex, n (%)	1	27782 (29.1)	4708 (31.7)	4577 (31.6)	18497 (27.9)
dm2_flag_1y_preindex, n (%)	1	26508 (27.7)	4435 (29.9)	4379 (30.2)	17694 (26.7)
dm_c_flag_preindex, n (%)	1	11328 (11.9)	2374 (16.0)	1724 (11.9)	7230 (10.9)

dm_c_flag_1y_preindex, n (%)	1	9897 (10.4)	2123 (14.3)	1474 (10.2)	6300 (9.5)
dm_uc_flag_preindex, n (%)	1	26811 (28.1)	4431 (29.8)	4513 (31.1)	17867 (27.0)
dm_uc_flag_1y_preindex, n (%)	1	25374 (26.6)	4085 (27.5)	4316 (29.8)	16973 (25.6)
dm_cc_flag_preindex, n (%)	1	9969 (10.4)	2121 (14.3)	1501 (10.4)	6347 (9.6)
dm_cc_flag_1y_preindex, n (%)	1	8727 (9.1)	1889 (12.7)	1275 (8.8)	5563 (8.4)
dm_ucc_flag_preindex, n (%)	1	27762 (29.1)	4646 (31.3)	4596 (31.7)	18520 (28.0)
dm_ucc_flag_1y_preindex, n (%)	1	26748 (28.0)	4398 (29.6)	4438 (30.6)	17912 (27.1)
dud_flag_preindex, n (%)	1	2644 (2.8)	608 (4.1)	349 (2.4)	1687 (2.5)
dud_flag_1y_preindex, n (%)	1	1558 (1.6)	368 (2.5)	210 (1.4)	980 (1.5)
fluid_flag_preindex, n (%)	1	10269 (10.7)	2216 (14.9)	1632 (11.3)	6421 (9.7)
fluid_flag_1y_preindex, n (%)	1	6275 (6.6)	1411 (9.5)	985 (6.8)	3879 (5.9)
hcabg_flag_preindex, n (%)	1	1689 (1.8)	784 (5.3)	142 (1.0)	763 (1.2)
hemiplegia_flag_preindex, n (%)	1	768 (0.8)	180 (1.2)	132 (0.9)	456 (0.7)
hemiplegia_flag_1y_preindex, n (%)	1	565 (0.6)	143 (1.0)	99 (0.7)	323 (0.5)
hmi_flag_preindex, n (%)	1	1980 (2.1)	805 (5.4)	176 (1.2)	999 (1.5)
hpci_flag_preindex, n (%)	1	2665 (2.8)	1381 (9.3)	181 (1.2)	1103 (1.7)
htn_c_flag_preindex, n (%)	1	5796 (6.1)	1237 (8.3)	985 (6.8)	3574 (5.4)
htn_c_flag_1y_preindex, n (%)	1	4760 (5.0)	1008 (6.8)	828 (5.7)	2924 (4.4)
htn_uc_flag_preindex, n (%)	1	57971 (60.7)	10069 (67.8)	8783 (60.6)	39119 (59.1)
htn_uc_flag_1y_preindex, n (%)	1	55243 (57.8)	9553 (64.3)	8363 (57.7)	37327 (56.4)
hypo_flag_preindex, n (%)	1	12644 (13.2)	1799 (12.1)	2207 (15.2)	8638 (13.0)
hypo_flag_1y_preindex, n (%)	1	11609 (12.1)	1629 (11.0)	2031 (14.0)	7949 (12.0)
hypo_bg_flag_preindex, n (%)	1	12325 (12.9)	1767 (11.9)	2142 (14.8)	8416 (12.7)
hypo_bg_flag_1y_preindex, n (%)	1	11210 (11.7)	1565 (10.5)	1960 (13.5)	7685 (11.6)
ld_flag_preindex, n (%)	1	2885 (3.0)	508 (3.4)	502 (3.5)	1875 (2.8)
ld_flag_1y_preindex, n (%)	1	1135 (1.2)	192 (1.3)	216 (1.5)	727 (1.1)
ld_mild_flag_preindex, n (%)	1	4994 (5.2)	885 (6.0)	853 (5.9)	3256 (4.9)
ld_mild_flag_1y_preindex, n (%)	1	2975 (3.1)	514 (3.5)	536 (3.7)	1925 (2.9)
ld_ms_flag_preindex, n (%)	1	416 (0.4)	61 (0.4)	89 (0.6)	266 (0.4)
ld_ms_flag_1y_preindex, n (%)	1	249 (0.3)	32 (0.2)	60 (0.4)	157 (0.2)
lymphoma_flag_preindex, n (%)	1	751 (0.8)	139 (0.9)	133 (0.9)	479 (0.7)
lymphoma_flag_1y_preindex, n (%)	1	546 (0.6)	103 (0.7)	98 (0.7)	345 (0.5)
met_flag_preindex, n (%)	1	986 (1.0)	230 (1.5)	149 (1.0)	607 (0.9)
met_flag_1y_preindex, n (%)	1	653 (0.7)	153 (1.0)	96 (0.7)	404 (0.6)
mi_flag_preindex, n (%)	1	1026 (1.1)	397 (2.7)	104 (0.7)	525 (0.8)
mi_flag_1y_preindex, n (%)	1	737 (0.8)	304 (2.0)	67 (0.5)	366 (0.6)
mi_inp_flag_preindex, n (%)	1	268 (0.3)	124 (0.8)	30 (0.2)	114 (0.2)
mi_inp_flag_1y_preindex, n (%)	1	238 (0.2)	116 (0.8)	23 (0.2)	99 (0.1)
mi_init_flag_preindex, n (%)	1	3495 (3.7)	1862 (12.5)	221 (1.5)	1412 (2.1)
mi_init_flag_1y_preindex, n (%)	1	2903 (3.0)	1604 (10.8)	180 (1.2)	1119 (1.7)
mi_init_inp_flag_preindex, n (%)	1	1097 (1.1)	601 (4.0)	80 (0.6)	416 (0.6)
mi_init_inp_flag_1y_preindex, n (%)	1	1022 (1.1)	581 (3.9)	74 (0.5)	367 (0.6)
mus_oth_flag_preindex, n (%)	1	11148 (11.7)	1919 (12.9)	1825 (12.6)	7404 (11.2)
mus_oth_flag_1y_preindex, n (%)	1	4140 (4.3)	739 (5.0)	693 (4.8)	2708 (4.1)
myal_flag_preindex, n (%)	1	5823 (6.1)	800 (5.4)	1119 (7.7)	3904 (5.9)
myal_flag_1y_preindex, n (%)	1	2236 (2.3)	265 (1.8)	475 (3.3)	1496 (2.3)
myo_flag_preindex, n (%)	1	174 (0.2)	35 (0.2)	39 (0.3)	100 (0.2)
myo_flag_1y_preindex, n (%)	1	76 (0.1)	16 (0.1)	16 (0.1)	44 (0.1)

myosi_flag_preindex, n (%)	1	265 (0.3)	50 (0.3)	45 (0.3)	170 (0.3)
myosi_flag_1y_preindex, n (%)	1	79 (0.1)	18 (0.1)	16 (0.1)	45 (0.1)
neuro_flag_preindex, n (%)	1	4791 (5.0)	1075 (7.2)	727 (5.0)	2989 (4.5)
neuro_flag_1y_preindex, n (%)	1	3599 (3.8)	814 (5.5)	549 (3.8)	2236 (3.4)
obesity_flag_preindex, n (%)	1	21211 (22.2)	3571 (24.0)	3266 (22.5)	14374 (21.7)
obesity_flag_1y_preindex, n (%)	1	15092 (15.8)	2576 (17.3)	2258 (15.6)	10258 (15.5)
paralysis_flag_preindex, n (%)	1	766 (0.8)	180 (1.2)	131 (0.9)	455 (0.7)
paralysis_flag_1y_preindex, n (%)	1	563 (0.6)	143 (1.0)	97 (0.7)	323 (0.5)
pci_flag_preindex, n (%)	1	1644 (1.7)	1052 (7.1)	80 (0.6)	512 (0.8)
pci_flag_1y_preindex, n (%)	1	1457 (1.5)	978 (6.6)	52 (0.4)	427 (0.6)
pci_inp_flag_preindex, n (%)	1	384 (0.4)	243 (1.6)	19 (0.1)	122 (0.2)
pci_inp_flag_1y_preindex, n (%)	1	382 (0.4)	243 (1.6)	17 (0.1)	122 (0.2)
pg_flag_preindex, n (%)	1	1672 (1.7)	233 (1.6)	308 (2.1)	1131 (1.7)
pg_flag_1y_preindex, n (%)	1	337 (0.4)	50 (0.3)	61 (0.4)	226 (0.3)
poi_flag_preindex, n (%)	1	57 (0.1)	3 (0.0)	15 (0.1)	39 (0.1)
poi_flag_1y_preindex, n (%)	1	29 (0.0)		8 (0.1)	21 (0.0)
psy_flag_preindex, n (%)	1	1672 (1.7)	272 (1.8)	301 (2.1)	1099 (1.7)
psy_flag_1y_preindex, n (%)	1	1083 (1.1)	166 (1.1)	208 (1.4)	709 (1.1)
pud_flag_preindex, n (%)	1	1392 (1.5)	273 (1.8)	243 (1.7)	876 (1.3)
pud_flag_1y_preindex, n (%)	1	741 (0.8)	148 (1.0)	122 (0.8)	471 (0.7)
pud_eb_flag_preindex, n (%)	1	1208 (1.3)	239 (1.6)	213 (1.5)	756 (1.1)
pud_eb_flag_1y_preindex, n (%)	1	636 (0.7)	127 (0.9)	109 (0.8)	400 (0.6)
pul_flag_preindex, n (%)	1	3252 (3.4)	752 (5.1)	519 (3.6)	1981 (3.0)
pul_flag_1y_preindex, n (%)	1	2362 (2.5)	557 (3.7)	350 (2.4)	1455 (2.2)
pvd_flag_preindex, n (%)	1	7925 (8.3)	2041 (13.7)	1078 (7.4)	4806 (7.3)
pvd_flag_1y_preindex, n (%)	1	5811 (6.1)	1596 (10.7)	732 (5.1)	3483 (5.3)
renal_flag_preindex, n (%)	1	8557 (9.0)	1718 (11.6)	1550 (10.7)	5289 (8.0)
renal_flag_1y_preindex, n (%)	1	7498 (7.8)	1490 (10.0)	1374 (9.5)	4634 (7.0)
rhabdo_flag_preindex, n (%)	1	163 (0.2)	43 (0.3)	30 (0.2)	90 (0.1)
rhabdo_flag_1y_preindex, n (%)	1	71 (0.1)	18 (0.1)	17 (0.1)	36 (0.1)
rheu_flag_preindex, n (%)	1	6570 (6.9)	958 (6.4)	1239 (8.5)	4373 (6.6)
rheu_flag_1y_preindex, n (%)	1	3491 (3.7)	520 (3.5)	679 (4.7)	2292 (3.5)
tia_flag_preindex, n (%)	1	2639 (2.8)	569 (3.8)	413 (2.8)	1657 (2.5)
tia_flag_1y_preindex, n (%)	1	1836 (1.9)	390 (2.6)	291 (2.0)	1155 (1.7)
tia_inp_flag_preindex, n (%)	1	338 (0.4)	84 (0.6)	50 (0.3)	204 (0.3)
tia_inp_flag_1y_preindex, n (%)	1	296 (0.3)	76 (0.5)	47 (0.3)	173 (0.3)
ua_flag_preindex, n (%)	1	1637 (1.7)	791 (5.3)	124 (0.9)	722 (1.1)
ua_flag_1y_preindex, n (%)	1	1247 (1.3)	645 (4.3)	83 (0.6)	519 (0.8)
ua_inp_flag_preindex, n (%)	1	324 (0.3)	156 (1.1)	34 (0.2)	134 (0.2)
ua_inp_flag_1y_preindex, n (%)	1	299 (0.3)	151 (1.0)	28 (0.2)	120 (0.2)
valvular_flag_preindex, n (%)	1	7664 (8.0)	1923 (12.9)	1121 (7.7)	4620 (7.0)
valvular_flag_1y_preindex, n (%)	1	5620 (5.9)	1476 (9.9)	791 (5.5)	3353 (5.1)
vitd_def_flag_preindex, n (%)	1	419 (0.4)	80 (0.5)	66 (0.5)	273 (0.4)
vitd_def_flag_1y_preindex, n (%)	1	196 (0.2)	32 (0.2)	43 (0.3)	121 (0.2)
weight_loss_flag_preindex, n (%)	1	3376 (3.5)	675 (4.5)	586 (4.0)	2115 (3.2)
weight_loss_flag_1y_preindex, n (%)	1	1716 (1.8)	350 (2.4)	310 (2.1)	1056 (1.6)
ascvd_flag_preindex, n (%)	1	2858 (3.0)	1230 (8.3)	326 (2.2)	1302 (2.0)
ascvd_flag_1y_preindex, n (%)	1	2658 (2.8)	1171 (7.9)	302 (2.1)	1185 (1.8)

mace_4p_flag_preindex, n (%)	1	2957 (3.1)	1280 (8.6)	331 (2.3)	1346 (2.0)
mace_4p_flag_1y_preindex, n (%)	1	2712 (2.8)	1205 (8.1)	299 (2.1)	1208 (1.8)
sas_flag_preindex, n (%)	1	31797 (33.3)	5499 (37.0)	5207 (35.9)	21091 (31.9)
sas_flag_1y_preindex, n (%)	1	29088 (30.4)	4934 (33.2)	4793 (33.1)	19361 (29.2)
sams_icd_flag_preindex, n (%)	1	15367 (16.1)	2458 (16.5)	2668 (18.4)	10241 (15.5)
sams_icd_flag_1y_preindex, n (%)	1	6160 (6.4)	978 (6.6)	1126 (7.8)	4056 (6.1)
age_adjusted_charlson_deyo_score_preindex, mean (SD)		7.2 (17.5)	9.3 (20.3)	7.7 (19.3)	6.6 (16.4)
charlson_quan_score_preindex, mean (SD)		3.4 (15.7)	4.6 (17.6)	3.7 (17.7)	3.0 (14.8)
elixhauser_quan_score_preindex, mean (SD)		6.4 (62.2)	10.9 (67.7)	7.2 (71.1)	5.2 (58.6)
any_lab_recent_days_preindex, mean (SD)		-80.2 (165.3)	-88.1 (181.5)	-68.1 (146.7)	-81.1 (165.2)
any_lab_collected_preindex, n (%)	1	95555 (100.0)	14855 (100.0)	14492 (100.0)	66208 (100.0)
any_lab_collected_6mo_preindex, n (%)	1	81814 (85.6)	12516 (84.3)	12786 (88.2)	56512 (85.4)
any_lab_collected_1yr_preindex, n (%)	1	90334 (94.5)	14013 (94.3)	13885 (95.8)	62436 (94.3)
any_lab_count_preindex, mean (SD)		173.6 (432.2)	196.5 (396.0)	205.6 (566.5)	161.4 (404.7)
ldl_collected_1y_preindex, n (%)	1	95555 (100.0)	14855 (100.0)	14492 (100.0)	66208 (100.0)
hdl_collected_1y_preindex, n (%)	1	94033 (98.4)	14612 (98.4)	14253 (98.4)	65168 (98.4)
tchol_collected_1y_preindex, n (%)	1	93953 (98.3)	14614 (98.4)	14231 (98.2)	65108 (98.3)
tg_collected_1y_preindex, n (%)	1	93722 (98.1)	14583 (98.2)	14178 (97.8)	64961 (98.1)
ck_collected_1y_preindex, n (%)	1	4627 (4.8)	804 (5.4)	773 (5.3)	3050 (4.6)
allergy_flag_1y_preindex, n (%)	1	414 (0.4)	47 (0.3)	106 (0.7)	261 (0.4)

A.2 Figures

A.2.1 Consort Diagram

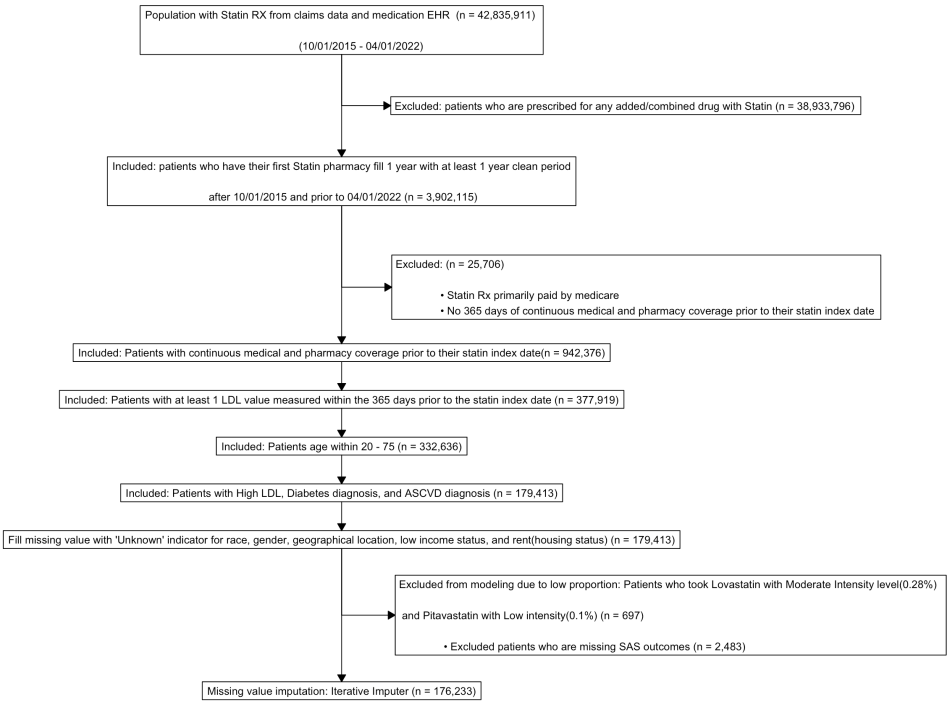


Figure A.1: Consort Diagram of OptumLabs data with inclusion/exclusion criteria

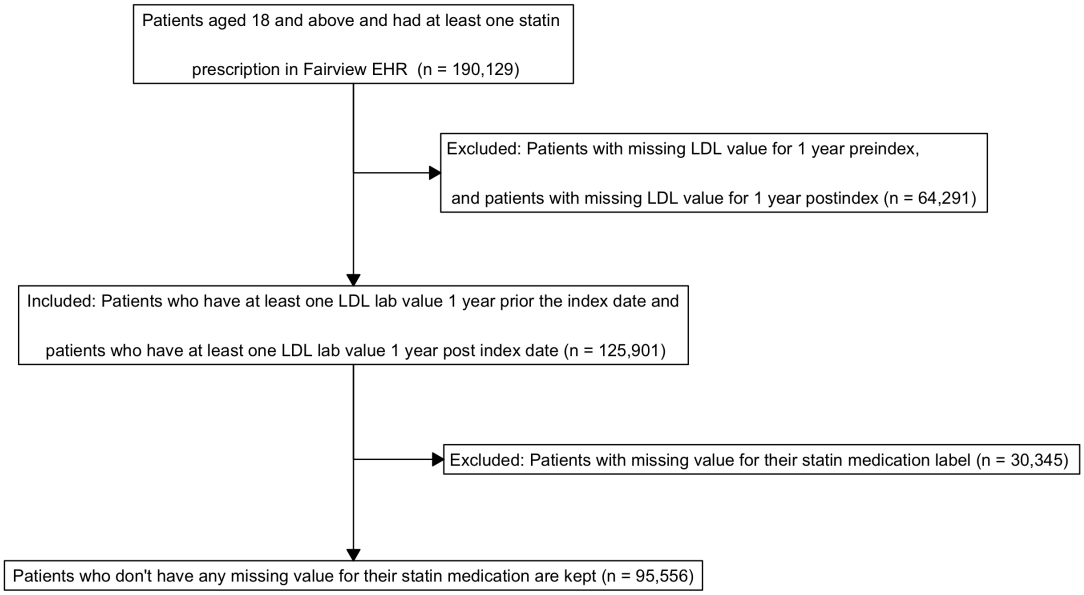
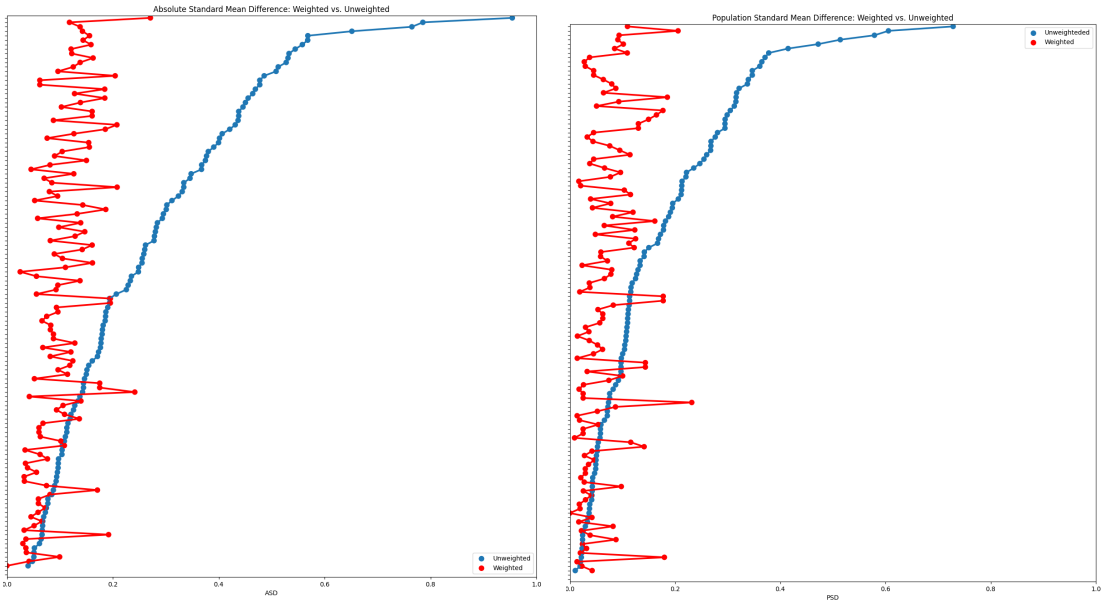


Figure A.2: Consort Diagram of UMN Fairview data with inclusion/exclusion criteria

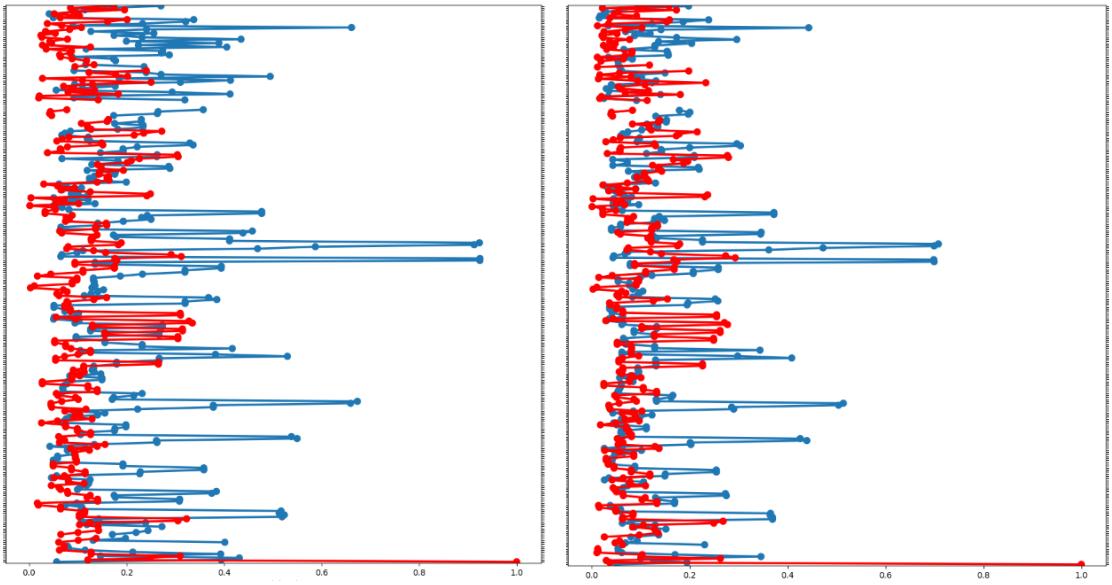
A.2.2 Balancing check



(a) ASD of sorted unweighted covariates for OptumLabs data

(b) PSD of sorted unweighted covariates for OptumLabs data

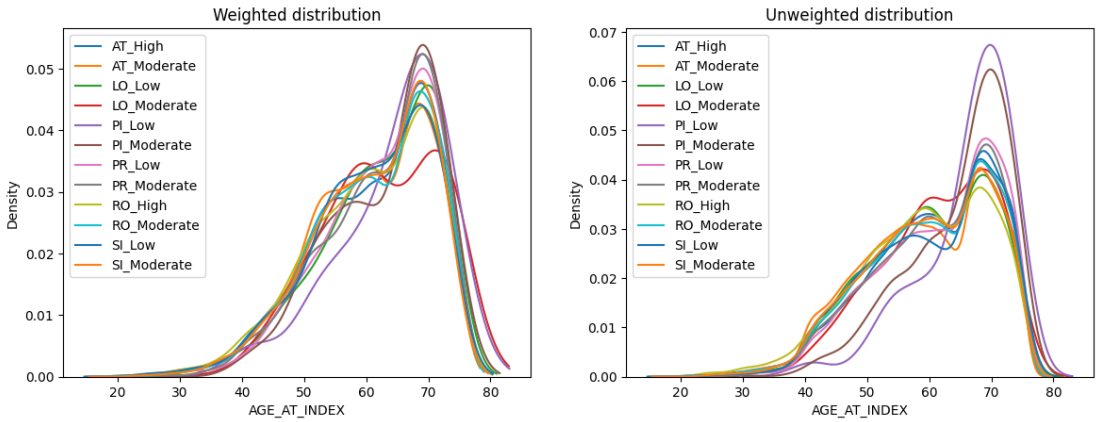
Figure A.3: ASD and PSD of the sorted covariates



(a) ASD of sorted unweighted covariates for AHC-IE data (b) PSD of sorted unweighted covariates for AHC-IE data

Figure A.4: ASD and PSD of the sorted covariates

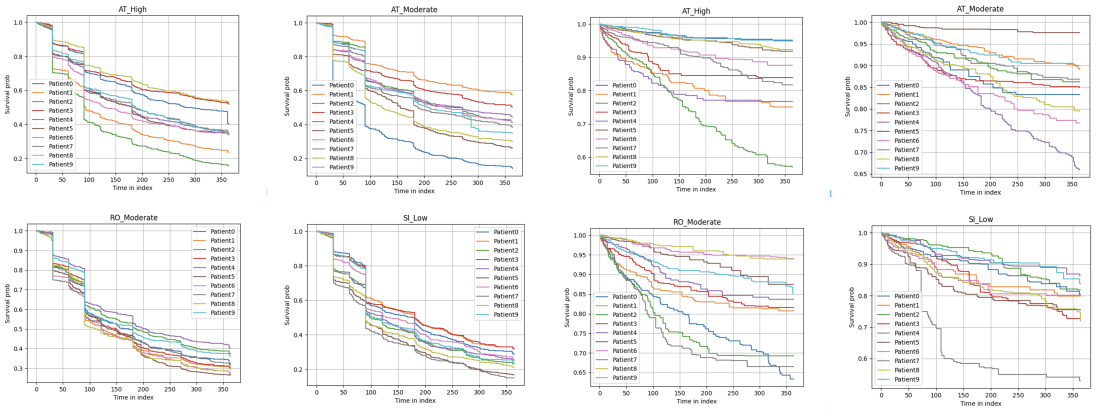
A.2.3 Overlap check



(a) Weighted distribution of patients' age at index date (b) Unweighted distribution of patients' age at index date

Figure A.5: Weighted and unweighted distribution for age variale at index date

A.2.4 SAS/Discontinuation counterfactual survival prediction

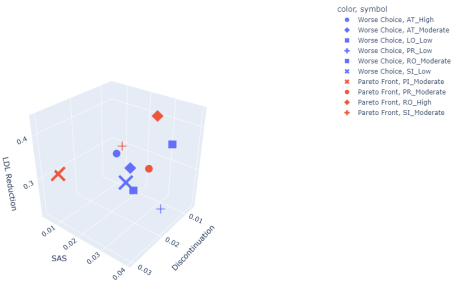


(a) Counterfactual survival plot of SAS for randomly simulated patients for Atorvastatin High/Moderate intensity and Simivastatin Low intensity and Rosuvastatin Moderate intensity

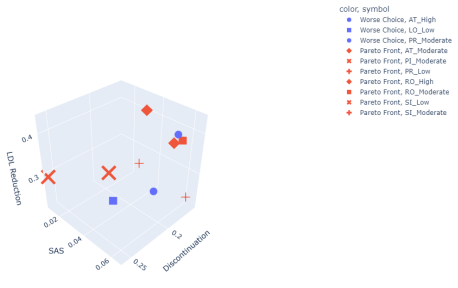
(b) Counterfactual survival plot of Discontinuation for randomly simulated patients for Atorvastatin High/Moderate intensity and Simivastatin Low intensity and Rosuvastatin Moderate intensity

Figure A.6: SAS/Discontinuation counterfactual survival prediction

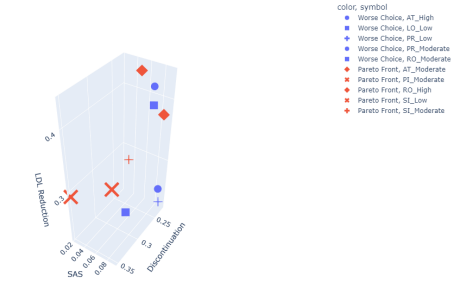
A.2.5 3d Multi-Objective Optimization showcase



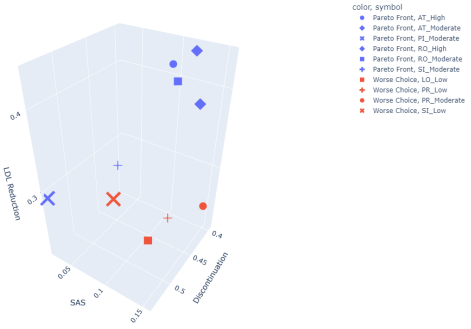
(a) PSTP optimal solutions at 30 days



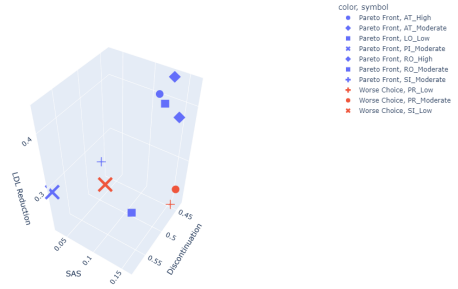
(b) PSTP optimal solutions at 60 days



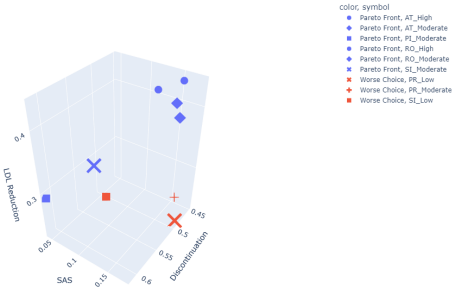
(c) PSTP optimal solutions at 90 days



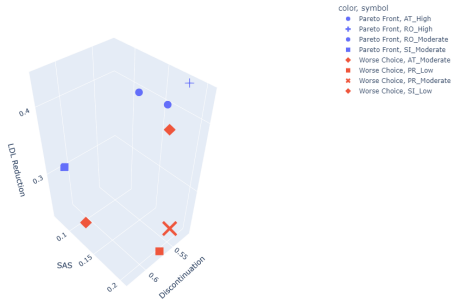
(d) PSTP optimal solutions at 120 days



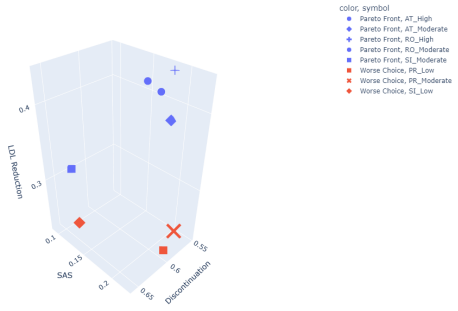
(e) PSTP optimal solutions at 150 days



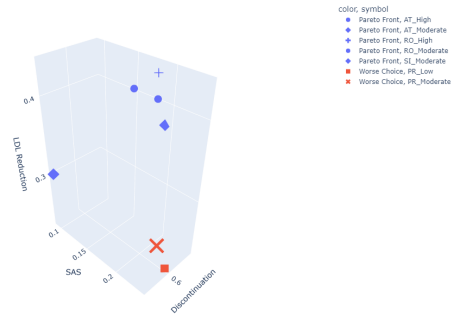
(f) PSTP optimal solutions at 180 days



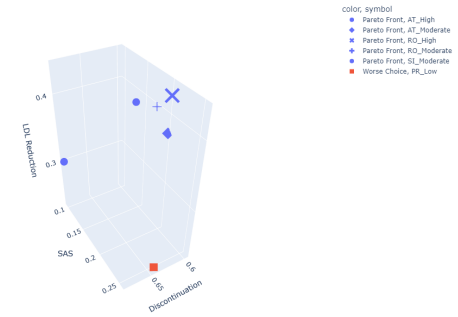
(g) PSTP optimal solutions at 210 days



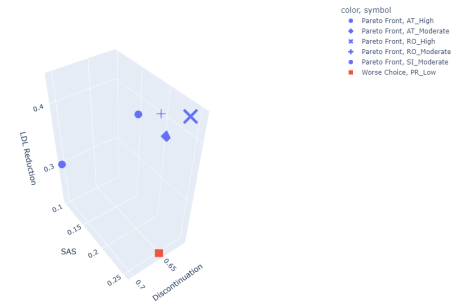
(h) PSTP optimal solutions at 240 days



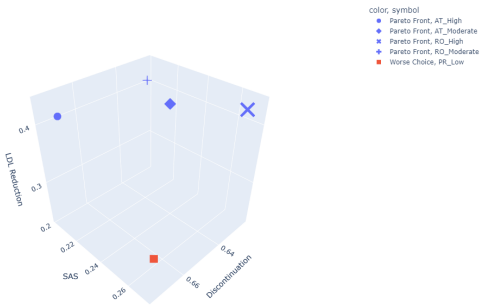
(i) PSTP optimal solutions at 270 days



(j) PSTP optimal solutions at 300 days



(k) PSTP optimal solutions at 330 days



(l) PSTP optimal solutions at 360 days

Figure A.6: Example plot of 3D Pareto Optimization for selecting out PSTP treatment for a simulated patient for 30 days to 360 days survival risks of SAS, discontinuation, and LDL-C reduction. Purple dot means that it is pareto front options. Orange dot is worse choice

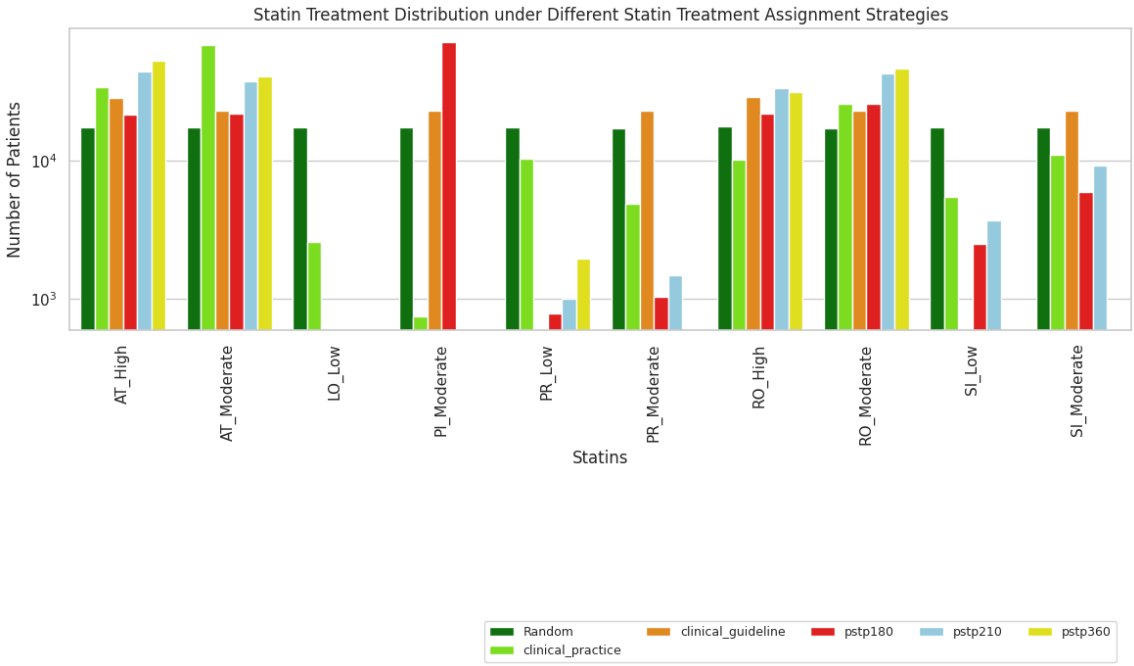
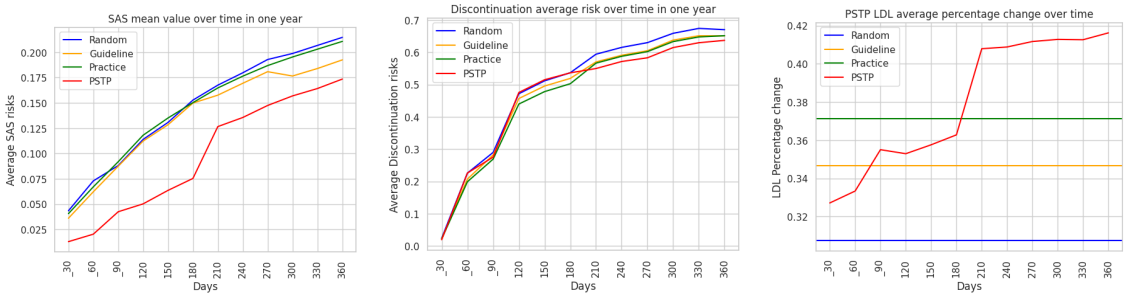


Figure A.7: Summarization of the CTS result with multiple survival days for PSTP predictions result



(a) Summarized SAS survival risks over time for each CTS arm

(b) Summarized Disc survival risks over time for each CTS arm

(c) Summarized LDL-C percentage reduction value over time for each CTS arm

Figure A.8: Summarised SAS/Disc/LDL reduction mean change overtime

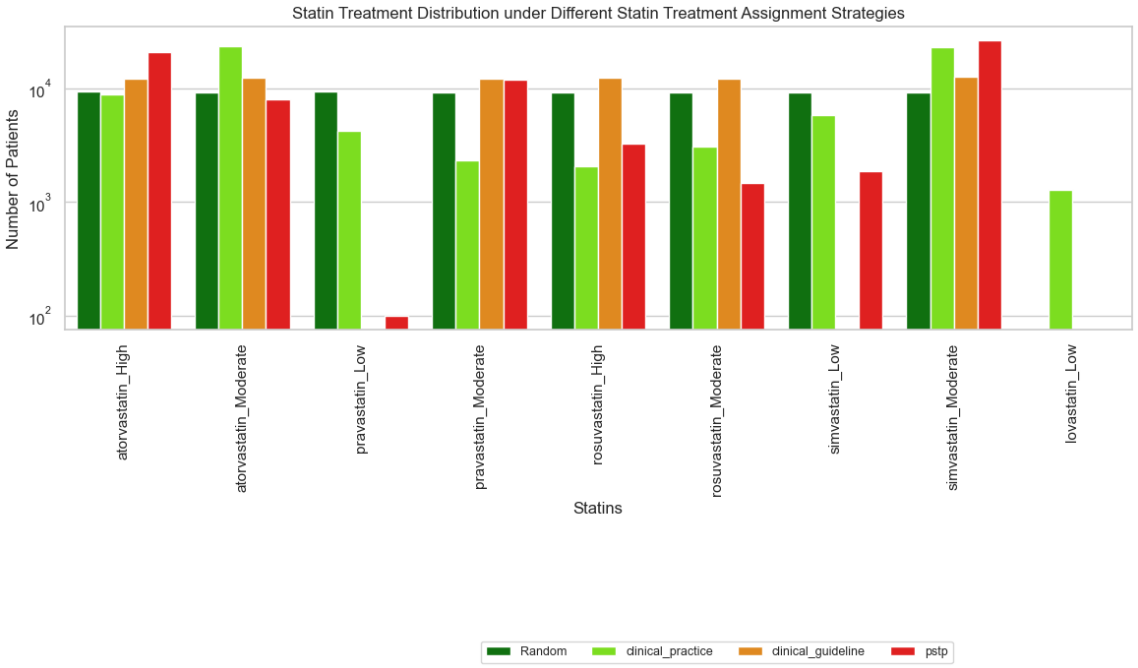


Figure A.9: Summarization of the CTS result using AHC-IE data with baseline model PSTP

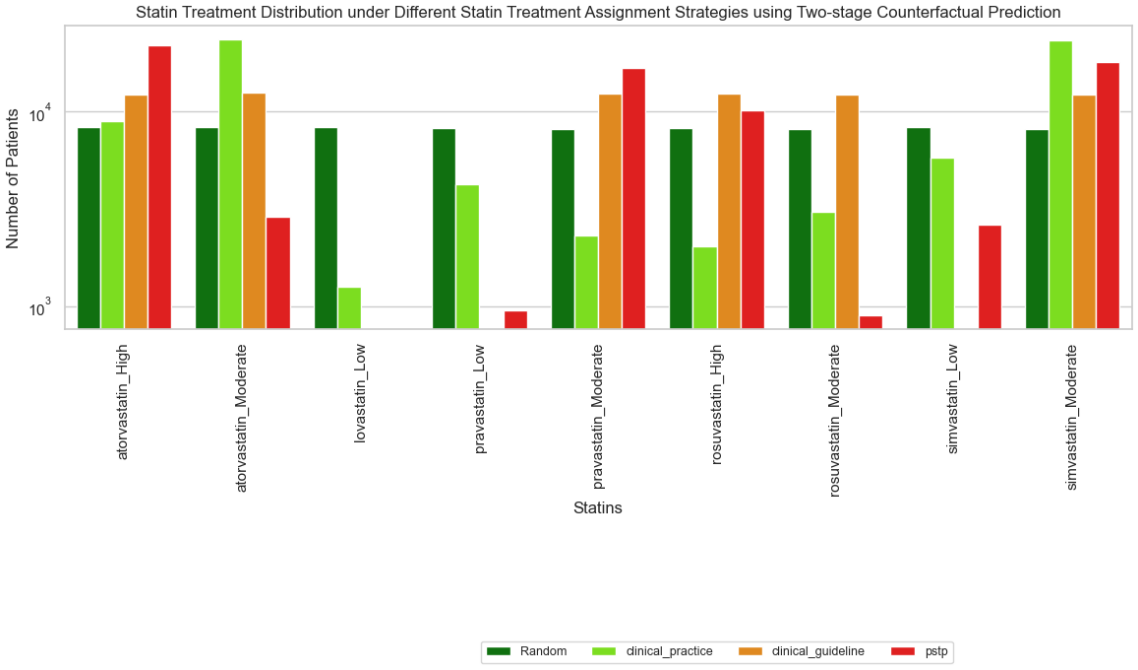
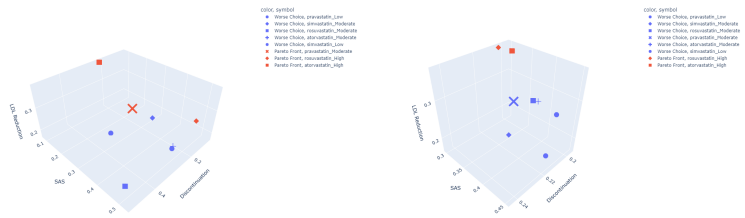


Figure A.10: Summarization of the CTS result using AHC-IE data with two-stage model PSTP



(a) Example of individual plot of pareto optimization for baseline model using AHC-IE data

(b) Example of individual plot of pareto optimization for twostage model using AHC-IE data

Figure A.11: Summarised SAS/Disc/LDL reduction mean change overtime

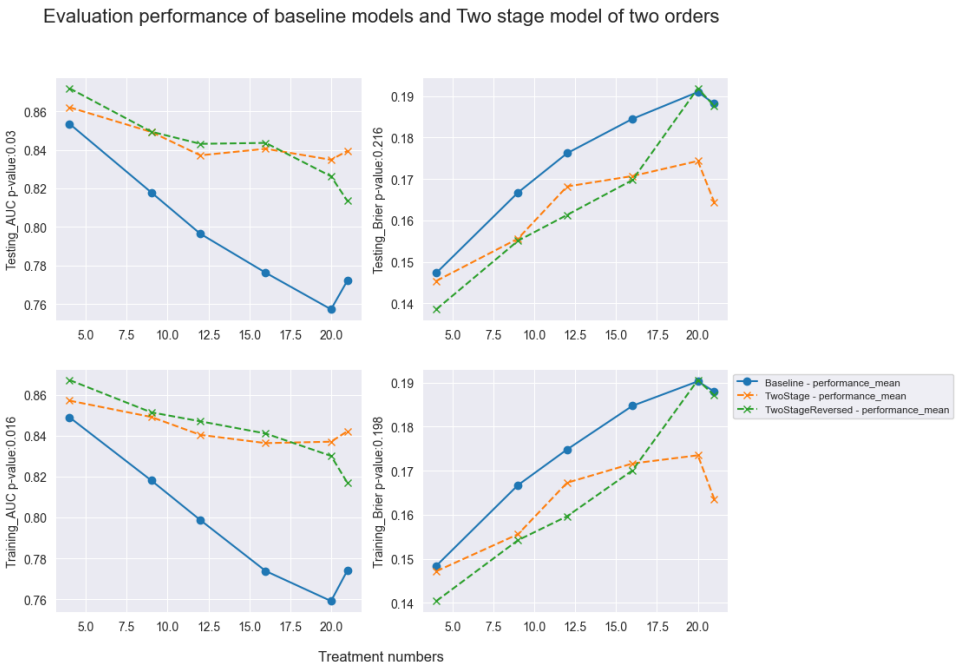


Figure A.12: Mean of model performance across all treatment options, for Additive Treatment effects with NO interactions terms

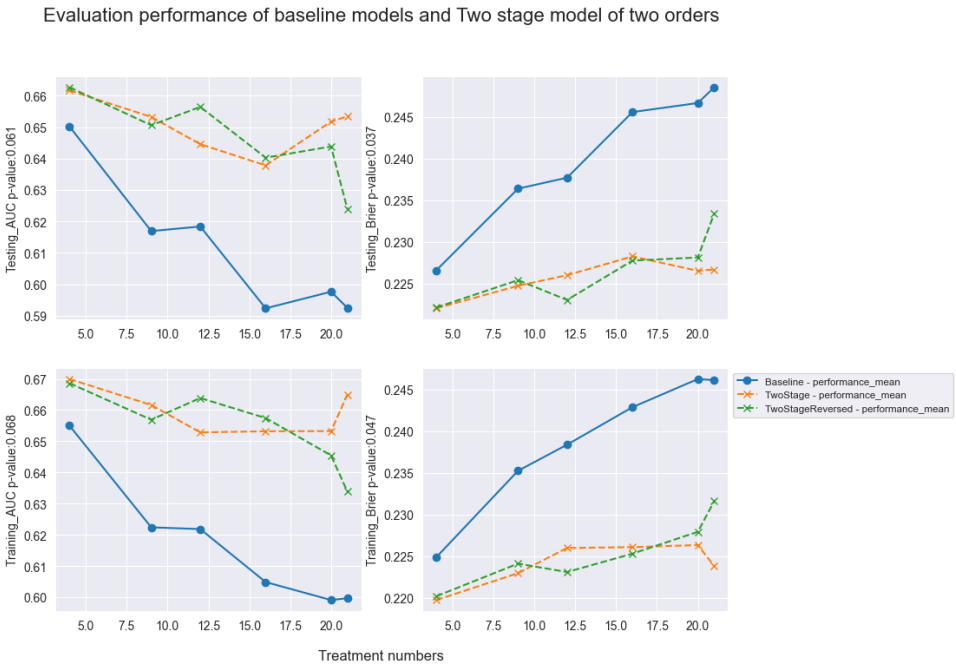


Figure A.13: Mean of model performance across all treatment options, for Additive Treatment effects with interactions terms

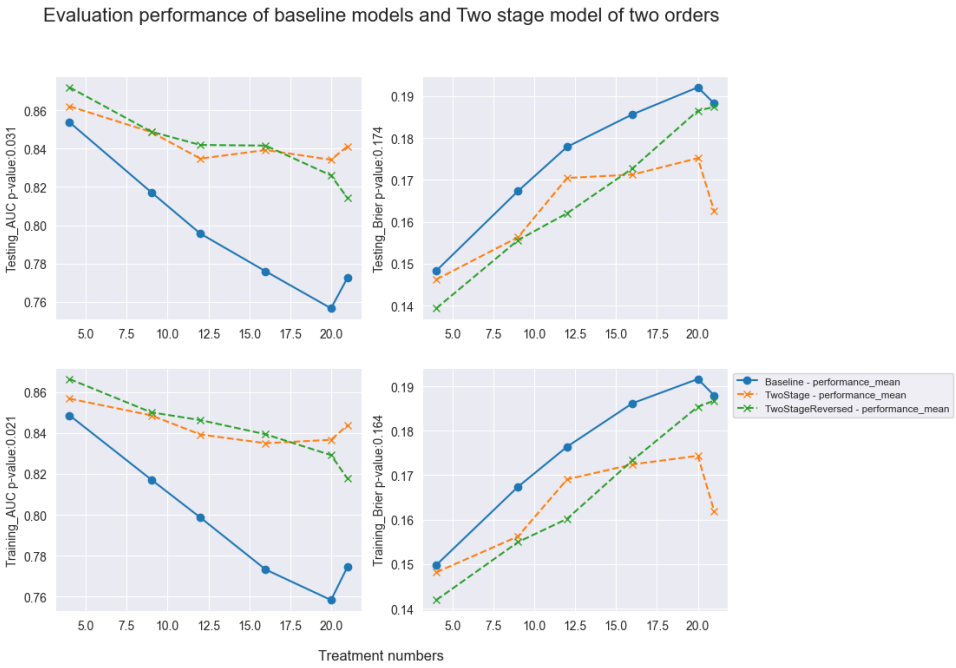


Figure A.14: Mean of model performance across all treatment options, for Overlapped Treatment effects with NO interactions terms

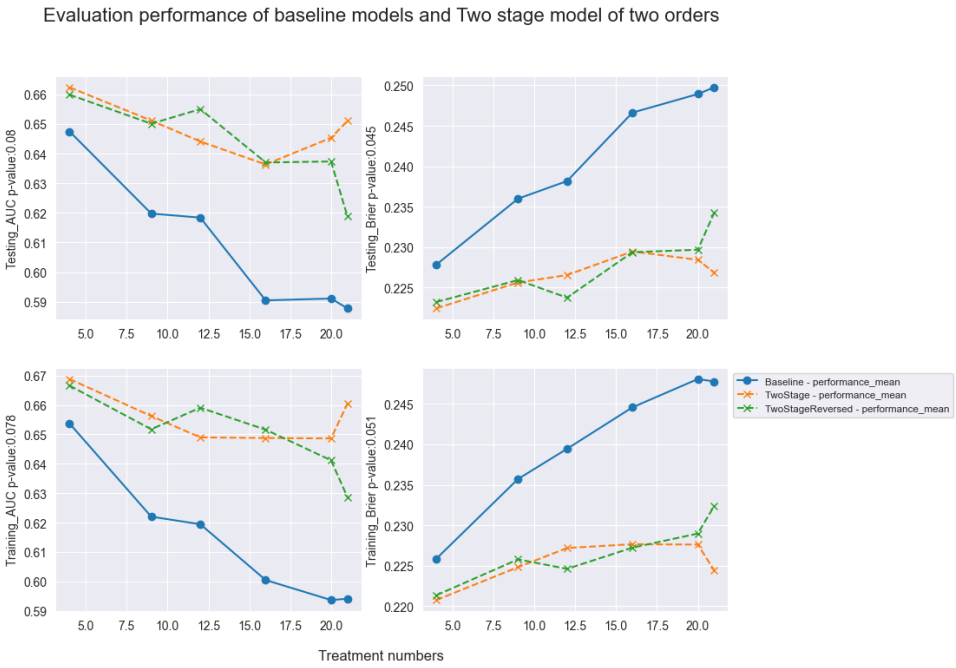


Figure A.15: Mean of model performance across all treatment options, for Overlapped Treatment effects with interactions terms

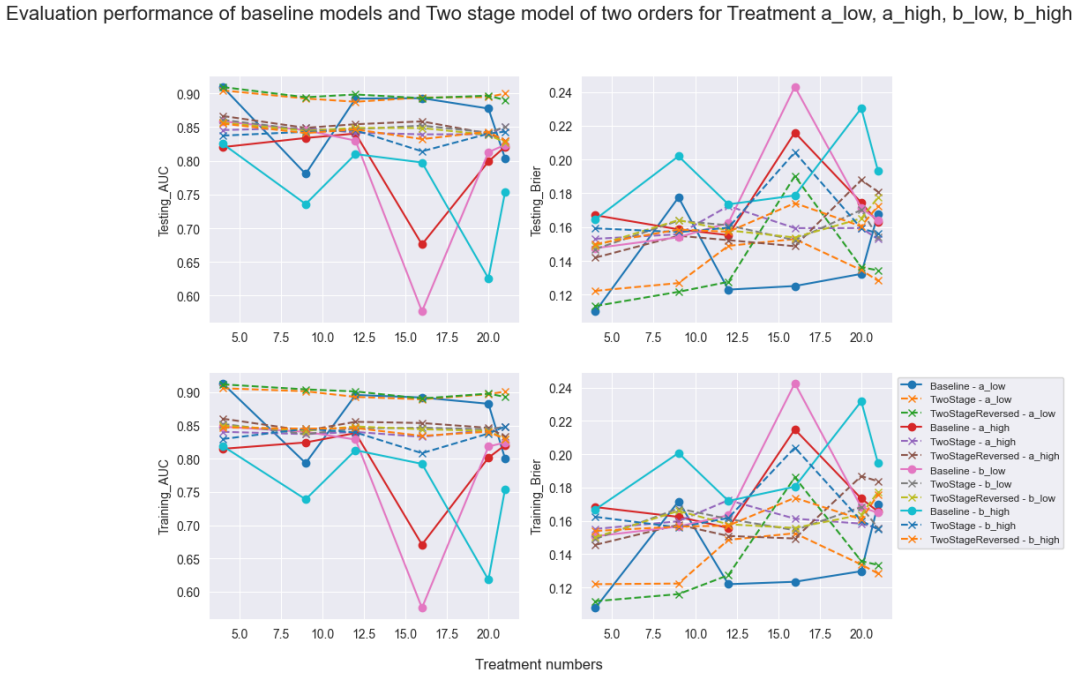


Figure A.16: 4 treatments examples of model performance across all treatment options, for Additive Treatment effects with No interactions terms

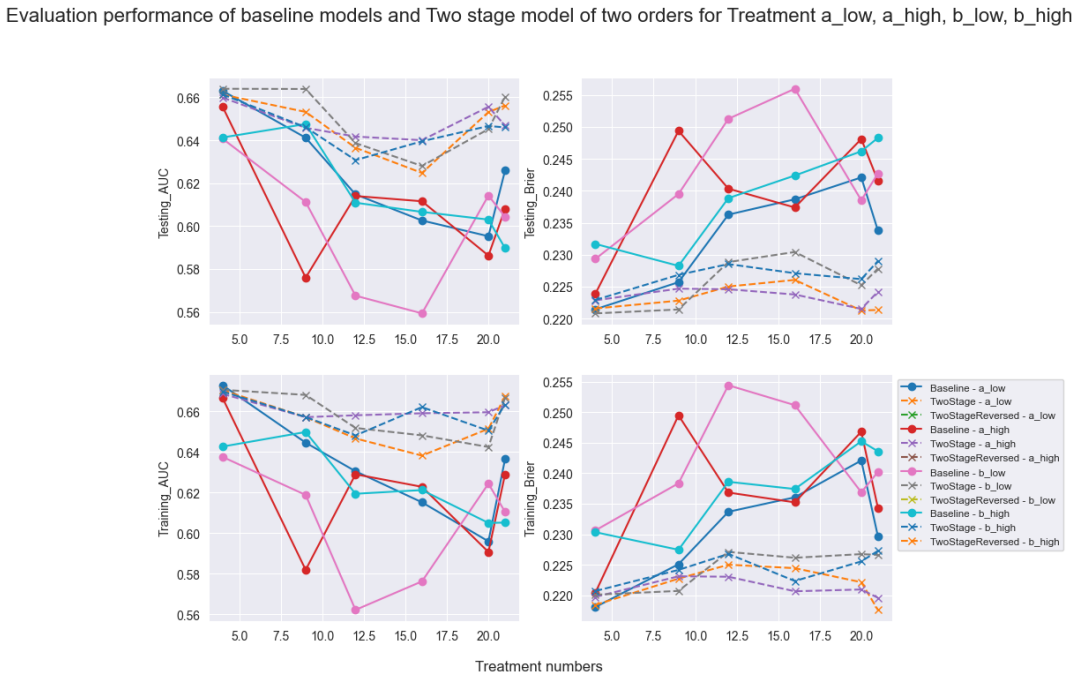


Figure A.17: 4 treatments examples of model performance across all treatment options, for Additive Treatment effects with interactions terms

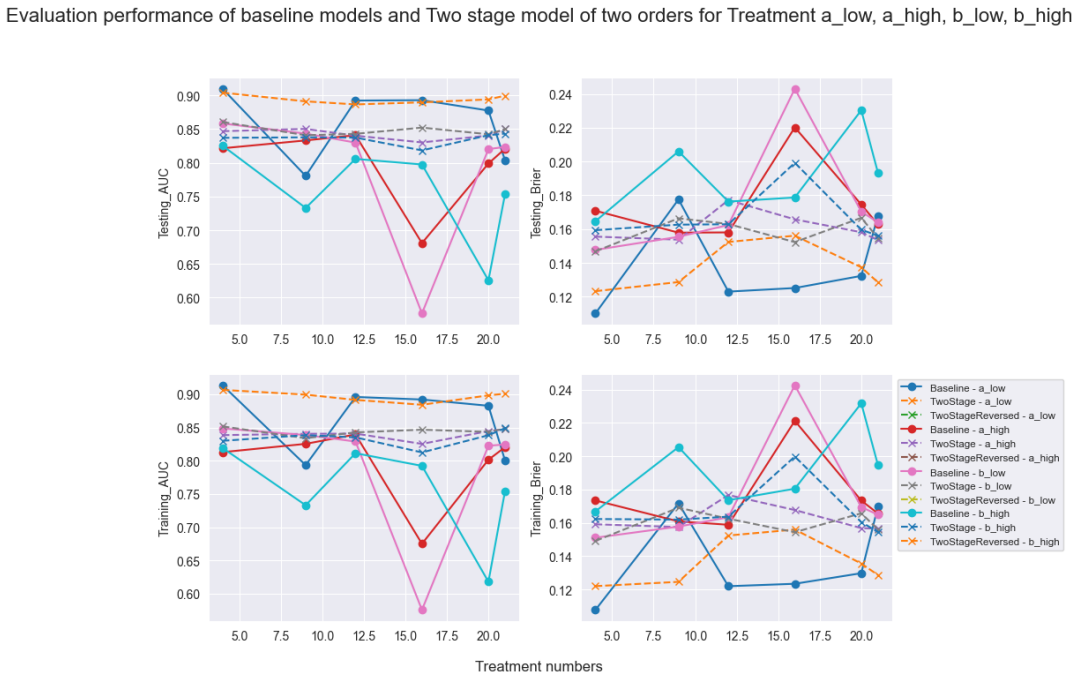


Figure A.18: 4 treatments examples of model performance across all treatment options, for Overlapped Treatment effects with No interactions terms

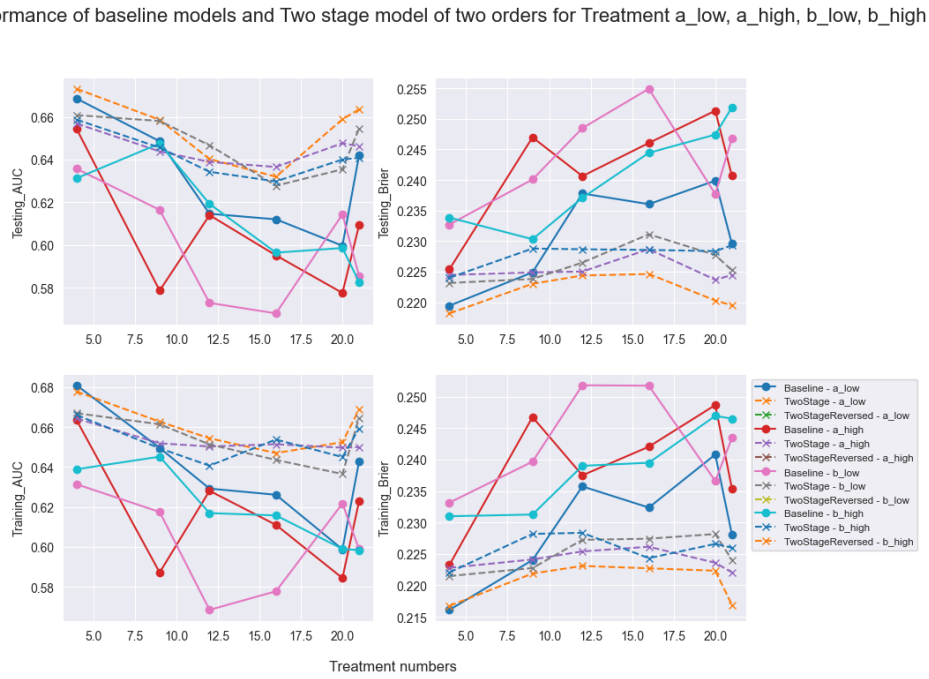


Figure A.19: 4 treatments examples of model performance across all treatment options, for Overlapped Treatment effects with interactions terms

A.3 Algorithms

A.3.1 Algorithm: Data Simulation for Additive Treatment Effects (No Interaction Terms)

Algorithm 6 Data Simulation for Additive Treatment Effects

Input: Number of patients N, Number of features p, Number of treatments T, Number of confounders C

Output: Simulated dataset with additive treatment effects

Set random seed: seed $\leftarrow 2$

Initialize simulation parameters:

$p \leftarrow 186$ ▷ Number of patient features

$C \leftarrow 100$ ▷ Number of confounders

Step 1: Generate Patient Features and Confounders

Generate patient feature matrix:

$$\mathbf{X} \sim \mathcal{N}(0,1) \in \mathbb{R}^{N \times p}$$

Generate confounders matrix:

$$\mathbf{Z} \sim \mathcal{N}(0,1) \in \mathbb{R}^{N \times C}$$

Step 2: Generate Treatment Assignment

Brainstore some coefficients

$$\boldsymbol{\beta} \in \mathbb{R}^{C \times T}$$

Calculate logits:

$$\text{logits} = \mathbf{Z} \cdot \boldsymbol{\beta}$$

Apply softmax to calculate treatment probabilities:

$$\mathbf{P} = \frac{\exp(\text{logits})}{\sum \exp(\text{logits})}$$

Assign treatments:

$$\text{Labels} \leftarrow \text{Sample based on } \mathbf{P}$$

Step 3: Define Treatment Effects

Set base treatment effect:

$$\tau \leftarrow 0.015$$

Define treatment effects for each treatment level:

$$\boldsymbol{\tau} = \tau \times (1, 2, \ldots, T)$$

Step 4: Calculate Baseline Outcomes

Generate outcome model coefficients:

$$\boldsymbol{\gamma} \in \mathbb{R}^{p+C+1}$$

Combine features, confounders, and treatment labels:

$$\mathbf{XZL} = [\mathbf{X}, \mathbf{Z}, \text{Labels}]$$

Calculate baseline logits:

$$\text{Baseline\_logits} = \mathbf{XZL} \cdot \boldsymbol{\gamma}$$

Step 5: Simulate Potential Outcomes

for *each treatment t* **do**

Calculate outcome probabilities:

$$\text{Outcome\_probabilities} = \sigma\left(\text{Baseline\_logits} + \mathcal{N}(0, 0.1)\right)$$

Adjust probabilities if necessary and simulate outcomes using Bernoulli trials

Step 6: Simulate Potential Outcomes

for *each treatment t* **do**

Calculate outcome probabilities:

$$p^{t}_{outcome} \leftarrow \sigma\left(\text{Baseline\_logits} + \mathcal{N}(0, 0.1)\right)$$

Scale the probabilities:

$$p^{t'}_{outcome} \leftarrow \text{ifelse}(p_{outcome} > 0.7 - \boldsymbol{\tau}[t], p_{outcome}, \text{scale\_proba}(p_{outcome}, \min = 0.00001, \max = 1 - (0.7 - \boldsymbol{\tau}[t])))$$

Sample the outcome variable $Y^{[t]}$

$$Y^{[t]} \leftarrow \text{bernoulli.rvs}(p^{t'}_{outcome})$$

Step 7: Use consistency to select the observed outcomes Map observed outcomes to treatments assigned to ensure consistency.

Step 8: Split Data into Training and Testing Sets

Split data into training (80%) and testing (20%) sets.

A.3.2 Algorithm: Data Simulation for Overlapped Treatment Effects (No Interaction Terms)

Algorithm 7 Data Simulation for Overlapped Treatment Effects Without Interaction Terms

Input: Number of patients N, Number of features p, Number of treatments T, Number of confounders C

Output: Simulated dataset with additive treatment effects

Set random seed: seed $\leftarrow 2$

Initialize simulation parameters:

$p \leftarrow 186$ ▷ Number of patient features

$C \leftarrow 100$ ▷ Number of confounders

Step 1: Generate Patient Features and Confounders

Generate patient feature matrix:

$$\mathbf{X} \sim \mathcal{N}(0,1) \in \mathbb{R}^{N \times p}$$

Generate confounders matrix:

$$\mathbf{Z} \sim \mathcal{N}(0,1) \in \mathbb{R}^{N \times C}$$

Step 2: Generate Treatment Assignment

Generate random coefficients for confounders and treatments:

$$\boldsymbol{\beta} \sim \mathcal{N}(0,0.1) \in \mathbb{R}^{C \times T}$$

Calculate logits:

$$\text{logits} = \mathbf{Z} \cdot \boldsymbol{\beta}$$

Apply softmax to calculate treatment probabilities:

$$\mathbf{P} = \frac{\exp(\text{logits})}{\sum \exp(\text{logits})}$$

Assign treatments:

$$\text{Labels} \leftarrow \text{Sample based on } \mathbf{P}$$

Step 3: Define Treatment Effects

Set base treatment effect:

$$\tau \leftarrow 0.015$$

Define treatment effects for each treatment level:

$$\boldsymbol{\tau} = \tau \times (1,2,\ldots,T)$$

Randomly select treatment pairs for overlapping:

Create overlap dictionary overlap_effects with selected pairs

Adjust treatment effects for overlap:

$$\boldsymbol{\tau}' = \text{Adjust } \boldsymbol{\tau} \text{ based on overlap effects}$$

For example:

$$\boldsymbol{\tau}' = \tau \times (1,2,2,4,4,\ldots,T)$$

Step 4: Calculate Baseline Outcomes

Generate outcome model coefficients:

$$\boldsymbol{\gamma} \sim \mathcal{N}(0, 0.1)^{p+C+1}$$

Combine features, confounders, and treatment labels:

$$\mathbf{XZL} = [\mathbf{X}, \mathbf{Z}, \text{Labels}]$$

Calculate baseline logits:

$$\text{Baseline\_logits} = \mathbf{XZL} \cdot \boldsymbol{\gamma}$$

Step 5: Simulate Potential Outcomes for *each treatment t* **do**

Calculate outcome probabilities:

$$\text{Outcome\_probabilities} = \sigma\left(\text{Baseline\_logits} + \mathcal{N}(0, 0.1)\right)$$

Adjust probabilities if necessary and simulate outcomes using Bernoulli trials

Step 6: Simulate Potential Outcomes

for *each treatment t* **do**

Calculate outcome probabilities:

$$p^{t}_{outcome} \leftarrow \sigma\left(\text{Baseline\_logits} + \mathcal{N}(0, 0.1)\right)$$

Scale the probabilities:

$$p^{t'}_{outcome} \leftarrow \text{ifelse}(p_{outcome} > 0.7 - \boldsymbol{\tau}'[t], p_{outcome}, \text{scale\_proba}(p_{outcome}, \min = 0.00001, \max = 1 - (0.7 - \boldsymbol{\tau}'[t])))$$

Sample the outcome variable $Y^{[t]}$

$$Y^{[t]} \leftarrow \text{bernoulli.rvs}(p^{t'}_{outcome})$$

Step 7: Use consistency to select the observed outcomes Map observed outcomes to treatments assigned to ensure consistency.

Step 8: Split Data into Training and Testing Sets

Split data into training (80%) and testing (20%) sets.

A.3.3 Algorithm: Data Simulation for Overlapped Treatment Effects with Interaction Terms

Algorithm 8 Data Simulation for Overlapped Treatment Effects with Interaction Terms

Input: Number of patients N, Number of features p, Number of treatments T, Number of confounders C, Number of polynomial features p_{poly}, Number of polynomial confounder features C_{poly}

Output: Simulated dataset with overlapped treatment effects and interaction terms

Set random seed: seed $\leftarrow 2$

Initialize simulation parameters:

$p \leftarrow 186$ ▷ Number of patient features

$C \leftarrow 100$ ▷ Number of confounders

$p_{poly} \leftarrow 50$ ▷ Number of polynomial features for patient data

$C_{poly} \leftarrow 30$ ▷ Number of polynomial features for confounders

Step 1: Generate Patient Features and Confounders with Polynomial Features

Generate patient feature matrix:

$$\mathbf{X} \sim \mathcal{N}(0,1) \in \mathbb{R}^{N \times p}$$

Select indices for polynomial features: $I_{poly} \leftarrow$ Randomly select p_{poly} indices from $\{1,\ldots,p\}$

Extract subset of features: $\mathbf{X}_{poly} \leftarrow \mathbf{X}_{[:,I_{poly}]}$

Generate polynomial features:

$$\mathbf{X}_{poly} = \text{PolynomialFeatures}(\mathbf{X}_{poly}, \text{degree} = 2)$$

Combine original and polynomial features:

$$\mathbf{X}_{final} = [\mathbf{X}_{[:,\text{remaining indices}]}, \mathbf{X}_{poly}]$$

Generate confounder matrix:

$$\mathbf{Z} \sim \mathcal{N}(0,1) \in \mathbb{R}^{N \times C}$$

Select indices for polynomial confounders: $I_{poly}^{conf} \leftarrow$ Randomly select C_{poly} indices from $\{1,\ldots,C\}$

Extract subset of confounders: $\mathbf{Z}_{poly} \leftarrow \mathbf{Z}_{[:,I_{poly}^{conf}]}$

Generate polynomial features for confounders:

$$\mathbf{Z}_{poly} = \text{PolynomialFeatures}(\mathbf{Z}_{poly}, \text{degree} = 2)$$

Combine original and polynomial confounders:

$$\mathbf{Z}_{final} = [\mathbf{Z}_{[:,\text{remaining indices}]}, \mathbf{Z}_{poly}] \in \mathbb{R}^{N \times C_{final}}$$

Step 2: Generate Treatment Assignment

Generate random coefficients for confounders and treatments:

$$\boldsymbol{\beta}_{\text{conf}} \sim \mathcal{N}(0, 0.1) \in \mathbb{R}^{C_{final} \times T}$$

Calculate logits:

$$\text{logits} = \mathbf{Z}_{final} \cdot \boldsymbol{\beta}_{\text{conf}}$$

Apply softmax to calculate treatment probabilities:

$$\mathbf{P} = \frac{\exp(\text{logits})}{\sum \exp(\text{logits})}$$

Assign treatments:

$$\text{Labels} \leftarrow \text{Sample based on } \mathbf{P}$$

Step 3: Define Treatment Effects with Overlaps

Set base treatment effect:

$$\tau \leftarrow 0.015$$

Define initial treatment effects:

$$\boldsymbol{\tau} = \tau \times (1, 2, \ldots, T)$$

Randomly select treatment pairs for overlapping:

Create overlap dictionary overlap_effects with selected pairs

Adjust treatment effects for overlap:

$$\boldsymbol{\tau}' = \text{Adjust } \boldsymbol{\tau} \text{ based on overlap effects}$$

For example:

$$\boldsymbol{\tau}' = \tau \times (1, 2, 2, 4, 4, \ldots, T)$$

Step 4: Calculate Baseline Outcomes with Interaction Terms

Generate outcome model coefficients:

$$\boldsymbol{\gamma} \sim \mathcal{N}(0, 0.1) \in \mathbb{R}^{p_{final} + C_{final} + 1}$$

Combine features, confounders, and treatment labels:

$$\mathbf{XZL} = [\mathbf{X}_{final}, \mathbf{Z}_{final}, \text{Labels}]$$

Calculate baseline logits with interaction terms:

$$\text{Baseline\_logits} = \mathbf{XZL} \cdot \boldsymbol{\gamma}$$

Step 5: Simulate Potential Outcomes

for *each treatment t* **do**

Calculate outcome probabilities:

$$\text{Outcome\_probabilities} = \sigma\left(\text{Baseline\_logits} + \mathcal{N}(0, 0.1)\right)$$

Adjust probabilities if necessary and simulate outcomes using Bernoulli trials

Step 6: Simulate Potential Outcomes

for *each treatment t* **do**

Calculate outcome probabilities:

$$p^{t}_{outcome} \leftarrow \sigma\left(\text{Baseline\_logits} + \mathcal{N}(0, 0.1)\right)$$

Scale the probabilities:

$$p^{t'}_{outcome} \leftarrow \text{ifelse}(p_{outcome} > 0.7 - \boldsymbol{\tau}'[t], p_{outcome}, \text{scale\_proba}(p_{outcome}, \min = 0.00001, \max = 1 - (0.7 - \boldsymbol{\tau}'[t])))$$

Sample the outcome variable $Y^{[t]}$

$$Y^{[t]} \leftarrow \text{bernoulli.rvs}(p^{t'}_{outcome})$$

Step 7: Use consistency to select the observed outcomes Map observed outcomes to treatments assigned to ensure consistency.

Step 8: Split Data into Training and Testing Sets

Split data into training (80%) and testing (20%) sets.