Data-Driven Decision Making in Healthcare

by

Wesley J. Marrero

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Industrial and Operations Engineering)
in the University of Michigan
2021

Doctoral Committee:

Associate Professor Mariel S. Lavieri, Chair
Associate Professor Eunshin Byon
Professor Rodney A. Hayward
Associate Professor David W. Hutton
Assistant Professor Neehar Parikh
Associate Professor Ambuj Tewari

Wesley J. Marrero
wmarrero@umich.edu
ORCID iD: 0000-0002-7092-2292

Dedication

This dissertation is dedicated to my parents and sister.

Acknowledgments

I would like to acknowledge the people who provided their support throughout my time working on this dissertation.

First, I express my enormous gratitude to my advisor, Mariel Lavieri. Her guidance and support have helped me grow tremendously as a researcher and will continue to shape my academic future. It is my hope that my future students gain as much from me as I have gained from her. My achievements and success over the course of my Ph.D. reflect her dedication as an advisor and mentor.

Second, I would like to thank those who have collaborated with me and provided feedback on the work in this dissertation. I am grateful for the guidance and support from Eunshin Byon, Rodney Hayward, David Hutton, Jeremy Sussman, Neehar Parikh, and Ambuj Tewari. Your contributions on this research have been invaluable. I am inspired by the rigor you bring to every one of your projects.

I am grateful for the many mentors that I have had during my undergraduate and graduate education. I am incredibly appreciative of my undergraduate mentors, Marlio Paredes, José Santivañez, Jannette Perez, and Martha Centeno, who encouraged me to pursue a Ph.D. and helped me in my application process to doctoral programs. During my time in graduate school, I have been extremely fortunate to be surrounded by a supportive group of faculty members. I especially want to thank Brian Denton, Marina Epelman, Larry Seiford, and Monroe Keyserling for providing their support, advice, and guidance throughout my graduate studies.

I also want to thank my friends and family. My friends have made my time at Michigan much more enjoyable that I could have ever imagined. I attribute much of my personal growth in the past few years to the quality of the friendships I have developed throughout graduate school. My family has been a constant source of support and motivation over the course of my Ph.D. Thanks mamá, papá, and Win for your unconditional love, support, and encouragement. I am especially grateful for my partner, Amy Newman. Thanks for showing me your love and support on every step of this journey.

Lastly, I am fortunate to have received financial support for my dissertation research. I want to acknowledge the Rackham Merit Fellowship and its community. This dissertation is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE 1256260.

Table of Contents

Dedication . . . . . **ii**

Acknowledgments . . . . . **iii**

List of Figures . . . . . **viii**

List of Tables . . . . . **xiv**

List of Appendices . . . . . **xv**

List of Abbreviations . . . . . **xvi**

Abstract . . . . . **xxi**

Chapter

1 Introduction **1**

1.1 Using Population-Level Data to Model the Supply and Demand of Organs for Transplantation 1
1.2 Using Patient-Level Data for the Management of Cardiovascular Diseases 2
1.3 Organization and Contributions 3
1.3.1 Part 1: Using Population-Level Data to Model the Supply and Demand of Organs for Transplantation 5
1.3.2 Part 2: Using Patient-Level Data for the Management of Cardiovascular Diseases 5

2 Modeling Supply, Demand, and Allocation in Liver Transplantation **8**

2.1 Background 8
2.2 Organization of the Chapter 9
2.3 Literature Review 9
2.3.1 Predictive Models for Organ Supply and Demand 10
2.3.2 Application of Operations Research Models to Organ Transplantation 11
2.4 Modeling Liver Supply 12
2.4.1 Data Sources 12
2.4.2 Projection Development 13
2.4.3 Sensitivity Analysis 13
2.4.4 Monte Carlo Simulation 14

2.4.5 Liver Supply Projections Results . . . 14
2.4.6 Discussion . . . 17
2.5 Modeling Liver Demand . . . 18
2.5.1 Data Sources . . . 19
2.5.2 Model Selection . . . 20
2.5.3 Projections and Stochastic Simulation . . . 21
2.5.4 Model Accuracy Measures . . . 22
2.5.5 Liver Demand Projection Results . . . 23
2.5.6 Discussion . . . 26
2.6 Modeling Liver Supply in Allocation Models . . . 27
2.6.1 Liver Donor Projections by District Model . . . 28
2.6.2 Exploratory Analysis of Geographic Inequity . . . 30
2.6.3 Liver Allocation Results . . . 30
2.6.4 Discussion . . . 32
2.7 Improving the Prediction of Deceased Donor Organ Yield . . . 35
2.7.1 Methods . . . 36
2.7.2 Results . . . 39
2.7.3 Discussion . . . 49
2.8 Conclusions . . . 52

3 Understanding the Role of Genetic Information in Cardiovascular Treatment Planning . . . 54

3.1 Background . . . 54
3.2 Organization of the Chapter . . . 56
3.3 Literature review . . . 56
3.4 Cost-Effectiveness of Risk Threshold-Based Treatment Plans Informed with Genetic Information . . . 59
3.4.1 Methods . . . 59
3.4.2 Results . . . 66
3.4.3 Discussion . . . 72
3.5 Cost-Effectiveness of Optimal Cholesterol Treatment Plans Informed with Genetic Information . . . 73
3.5.1 Simulation Framework . . . 74
3.5.2 Results . . . 79
3.5.3 Discussion . . . 82
3.6 Optimal Cholesterol Treatment and Genetic Testing Strategies . . . 83
3.6.1 Modeling Framework . . . 84
3.6.2 Simulation Framework . . . 93
3.6.3 Results . . . 95
3.6.4 Discussion . . . 104
3.7 Conclusions . . . 106

4 Data-Driven Ranges of Near-Optimal Actions for Finite Markov Decision Processes . . . 108

4.1 Background . . . 108
4.1.1 Applications of Markov Decision Processes to Medical Decision-Making . . . 109
4.1.2 Modeling Approach . . . 111
4.2 Organization of the Chapter . . . 111
4.3 Literature Review . . . 112
4.4 Preliminaries . . . 114
4.4.1 Markov Decision Processes . . . 114
4.4.2 Multiple Comparisons With a Control . . . 116
4.5 Ranges of Near-Optimal Actions . . . 117
4.6 Solution Approach . . . 117
4.6.1 Simulation-Based Backwards Induction . . . 119
4.6.2 Simulation-Based Parallel Multiple Comparisons with a Control . 123
4.7 Case study: Personalized Hypertension Treatment Plans . . . 130
4.7.1 Background on Hypertension Treatment . . . 130
4.7.2 Markov Decision Process Formulation . . . 130
4.7.3 Data Source . . . 133
4.7.4 Ordering of States and Actions . . . 135
4.7.5 Simulation Framework . . . 135
4.7.6 Selection of Number of Batches and Observations per Batch . . . 136
4.7.7 Analysis . . . 137
4.7.8 Sensitivity Analyses . . . 137
4.7.9 Numerical Results . . . 138
4.8 Discussion . . . 146
4.9 Conclusions . . . 148

5 Conclusions and Future Work . . . 149

5.1 Future Work: Evaluating the Impact of Interventions to Narrow the Gap Between Organ Supply and Demand . . . 150
5.2 Future Work: Understanding the Role of Improvements in the Quality of Information in Genetic Testing . . . 151
5.3 Future Work: Flexible Decision Support Under Input Data Uncertainty . . . 152
5.4 Future Work: Synergies Across Chapters . . . 153
5.5 Final Remarks . . . 154

Appendices . . . 155

Bibliography . . . 230

List of Figures

1.1 Organization of dissertation and connections between chapters. 4

2.1 Organization and connections between sections in Chapter 2. 10
2.2 Projections of liver organ availability. 15
2.3 Results of sensitivity analyses. 16
2.4 Monte Carlo simulation of liver availability. 17
2.5 Relationship between NASH additions to the waiting list and obese population in the US using a 9-year lag. 23
2.6 Point estimates and prediction intervals of NASH additions to the waiting list from 2016 to 2030. 24
2.7 Point estimates and prediction intervals of NASH additions to the waiting list from 2024 to 2030 given plateau in obesity prevalence. 25
2.8 Liver allocation systems. 29
2.9 Projected National Changes in D/100K US Population and Liver Donors Available from 2016-2025. 31
2.10 Projected Change in Regional Variation in D/100K US Population in each Allocation Model from 2016 to 2025. 31
2.11 MAE of potential models to predict overall deceased donor organ yield. Points in the center of the boxes represent the mean MAE of the statistical models and asterisks represent outliers. The MAE of the ANN models are excluded for illustration purposes. ANN presented a median MAE of 2.766 with an interquartile range of 0.852. 44
2.12 MSE of potential models to predict overall deceased donor organ yield. Points in the center of the boxes represent the mean MSE of the statistical models and asterisks represent outliers. The MSE of the ANN models are excluded for illustration purposes. ANN presented a median MSE of 11.449 with an interquartile range of 5.531. 44
2.13 Partial dependence plots of the main categorical predictors of overall deceased donor organ yield. 46
2.14 Partial dependence plots of the main numerical predictors of deceased donor organ yield. Points in the horizontal axis of the plots represent the empirical quartiles of each predictor. 47

2.15 MAE of potential models to predict overall deceased donor organ yield using imputed dataset. Points in the center of the boxes represent the mean MAE of the statistical models and asterisks represent outliers. The MAE of the ANN models are excluded for illustration purposes. ANN presented a median MAE of 2.670 with an interquartile range of 1.028. 48
2.16 MSE of potential models to predict overall deceased donor organ yield using imputed dataset. Points in the center of the boxes represent the mean MSE of the statistical models and asterisks represent outliers. The MSE of the ANN models are excluded for illustration purposes. ANN presented a median MSE of 9.471 with an interquartile range of 8.206. 48
2.17 MAE of potential models to predict overall deceased donor organ yield using imputed dataset. Points in the center of the boxes represent the mean MAE of the statistical models and asterisks represent outliers. The MAE of the ANN models are excluded for illustration purposes. The median (interquartile range) MAE of the ANN models in cohorts 2013-2014, 2015-2016, and 2017-2018 are: 2.431 (0.615), 2.748 (0.719), and 2.398 (0.431), respectively. 49
2.18 MSE of potential models to predict overall deceased donor organ yield using imputed dataset. Points in the center of the boxes represent the mean MSE of the statistical models and asterisks represent outliers. The MSE of the ANN models are excluded for illustration purposes. The median (interquartile range) MSE of the ANN models in cohorts 2013-2014, 2015-2016, and 2017-2018 are: 8.126 (3.371), 9.971 (4.611), and 7.957 (2.483), respectively. 50

3.1 Organization and connections between sections in Chapter 3. 56
3.2 Flow diagram of the study dataset. 61
3.3 ICER of GenePCE (compared to the PCE strategy) under different number of replications at a genetic testing costs of $200 per test. Points represent the mean ICER using from 1 to 1,500 replications. 66
3.4 Comparison of PCE and GenePCE risks. The population size is represented with the color gradient (dark blue indicates the smallest population sizes and dark orange indicates the biggest population sizes). The green long dashed vertical lines indicate the population within the 1 SD genetic testing range. The blue short dashed lines represent the treatment thresholds. People with GenePCE risk scores above the horizontal blue short dashed line are treated according to the GenePCE strategy. People with PCE risk scores to the right of the vertical blue short dashed line are treated according to the PCE strategy. A shows the risk scores for the whole population of 93.6 million people (after LDL exclusions). B shows the risk scores of the people that received a genetic test during the first year of our study (16.17 million people). 68
3.5 ICER of GenePCE (compared to the PCE strategy) using multiple genetic testing scenarios and costs of genetic testing. Points indicate ICER of GenePCE at genetic testing costs of $0, $50, $100, $200, $400, $800, and $2,000 per test. . . 72
3.6 Summary of simulation framework for a single patient. 74

3.7 Convergence of QALYs saved and cost incurred due to genetic testing in simulation. Figure 3.7(a) illustrates the average QALYs saved due to genetic testing using from 1 to 1,500 replications. Points in Figure 3.7(b) represent the average cost incurred due to genetic testing using from 1 to 1,500 replications. The red horizontal lines in plots show the average QALYs saved and cost incurred at 2,000 replications. 80
3.8 Comparison of PCE and GenePCE risk scores. The population size is represented with the color gradient (dark blue indicates the smallest population sizes and dark orange indicates the biggest population sizes). Figure 3.8(a) shows the risk scores of the whole population (93.55 million people). Figure 3.8(b) shows the risk scores of the patients that receive a genetic test during the first year of our study (16.17 million people). 81
3.9 ICER of genetic information in different populations and costs of genetic testing. Points indicate the ICER of genetic information at genetic testing costs of $50, $100, $200, and $400 per test. The base population represents performing genetic testing according to clinical expertise and the USPSTF statin guidelines. A base-10 logarithmic scale is used in the vertical axis for illustration purposes. Minor tick marks in the vertical axis represent 10% increases between the major tick marks. 83
3.10 Summary of modeling framework for a single patient. The index g represents a realization of genetic information, t represents the year in the planning horizon, and k represents a realization of the health trajectory of the patient. . . 85
3.11 Distribution of the initial risk for ASCVD events of patients that receive genetic testing by age group. 97
3.12 Distribution of VoI of genetic testing with respect to the initial risk for ASCVD events by age group in the population that received genetic testing. Each point represents 5,000 patients. Smoothed line are obtained using second degree local regression with a span of 90%. Shaded areas around the smoothed lines represent 95% confidence intervals around the mean. 98
3.13 Distribution of the expected year of genetic testing with respect to the initial risk for ASCVD events by age group. Each point represents 5,000 patients. Smoothed line are obtained using second degree local regression with a span of 90%. Shaded areas around the smoothed lines represent 95% confidence intervals around the mean. 99
3.14 Comparison of PCE and GenePCE risk scores. The population size is represented with the color gradient (the darker the color the bigger the population). Equal risk scores up to the fifth decimal point are compressed into single point. 100
3.15 Proportion of patients receiving each treatment intensity by treatment policy over the planning horizon. Points represent the ratio of average number of patients treated with each statin intensity and the total population that receives genetic testing. 101
3.16 Total cumulative cost-savings due to genetic testing. Points represent the total cumulative cost savings by each year. 102

4.1 Summary of solution approach analysis. Font in italics gives the assumption number each result relies on followed by a short description of the assumption. Boldface font gives the result number in the chapter followed by a brief description of the result. A result uses an assumption if it is placed below the box of the assumption. We divide our theoretical results into three categories: finite sample properties (underlined with a wavy line), general asymptotic properties (underlined with a solid line), and asymptotic structural properties (underlined with a dashed line). 118
4.2 Summary of simulation framework for a single patient. The index t represents represents the year in the planning horizon (10 years). 136
4.3 Ranges of near-optimal treatment choices per patient profile. The ranges are highlighted with the gray shaded area in each profile. The labels "SD" and "HD" denote antihypertensive medications at standard dosage and half dosage, respectively. 139
4.4 Distribution of treatment at year 1 and year 10 of the study. BP categories are made based on patients' characteristics at year 1. 141
4.5 Life-years saved by each treatment policy compared to no treatment per BP group over the planning horizon. 143

B.1 Distribution of risk scores over the 10-year planning horizon (only includes population that received genetic testing). 161
B.2 Convergence of QALYs and cost saved in our simulation over the number of health trajectory replications under a single GRS realization per patient. Red line represents the QALYs and cost saved at 2,000 health trajectory replications. 164
B.3 Convergence of QALYs and cost saved in our simulation over the number of GRS realizations with a fixed number of heath trajectory replications. Red line represents the QALYs and cost saved at 500 GRS realizations and 500 health trajectory replications. 165
B.4 Convergence of QALYs and cost saved in our simulation over the number of health trajectory replications under 100 GRS realizations per patient. Red line represents the QALYs and cost saved at 750 health trajectory replications and 100 GRS realizations. 165
B.5 Testing cost sensitivity analysis results by age group. Shaded boxplots represent the base case, asterisks represent outliers, and points in the center of the boxes represent the average testing year. 166
B.6 Treatment cost sensitivity analysis results by age group. Shaded boxplots represent the base case, asterisks represent outliers, and points in the center of the boxes represent the average testing year. 166
B.7 Treatment disutility sensitivity analysis results by age group. Shaded boxplots represent the base case, asterisks represent outliers, and points in the center of the boxes represent the average testing year. 167

C.1 Convergence of confidence interval width over the number of batches. Red line represents the confidence interval width using 1,000 batches (0.01). 219

C.2 Number of people by race, race, and BP category. BP groups are consistent with the BP categories of the 2017 Hypertension Clinical Practice Guidelines. The label "Elevated" denotes elevated BP, "Stage 1" denotes stage 1 hypertension, and the label "Stage 2" denotes stage 2 hypertension. 220
C.3 Distribution of treatment at year 1 and year 10 of the study by sex. BP groups are consistent with the BP categories of the 2017 Hypertension Clinical Practice Guidelines. The label "Elevated" denotes elevated BP, "Stage 1" denotes stage 1 hypertension, and the label "Stage 2" denotes stage 2 hypertension. 221
C.4 Distribution of treatment at year 1 and year 10 of the study by race. BP groups are consistent with the BP categories of the 2017 Hypertension Clinical Practice Guidelines. The label "Elevated" denotes elevated BP, "Stage 1" denotes stage 1 hypertension, and the label "Stage 2" denotes stage 2 hypertension. 222
C.5 Proportion of patients of treatment recommendations made by clinical guidelines contained in the ranges of near-optimal actions. BP groups are consistent with the BP categories of the 2017 Hypertension Clinical Practice Guidelines. The label "Elevated" denotes elevated BP, "Stage 1" denotes stage 1 hypertension, and the label "Stage 2" denotes stage 2 hypertension. 223
C.6 Life-years saved by each treatment policy compared to no treatment per sex and BP group over the planning horizon. BP groups are consistent with the BP categories of the 2017 Hypertension Clinical Practice Guidelines. The label "Elevated" denotes elevated BP, "Stage 1" denotes stage 1 hypertension, and the label "Stage 2" denotes stage 2 hypertension. 224
C.7 Life-years saved by each treatment policy compared to no treatment per race and BP group over the planning horizon. BP groups are consistent with the BP categories of the 2017 Hypertension Clinical Practice Guidelines. The label "Elevated" denotes elevated BP, "Stage 1" denotes stage 1 hypertension, and the label "Stage 2" denotes stage 2 hypertension. 225
C.8 Average range width and number of medications in base case, assuming normality in the action-value functions, using the action that corresponds to the median number of medications in next year's range, and using the action to corresponds to the fewest number of medications in next year's range. The label "'Median in Next Year's Range" denotes the median number of medications in next year's range and the label "Fewest in Next Year's Range" denotes the fewest number of medications in next year's range. Dashed lines represent the 5^{th} and 95^{th} quantile across the population of adults in the US with ages between 50 and 54. 226
C.9 Life-years saved by each treatment policy compared to no treatment per sex and BP group over the planning horizon in secondary population of adults with ages between 70 and 74. BP groups are consistent with the BP categories of the 2017 Hypertension Clinical Practice Guidelines. The label "Elevated" denotes elevated BP, "Stage 1" denotes stage 1 hypertension, and the label "Stage 2" denotes stage 2 hypertension. 227

C.10 Life-years saved by each treatment policy compared to no treatment per race and BP group over the planning horizon in secondary population of adults with ages between 70 and 74. BP groups are consistent with the BP categories of the 2017 Hypertension Clinical Practice Guidelines. The label "Elevated" denotes elevated BP, "Stage 1" denotes stage 1 hypertension, and the label "Stage 2" denotes stage 2 hypertension. 228
C.11 Proportion of patients whose treatment recommendations are contained in the ranges of near-optimal actions despite parameter misestimation. Each panel represents a different misestimation scenario: no misestimation (top left), patients' true risk for ASCVD events is half the estimated risk (top right), patients' true risk for ASCVD events is double the estimated risk (bottom left), and patients' true benefit from treatment is half the estimated benefit (bottom right). 229

List of Tables

2.1 Projected population and donor growth from 2014 to 2025. 15
2.2 Performance under different time lags. 24
2.3 Effect of standardizing liver utilization rates, liver donation rates, and both on geographic variation in D/100K. 32
2.4 Summary of all the predictors of overall deceased donor organ yield considered in our study ($n = 89,520$). 39

3.1 Input variables. 64
3.2 Description of patients at first year of study. 67
3.3 Treatment reclassification of patients at first year of study. Counts are presented in millions with percentages in terms of the population that received genetic testing. 69
3.4 Population health outcomes over 10-year time horizon. Includes only people who receive genetic testing during the first year of the study with 10-year PCE risk between 7.68% and 13.63% (16.17 million patients). 70
3.5 Sensitivity analyses. Compares the results of the GenePCE strategy to the PCE approach. 71
3.6 Population health outcomes at the end of the 10-year time horizon. The policies informed with clinical information only and with clinical and genetic information are compared to no treatment. Results are averages over 750 replications. Values within parenthesis indicate the standard deviation across replications. 82
3.9 Sensitivity analysis parameter values. 95
3.10 Summary of sensitivity analyses. 103

4.3 Base case parameters 134
4.4 Sensitivity analysis parameter values. 137
4.5 Summary of sensitivity analyses at the first year of our study. 144

B.1 Linear regression models coefficients. 160
B.2 Summary of GenePCE treatment thresholds. 162
B.3 Impact of genetic risk on relative benefit of treatment. 163

C.1 Linear regression models coefficients. 217

List of Appendices

A Appendix for Modeling Supply, Demand and Allocation in Liver Transplantation 155

B Appendix for Understanding the Role of Risk in Optimal Cardiovascular Treatment Planning 160

C Appendix for Data-Driven Ranges of Near-Optimal Actions for Finite Markov Decision Processes 168

1 Simulation-based backwards induction (SBBI) algorithm. 169
2 Simulation-based multiple comparisons with a control (SBMCC) algorithm. . 170
3 Combined SBBI and SBMCC algorithm. 171

List of Abbreviations

ADP approximate dynamic programming

a.s. almost surely

ANN artificial neural networks

ARR absolute risk reduction

ASCVD atherosclerotic cardiovascular disease

AST aspartate aminotransferase

ALT alanine aminotransferase

BART Bayesian additive regression trees

BMI body mass index

BP blood pressure

BPLTTC Blood Pressure Lowering Treatment Trialists' Collaboration

CART classification and regression trees

CHD coronary heart disease

CI confidence interval

CMS Centers for Medicare and Medicaid Services

DBP diastolic blood pressure

DM decision maker

D/100K donors per 100,000 population

DSAs donor service areas

GAM general additive models

GLM generalized linear models

GRS genetic risk score

HCC hepatocellular carcinoma

HDL high-density lipoprotein

HHS United States Department of Health and Human Services

ICER incremental cost-effectiveness ratio

KDRI kidney donor risk index

LDL low-density lipoprotein

LSAM liver simulated allocation model

LT liver transplantation

MAE mean absolute error

MDP Markov decision process

MARS multivariate adaptive regression splines

MCC multiple comparisons with a control

MELD Model of End-stage Liver Disease

MICE multivariate imputation by chained equations

MPSC Membership Professional Standards Committee

MSE mean squared error

NASH non-alcoholic steatohepatitis

NAFLD non-alcoholic fatty liver disease

NB negative binomial

NCHS National Center for Health Statistics

NHANES National Health and Nutrition Examination Survey

NIH National Institutes of Health

OPOs organ procurement organizations

OPTN Organ Procurement and Transplantation Network

OR operations research

POMDP partially observable Markov decision process

PCE pooled cohort equations

pO2 partial pressure of oxygen

pO2/FiO2 the ratio of pO2 to the fraction of inspired oxygen

QALYs quality-adjusted life years

QoL quality of life

RL reinforcement learning

rMSE root mean squared error

RRR relative risk reduction

SBBI simulation-based backwards induction

SBDP simulation-based dynamic programming

SBMCC simulation-based multiple comparisons with a control

SBP systolic blood pressure

SD standard deviation

SRTR Scientific Registry of Transplant Recipients

TC total cholesterol

UNOS United Network for Organ Sharing

USPSTF US Preventive Services Task Force

VoI value of information

Abstract

The increasing availability of healthcare data has provided a great opportunity for the development of data-driven models to guide health policy and medical practice. The objective of this dissertation is to present new methods that use these data to make better healthcare decisions at a population and patient level. We first model the supply, demand, and allocation of organs for transplantation using data from the Organ Procurement and Transplantation Network and the US Census Bureau. Then, we introduce personalized treatment plans and genetic testing strategies for the management of cardiovascular diseases. We evaluate the clinical and policy implications of the treatment and testing strategies at a population level using data from the National Health and Nutrition Examination Survey. Lastly, we propose a modeling framework to consider physicians' judgment and patients' preferences in the implementation of treatment protocols. To illustrate how this method can be implemented in medical practice, we find ranges of near-optimal antihypertensive treatment choices for 16.72 million adults in the US. This research has the potential to improve healthcare practice by giving flexible and achievable guidelines to policymakers and medical professionals based on patient and population-level data.

Chapter 1

Introduction

The increasing accessibility of data in our society provides the opportunity to use these data to inform decisions across multiple domains, such as supply chain, transportation, and healthcare (Stobierski, 2019). Within the healthcare field, data-driven approaches have been proposed to predict the availability and need of resources, model patients' disease progression, support clinical and public health decisions, as well as to evaluate the effectiveness of interventions (Galetsi and Katsaliaki, 2020). More generally, these approaches can be categorized based on the level of granularity of the data. In this dissertation, we present new methods that use population and patient-level data to make better healthcare decisions. Using population-level data, we focus on modeling the supply and demand of organs for transplantation in the US. Based on patient-level data, we aim to personalize treatment and genetic testing strategies for the management of cardiovascular diseases.

1.1 Using Population-Level Data to Model the Supply and Demand of Organs for Transplantation

Population changes may impact the need and availability of resources. Within a healthcare system, resources may range from providers to organs available for transplantation. While demographic changes may affect various resources in our healthcare system, in this dissertation we center on the supply and demand of organs for transplantation.

Organ transplantation is a life-saving and cost-effective intervention for patients with organ failure. There are six types of solid organs that can be transplanted in the US: kidneys, liver, heart, lungs, pancreas, and intestines (Keller, 2015). Liver failure is the second most common reason to enroll in the organ transplantation waiting list (U.S. Organ Procurement and Transplantation Network, 2019).

Despite the number of organs available for liver transplantation (LT), the demand for livers greatly outweighs its supply. At the end of 2019, 9,675 livers were donated while 12,941 patients were added to the LT waiting list. There has been a gap between the amount of livers donated and the number of candidates added to the LT waiting list for more than 15 years. This has led to prolonged waiting times for LT and higher rates of waiting list dropout due to patient death or deteriorating medical conditions (U.S. Organ Procurement and Transplantation Network, 2019). There is a clear gap between the supply and demand of livers in the US. Moreover, there is a potential for further exacerbation of this disparity given the aging population and obesity epidemic in the US (Parikh et al., 2015).

In this dissertation, we present data-driven methods to understand how the expected population changes in the US may impact liver availability and demand in the future. These methods are built based on data from the Organ Procurement and Transplantation Network (OPTN), the US Census Bureau, and the University of Virginia's Weldon Cooper Center for Public Service.

1.2 Using Patient-Level Data for the Management of Cardiovascular Diseases

According to the National Vital Statistics, atherosclerotic cardiovascular disease (ASCVD), constituting coronary heart disease (CHD) and stroke, is the leading cause of death in the US (Kochanek et al., 2019). The Heart Disease and Stroke Statistics 2020 Update reports that CHD and stroke account for 42.6% and 17.0% of deaths attributable to cardiovascular diseases in the US, respectively (Virani et al., 2020). The management of ASCVD can be improved by (1) incorporating novel procedures, such as genetic testing, and (2) providing physicians and their patients with flexibility in the implementation of protocols.

Research has found that ASCVD has genetic and familial components (The CARDIoGRAMplusC4D Consortium, 2013; MacRae and Vasan, 2016). However, the role of genetic testing in the prevention and treatment of ASCVD is not well understood yet (Jarmul et al., 2018). Recently, a genetic risk score (GRS) that helps predict ASCVD was developed (Mega et al., 2015; Khera et al., 2016; Natarajan et al., 2017). This GRS may also help characterize which patients will likely receive the greatest benefit from cholesterol treatment. High blood cholesterol is one of the main controllable risk factors of ASCVD. Approximately 29.4% of adults in the US have high low-density lipoprotein (LDL) cholesterol (Virani et al., 2020). High LDL is the main concern while having high blood cholesterol. This

dissertation includes data-driven methods to determine who should receive genetic testing and how patients' cholesterol should be treated with or without genetic information. Using data from the National Health and Nutrition Examination Survey (NHANES), we illustrate the clinical and policy impact of genetic testing across different populations in the US.

Another main controllable risk factor of ASCVD is the patients' blood pressure (BP). Using the definition from the 2017 Hypertension Clinical Practice Guidelines, 45.6% of adults in the US have hypertension or high BP (Whelton et al., 2018). These guidelines have generated considerable controversy among practitioners (Ioannidis, 2018; Cohen and Townsend, 2018; Solberg and Miller, 2018; Wilt et al., 2018). In addition, the 2017 Hypertension Clinical Practice Guidelines (Whelton et al., 2018) provide conflicting recommendations regarding when to initiate pharmacological interventions with other guidelines such as Williams et al. (2018). Controversy and conflicting recommendations complicate the already difficult problem of deciding how to manage patients' BP. Further, physicians' opinion and patients' preferences may influence the selection of treatment plans, independently of the guidelines' suggestions (Cabana et al., 1999). To benefit from clinicians' expertise and account for any potentially conflicting recommendations, this dissertation introduces a data-driven method to obtain personalized ranges of near-optimal treatment options.

1.3 Organization and Contributions

The organization of this dissertation is shown in Figure 1.1. This dissertation is divided into two parts. The first part of the dissertation (Chapter 2) focuses on the development of models to predict the supply and demand of organs for transplantation, with an emphasis on liver transplantation. The second part of the dissertation (Chapters 3 and 4) centers on the management of ASCVD. In terms of the level of granularity of the data, the first part of the dissertation uses population-level data, and the second part uses patient-level data to address healthcare questions. The work in this dissertation is done in collaboration with experts from the Department of Industrial and Operations Engineering, the Department of Statistics, the Medical School, and the School of Public Health at the University of Michigan, as well as from the US Department of Veterans Affairs and the Department of Medicine at Harvard Medical School. We now briefly describe to contents of each chapter as well as our contributions.

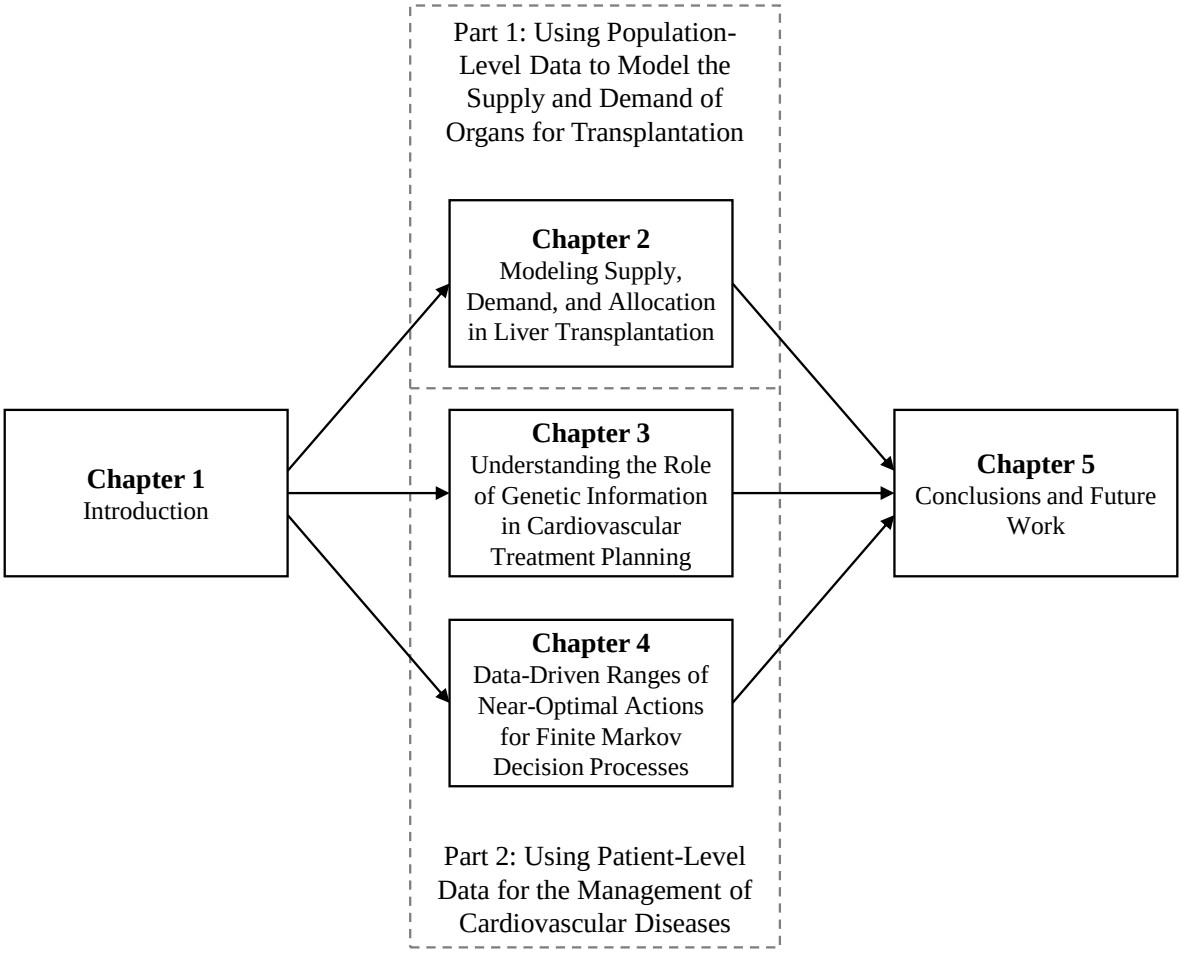


Figure 1.1: Organization of dissertation and connections between chapters.

1.3.1 Part 1: Using Population-Level Data to Model the Supply and Demand of Organs for Transplantation

In Chapter 2, we develop data-driven models to predict future organ supply and demand of organs for transplantation. We first use data on the expected demographic trends in the US and past donation rates to make projections on liver availability in the US. Then, we forecast the future burden of non-alcoholic steatohepatitis (NASH) related cirrhosis, one of the leading causes of additions to the LT waitlist in the US. Afterwards, we present a population-based approach to understand the impact of redistricting and demographic changes on the number of donors available in the US. Lastly, we aim to improve the prediction of deceased donor organ yield, a key performance metric in the organ donation system. Increasing organ yield is a potential way to reduce the gap between organ supply and demand. The main contributions of Chapter 2 are as follows.

- We design a heuristic model to make predictions based on population-level data.
- We present a modeling technique that integrates machine learning and stochastic simulation methods to make predictions.
- We forecast the supply and demand of livers for transplantation.
- We create a computationally efficient (through parallel computing) machine learning framework to predict how many organs one can expect to recover from a deceased donor.

The projection of liver supply was published in *Liver Transplantation* (Parikh et al., 2015), the forecast of NASH in *Hepatology* (Parikh et al., 2017), and the prediction of liver supply under the allocation policies was published in *Transplantation* (Parikh et al., 2017). A preliminary version of the estimation of the expected organ yield was published as an In Brief in *Transplantation* (Marrero et al., 2018). The complete organ yield analysis and results will be submitted to a medical journal.

1.3.2 Part 2: Using Patient-Level Data for the Management of Cardiovascular Diseases

In Chapter 3, we design cholesterol treatment plans and genetic testing strategies for the management of ASCVD. We first develop risk-based threshold treatment and testing strategies on the basis of the US Preventive Services Task Force (USPSTF) Guidelines (Bibbins-Domingo et al., 2016). Then, we develop cholesterol treatment plans using a

Markov decision process (MDP) model and perform genetic testing based on the USPSTF Guidelines as well as based on the demographic information of patients. Finally, we present a framework to obtain simultaneous cholesterol treatment plans and genetic testing strategies by combining dynamic programming with value of information (VoI) analysis. This chapter makes the following contributions:

- We develop a framework to incorporate genetic information to the risk of adverse events based on conventional factors.
- We create and thoroughly validate a simulation model to evaluate the implications of genetic testing in the US.
- We quantify of the impact of genetic information across different populations in the US.
- We design a framework to simultaneously make optimal cholesterol treatment and genetic testing decisions.

The risk-based threshold treatment and testing strategies is a working manuscript and will be submitted to a medical journal, the cholesterol treatment plans using MDP models and the simulation model to evaluate the effect of genetic information was published in the *Proceedings of the 2019 Winter Simulation Conference* (Marrero et al., 2019), and the framework to make simultaneous treatment and testing decisions is under review in *Health Care Management Science*.

In Chapter 4, we propose a new framework for identifying sets of near-optimal treatment choices. This framework integrates simulation-based dynamic programming (SBDP) and statistical multiple comparisons to provide clinicians' and their patients' with flexibility in the implementation of protocols. Overall, the contributions of this work are as follows.

- We develop a new SBDP algorithm, which we will refer to as the simulation-based backwards induction (SBBI).
- We provide finite sample, convergence, and asymptotic structural properties of the SBBI algorithm.
- We design a new multiple comparisons with a control (MCC) method, which we will refer to as the simulation-based multiple comparisons with a control (SBMCC) algorithm.

- We present a new notion of near-optimality by formulating stochastic optimization problems as hypothesis testing problems.
- We offer convergence and asymptotic structural properties of the SBMCC algorithm.
- We show the scalability of our approach to find ranges of near-optimal actions by applying our method to the management of hypertension.

This work will be submitted to an engineering journal.

Chapter 2

Modeling Supply, Demand, and Allocation in Liver Transplantation

In this chapter, we first present a heuristic model to forecast the future availability of livers in the US. We apply historical donation rates to the US Census Bureau population projections to estimate the future supply of livers. Then, we present a model based on linear regression and stochastic simulation to make projections of the future potential burden of NASH, one of the main causes of additions to the LT waiting list. We identify the population-based temporal relationship between obesity and NASH related cirrhosis requiring LT listing in the US. We also present a population-based approach to understand the impact of redistricting and demographic changes in the US on the number of livers donors available. Lastly, we compare several machine learning methods to predict overall deceased donor organ yield.

2.1 Background

There has been a decrease in the availability of deceased organ donors for LT in the US for more than a decade (Wong et al., 2014). The decrease in donor availability has increased the disparity among the number of patients listed for LT and the number of liver transplants performed (Dienstag and Cosimi, 2012). One factor contributing to the decrease in the availability of deceased organ donors for LT has been the decrease in the utilization of grafts. Factors contributing to the decrease in the utilization of grafts include the aging population, the obesity epidemic, and the increased prevalence of diabetes (Orman et al., 2013). The US population is expected to continue to age according to the projected demographic trends, thus likely exacerbating the availability of donors (US Census Bureau, 2011; Rayhill et al., 2007). Nonetheless, there is little understanding on the impact of demographic trends on future liver availability.

It is estimated that more than 20% of patients listed for LT will drop off the waiting list and approximately 17% will die before receiving a liver transplant each year (Northup et al., 2015; Kim et al., 2015). One of the main causes of addition to the waiting list for LT is NASH cirrhosis. NASH is the fastest growing indication for LT, becoming the 6th leading cause of additions to the waiting list for LT in 2014 and contributing 14.3% of the total additions (U.S. Organ Procurement and Transplantation Network, 2019). This condition resembles alcoholic liver disease, but occurs in people who drink little or no alcohol. NASH is characterized by fat in the liver, along with inflammation and damage. Given the historical rise of obesity in the US, the incidence of NASH is expected to increase. However, there are no studies on the future burden of NASH in the US.

Given the aging US population and better therapies for liver diseases, the demand for LT may fluctuate (Su et al., 2016). These factors vary geographically and have an impact on liver availability (Parikh et al., 2015). Estimating future liver availability as a function of the total population may provide additional insights into geographic inequities in organ availability.

2.2 Organization of the Chapter

This remainder of this chapter is organized as follows. Section 2.3 provides a review of the relevant literature. We study liver supply and demand in Sections 2.4 and 2.5. Then, we used our modeling approach for liver supply in Section 2.4 to understand the implications of geographic redistricting in Section 2.6. While we did not model the effect of geographic redistricting on liver demand, better allocation of livers could potentially reduce the liver transplantation waiting list. Finally, in Section 2.7, we aimed to improve the prediction of deceased donor organ yield. Better prediction of deceased donor organ yield may serve as an aid to improve future liver availability. The organization of this chapter, after the background and literature review, is shown in Figure 2.1.

2.3 Literature Review

The relevant literature in this chapter falls into two domains: (i) development of predictive models of organ supply and demand; and (ii) application of operations research (OR) models to organ transplantation.

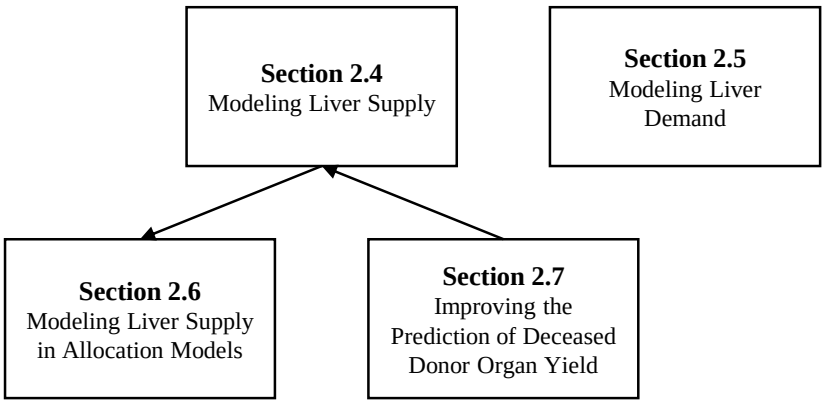


Figure 2.1: Organization and connections between sections in Chapter 2.

2.3.1 Predictive Models for Organ Supply and Demand

There is limited research regarding the use of predictive modeling to forecast future organ availability. Yee et al. (2010) developed a logistic regression model to predict the time of death after withdrawal of life-sustaining measures. This model could then be used to forecast the number of donors after cardiac death that would be available, which may be an indication for viable liver donors. However, this model does not forecast future liver availability itself. In addition, this model does not consider population changes or variation on donation rates to predict how many donors would be available in the future.

From the point of view of demand for organs, most of the work has been done in kidney, rather than liver. Tom and Kumar (2016) addressed the growth and style of cadaveric kidney demand in Kerala, India using a time series model. The results obtained during this study presented an increasing linear trend on the overall demand in Kerala. Abellan et al. (2004) addressed the problem of predicting the demand for kidney transplantation in Spain using discrete event simulation. Surprisingly, their results indicated that the waitlist was expected to decrease in the short term. Additionally, Roderick et al. (2004) created a discrete simulation model to estimate the future demand for renal replacement for a range of scenarios in England. According to their model, the acceptance rates and dialysis survival were the most influencing factors in future patient demand. Nevertheless, none of these models forecast demand for LT or consider demographic trends in their projections.

From an OR perspective, most of the predictive modeling work has been done in

graft survival, which can be used to determine which patients should have priority in the waiting lists when organs become available. Models to predict the graft survival of hearts, heart-lungs and kidney have been developed in several studies (Medved et al., 2016; Oztekin et al., 2009; Decruyenaere et al., 2015; Grams et al., 2012; Marrero et al., 2016; Rao et al., 2014). Raji and Chandra (2016) predicted the survival of liver grafts after transplantation using an artificial neural network. Their model showed an accuracy of 99.74% when evaluated using 10-fold cross-validation, while the current model for predicting liver graft survival presented an accuracy of 79.17%. The problem of predicting patient survival after LT was also addressed by Cruz-Ramírez et al. (2013). The authors used models based on artificial neural networks to develop a rule-based system to allocate organs. In addition, Kim et al. (2015) created classifiers based on several machine learning methods in an effort to improve the simulated allocation models for liver.

2.3.2 Application of Operations Research Models to Organ Transplantation

Several optimization models have been developed with the purpose of designing the optimal geographical partition of the US. Kong et al. (2010) created a mixed-integer programming model to maximize the efficiency of intraregional transplants by redesigning the liver allocation regions. Gentry et al. (2013, 2015) also used a mixed-integer programming model to divide the donor service areas (DSAs) into sharing districts that minimize the disparity in liver availability among districts. They evaluated and compared their models with a discrete-event simulation tool that represents liver allocation at a per-person, per-organ level. On the other hand, Akan et al. (2012) designed a multi-class fluid model of overloaded queues to model the waiting list for liver transplants. The authors used a bi-criteria objective function with the goal of minimizing the total number of patient deaths while waiting for LT and maximizing the total quality-adjusted life years (QALYs). Other optimization models in organ allocation are surveyed in Alagoz et al. (2009).

Other OR models related to liver transplantation are also available in literature. Alagoz et al. (2004) developed a model to determine the best time to accept a liver from a living donor with the goal of maximizing the patient's total reward (e.g. quality-adjusted life expectancy). The authors continued their work when they also addressed the problem of accepting or declining a liver of some quality from a patient's perspective (Alagoz et al., 2007b). The authors further extended their work by considering the problem of choosing between a liver from a living or a deceased donor from a patient's point of view (Alagoz et al., 2007a). While OR has provided many valuable insights in the field of

organ transplantation, none of the models directly address how to use the available data to understand changes in the supply and demand of organs.

2.4 Modeling Liver Supply

Although it is common perception that organ availability is decreasing, there are very few studies that quantify the expected organ availability. Moreover, little research has been done on the implications of demographic trends on the utilization of donors in the US over the next decade. Thus, our objective is to leverage the expected demographic trends and past donation rates to make projections on liver availability in the US over the next decade.

2.4.1 Data Sources

2.4.1.1 Historical Population Data

Data from the US Census Bureau is used to obtain 2002-2014 national population estimates. We use two US Census Bureau data sets for this purpose: (1) Intercensal Estimates of the Resident Population by Single Year of Age, Sex, Race, and Hispanic Origin for the United States: April 1, 2000 to July 1, 2010 and (2) Monthly Population Estimates by Age, Sex, Race, and Hispanic Origin for the United States: April 1, 2010 to July 1, 2014 (US Census Bureau, 2018, 2020). These data is stratified by race/ethnicity (white, black, Hispanic, and other), sex, and age group (18-34, 34-50, 50-64, and 65-84 years old). The population estimates by race/ethnicity, sex, and age group are used to calculate overall donation rates.

2.4.1.2 Historical National Population Obesity Prevalence Data

We segregate the historical population into obese and non-obese population using data from the NHANES from 1999 to 2010 (Flegal et al., 2012). The historical population data is then segregated by body mass index (BMI) (<30 or $\geq$ 30 kg/m2), race, sex, and age group. The data from Continuous NHANES is directly measured and collected in two-year cycles, thus we divide each cycle according to the time period the data is collected. We assume that data collected from November 1 through April 30 is representative of the first year of the cycle and data collected from May 1 to October 31 is representative of the second year of the cycle. We only include the records of adult individuals between ages 18-74 years with BMI less than 60 kg/m2 in our estimates of obesity prevalence. NHANES entries with any missing BMI data are excluded from this calculation (n = 869; 2.1%).

2.4.1.3 Donor Utilization Data

Liver utilization rates are obtained using the OPTN database from 2000 to 2012. We determine the utilization rates of whole and split livers for all donors with at least one organ transplanted. These rates are calculated as a percentage of the overall population per year.

2.4.1.4 Population Projections by Age, Sex and Race

Population projections in the US from 2014 to 2025 are derived using data from the US Census Bureau stratified by age, sex, and race (US Census Bureau, 2014). The population projections from the US Census Bureau provide three series of projections: the middle series, the high series, and the low series. We use the middle series of the population projections as the base case of our analysis.

2.4.2 Projection Development

We first forecast overall donor availability from 2014 to 2025 using the US population-based rates of donation (donors transplanted/total population) per year. We stratify the average donation rates from 2008 to 2012 by age group, race/ethnicity, sex, and BMI to obtain an estimate of overall donor availability. We then project future liver donor availability using the historic liver utilization rates (liver donors/total donors) segregated by age group, race/ethnicity, sex, and BMI. We use the average liver utilization rate from 2008 to 2012, and the best and worst liver utilization from 2000 to 2012 to forecast liver donors from 2014 to 2025.

2.4.3 Sensitivity Analysis

We perform four separate sensitivity analyses, varying the proportion of obese adults, proportion of Hispanic adults, overall US adult population growth over the projection period, and varying the change in liver utilization rate over time. For the first sensitivity analysis we vary the rate of BMI changes over the next decade. The ranges for the sensitivity analysis of BMI are derived from the average annual rate of increase in the US population BMI in the obese range (>= 30) from 1999-2010, which is approximately 0.75% per year (Flegal et al., 2012). The ranges for sensitivity analysis for the proportion of Hispanic adults are obtained from US Census Bureau high and low race/ethnicity projections (US Census Bureau, 2012). We focus the sensitivity analysis on the proportion of US Hispanics because this population is projected to be the most dynamic in the US

over the coming decade, relative to other racial/ethnic groups. For the overall population growth sensitivity analyses we use the high and low population projections from US Census Bureau data. Since liver utilization has been decreasing over the last 7 years and there has not been a clear nadir in utilization, we perform a sensitivity analysis varying the rate of liver utilization. We calculate the average change in liver utilization from 2007-2012, which is a mean decrease of 0.72% per year. We then conduct our projections increasing or decreasing the liver utilization by 0.72% per year.

2.4.4 Monte Carlo Simulation

We perform a Monte Carlo simulation to measure the combined impact of uncertainty in population projections, BMI, donation rates, and utilization rates on the projected donors. Each input variable in our model is defined a distribution. The population projections are modeled as a multivariate Gaussian distribution with mean as the middle series of the population projections stratified by age group, sex, and race, and standard deviation as one quarter of the difference between the high and low series. The parameters for the proportion of obese population (BMI $\geq$ 30 kg/m2) for each age, sex, and race group are sampled from normal distributions with means and standard deviations matching those in the NHANES study (Flegal et al., 2012). We sample with replacement the annual donation and utilization rates from 2000 to 2012.

2.4.5 Liver Supply Projections Results

The projected liver donor availability is depicted in Figure 2.2. Our base-case forecast is made using the average liver utilization rates from 2008 to 2012. We also use the best (2012) and worst (2004) liver utilization rates from 2002 to 2012 as alternative scenarios of future liver availability. All cases show a steady increase in the expected number of liver grafts available from 6,129 (6,047-6,462) in 2014 to 6,500 (6,429-6,833) in 2025. We forecast an overall increase in the percentage of available livers for transplantation of 6.8%.

Table 2.1 shows a comparison between the projected donor and adult population growth in the US. We estimate that the US population growth will surpass the growth in potential liver donors from 2014 to 2025. The number of utilized liver grafts is expected to increase in 6.1% and the number of available donors in 5.8%. On the other side, we estimate that the adult population with ages 18-75 in the US will increase in 7.1% from 2014 to 2025. This age range is considered acceptable at most US centers for receiving liver transplantation.

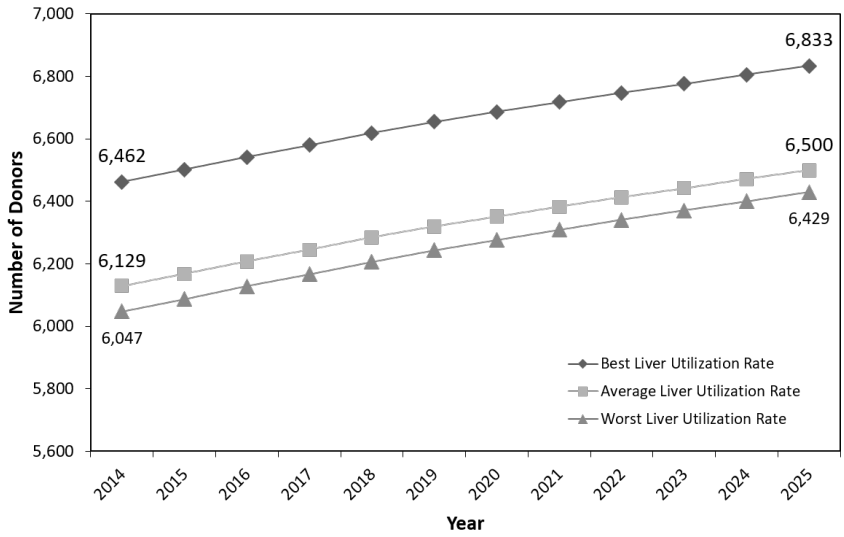
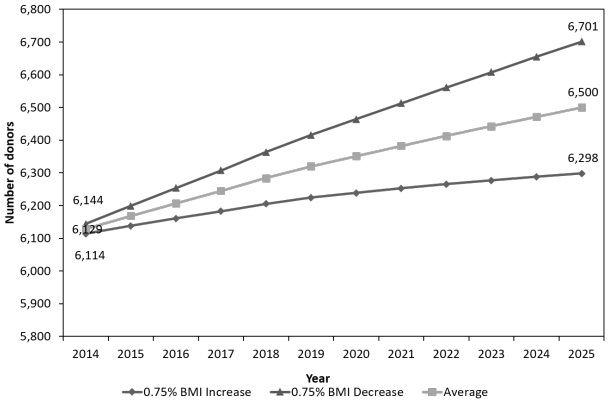


Figure 2.2: Projections of liver organ availability.

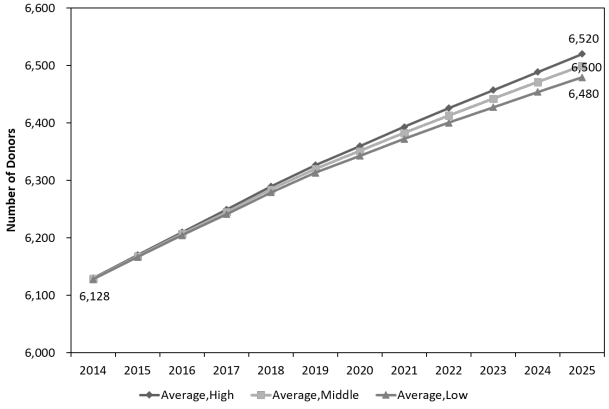
Table 2.1: Projected population and donor growth from 2014 to 2025.

Year	Number of Livers	% Liver Donor Growth (Average)	% Total Donor Growth	% Population Growth (18-75 Years Old)
2014	6129	0.7	0.7	0.9
2015	6168	0.6	0.6	0.8
2016	6207	0.6	0.6	0.8
2017	6245	0.6	0.6	0.8
2018	6284	0.6	0.6	0.8
2019	6320	0.6	0.5	0.6
2020	6351	0.5	0.5	0.6
2021	6383	0.5	0.5	0.6
2022	6413	0.5	0.4	0.6
2023	6442	0.5	0.4	0.3
2024	6471	0.5	0.4	0.4
2025	6500	0.4	0.4	0.4

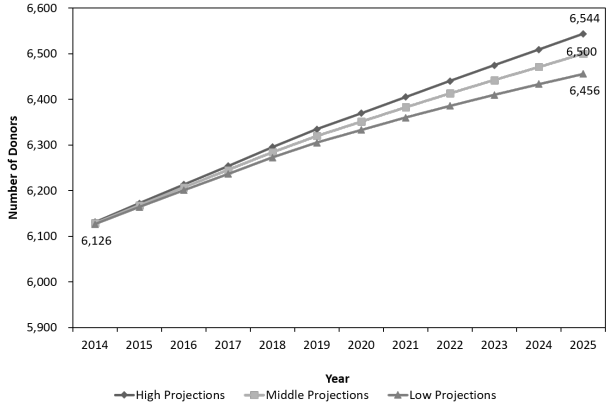
The results of the sensitivity analysis for BMI are shown in Figure 2.3a. The results show that donor availability is sensitive to increases or decreases in the proportion of adults with a BMI over 30 with a 6.2% range in projected liver donors. The results for the sensitivity analysis for changes in the proportion of the US Hispanic population are shown in Figure 2.3b. Changes in the Hispanic population do not result in a dramatic change in the projected number of liver grafts with a 0.62% range in projected liver donors. The impact of changes in total US population growth is shown in the sensitivity analysis in Figure 2.3c. The availability of donors is also fairly insensitive to changes in these parameters with a 1.4% range in projected liver donors. Figure 2.3d shows the sensitivity analysis of changes in liver utilization rate over time. With this projection, the 2025 utilization would be 90.12% in the utilization growth scenario and 73.32% in the utilization decline scenario. The results show that liver donor availability is sensitive to utilization with a 19.6% range at the end of the projection time period.



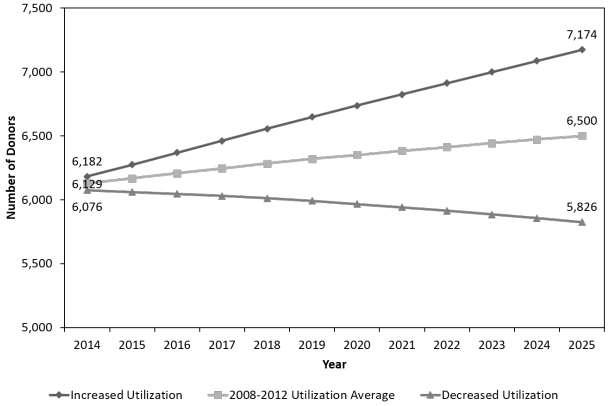
(a) BMI sensitivity analysis.



(b) Hispanic population sensitivity analysis.



(c) Population growth sensitivity analysis.



(d) Liver utilization sensitivity analysis.

Figure 2.3: Results of sensitivity analyses.

The Monte Carlo simulation is shown in Figure 2.4. Incorporating uncertainty in all

the input projection variables, the projections using the Monte Carlo simulation are very similar to our base case projections shown in Figure 2.2.

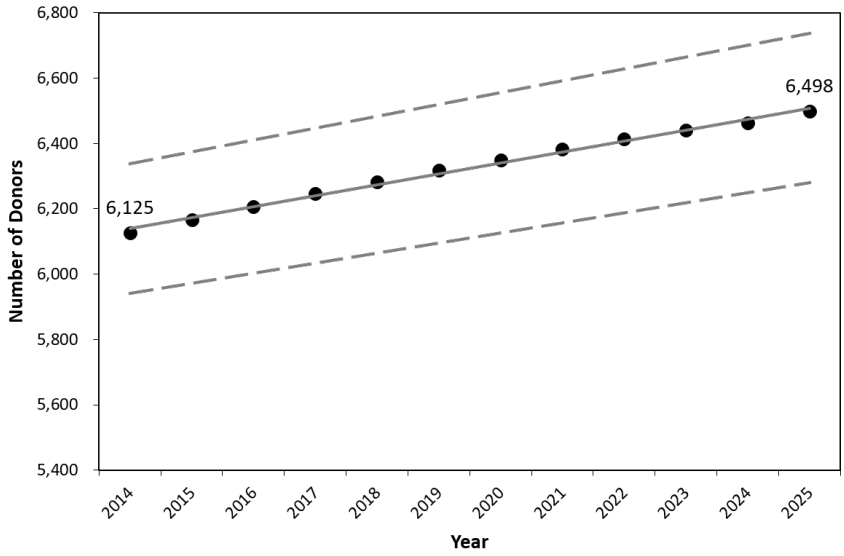


Figure 2.4: Monte Carlo simulation of liver availability.

2.4.6 Discussion

Liver transplantation remains the best life-saving therapy for patients with end-stage liver disease, however the ability to perform transplantation is limited by donor availability. While there is a perception that donor availability will continue to worsen with future demographic changes in the US population, this has not been objectively studied. We found that while the donor population will increase over the next 12 years, general population growth will outstrip projected donor growth, thus potentially exacerbating the donor shortage and increasing wait times for LT nationally. Sensitivity analyses were sensitive to population changes in BMI and liver utilization rate, but not changes in the proportion of the US Hispanic population or changes in the overall population growth.

Strategies to increase the donor pool are warranted to help alleviate the anticipated shortage in donors to decrease waitlist mortality. These include increasing donor enrollment and public awareness regarding organ donation (Salim et al., 2014; Li et al., 2013); optimizing processes of organ retrieval and allocation (Doyle et al., 2014); and utilization of technologies to increase viability of organs with prolonged cold ischemia time and donation after cardiac death donor organs (Graham and Guarrera, 2014). We otherwise do not anticipate a major expansion in the utilization of donation after cardiac death livers, given the inferior outcomes seen with these grafts. Expansion of living donor liver

transplantation could also help alleviate the projected shortage, however risks to donors remain a concern in the US (Hall et al., 2014; Abecassis et al., 2012). Broader regulatory changes that may lead to variance in risk aversion with respect to marginal grafts are more difficult to project; however with the increased focus in outcomes and quality of care and links to reimbursement, we anticipate that risk aversion may worsen moving forward (Bardach et al., 2013; Jha, 2013; Calikoglu et al., 2012).

There are many strengths and weaknesses of our study. We compiled unique data elements from many sources to complete our projections, and thus many of our input variables are dependent on single source data. The projections of organ availability assume no major changes in donor availability or breakthroughs in procurement technologies that could dramatically increase liver utilization rates. We accounted for incremental changes that would affect liver utilization with our final sensitivity analysis, varying the utilization rate over time. A significant breakthrough resulting in major changes in utilization could significantly alter these projections. However, we do not anticipate approval and widespread adoption of new technologies during the relatively short time-horizon of this study. In addition, BMI is an imperfect measure and is has decreased validity in elderly populations due in part to sarcopenic obesity (Janssen et al., 2005). Increased hepatic steatosis at lower BMIs in the elderly population may lead to lower than predicted availability of liver allografts. We also did not account for any other potential changes in public policy such as implied consent for donation that may increase the pool of donors available (Starzl, 1984).

Projecting the need for LT in the future is more uncertain, given the rapidly changing landscape of chronic hepatitis C and the emergence of non-alcoholic fatty liver disease (NAFLD). There is the possibility that the number of patients requiring transplantation will decrease, thus partially alleviating the donor shortage, despite our predictions. With new potent anti-viral agents against hepatitis C under development and recently approved (Afdhal et al., 2014), the need for LT in this population will likely diminish. However, NASH related cirrhosis has emerged as a leading indication for LT, and continued growth may cause further expansion of the LT recipient waitlist further (Wong et al., 2014a; Agopian et al., 2012). In the next section, we now study the possible impact of NASH in the LT waiting list.

2.5 Modeling Liver Demand

Obesity has become highly prevalent in the US. It is estimated that 38% of adults in the US have a BMI greater than 30 (Flegal et al., 2016). Obesity is associated with the development

of the metabolic syndrome, which includes dyslipidemia, hypertension, insulin resistance, and hepatic steatosis (Grundy, 2004). Individuals with hepatic steatosis can develop NASH, which over time leads to hepatic fibrosis and increased risk of development of hepatocellular carcinoma (HCC) (Tetri et al., 2008; Starley et al., 2010). Therefore, NASH is expected to become a leading indication for LT in the US and other countries with rising obesity rates (Flegal et al., 2016; Belli et al., 2016; Goldberg et al., 2017).

The metabolic syndrome and obesity are positively correlated with the development of NASH related cirrhosis and HCC (Hassan et al., 2010, 2015). Nonetheless, the temporal relationship between the prevalence of obesity in the population and the rise in NASH related cirrhosis is still unclear. Several developed and developing countries worldwide are experimenting increases in the prevalence of obesity. However, the lag in years between development of obesity in the population and increases in the number of patients that present for LT due to NASH is not yet characterized. Thus, we propose to identify the population-based temporal relationship between obesity and NASH related cirrhosis requiring LT listing in the US. In addition, we aim to forecast future additions to the LT waitlist due to NASH related cirrhosis.

2.5.1 Data Sources

2.5.1.1 Waiting List Data

We use data from the OPTN database from 2000 to 2014 to obtain the historic number of additions to the waiting list due to NASH each year. Since previous studies have shown that a significant proportion of the patients diagnosed with cryptogenic cirrhosis likely represented undiagnosed NASH cirrhosis, we use a modified definition of NASH that includes obese cryptogenic patients (Wong et al., 2014b). We include all adult waiting list additions between 18 and 74 years of age whose BMI is less than 60 kg/m2 in our analysis.

2.5.1.2 Historical National Population Obesity Prevalence Data

We once again use data from Continuous NHANES from 1999 to 2014 to estimate the national obesity prevalence per year. Please refer to Section 2.4 for details. We stratify the data from Continuous NHANES according to the following BMI categories 30 to <35, 35 to <40, and 40+ kg/m2. These categories are consistent with the obesity categories used by the CDC (Kuczmarski and Flegal, 2000).

2.5.1.3 Historical Population Data

In this section, we also we use data from the US Census Bureau to obtain 2002-2014 national population estimates (see Section 2.4 for details). We apply the obesity prevalence percentages obtained from Continuous NHANES to the national population estimates from the US Census Bureau to obtain an estimate of the national obese population. We use the Continuous NHANES data on obesity prevalence stratified by BMI categories to estimate the obese population by BMI category from the US Census Bureau national population estimates.

2.5.2 Model Selection

We examine the relationship among obese population estimates and NASH additions to the waiting list from 0 to 10 lagged years (i.e. a rise in obesity led to a rise in NASH additions 0 to 10 years later) using scatter diagrams. This range of lag years is limited by the availability of data on NASH additions to the waiting list. Scatter diagrams show that there is a linear association between obesity and NASH additions to the waiting list at all the time lags considered. Hence, we use a linear regression model to predict NASH additions to the waitlist based on obese population.

2.5.2.1 National Analysis

We evaluate the association between obese population and NASH additions to the waiting list using a linear regression model. The predictive performance of national obese population to forecast NASH additions to the waiting list is assessed using leave-one-out cross-validation methodology under different lag times (0 to 10 years). We choose the lag time that minimizes the internal validation root mean squared error (rMSE) (Refaeilzadeh et al., 2009).

Based on historical data, the adult obese population in the US increased linearly from 2000 to 2014. We therefore assume that time is a predictor for obese population. The adult obese population in the US is modeled as a function of time for future projections using linear regression.

2.5.2.2 Sensitivity Analysis on Obese Population

Although our data does not provide direct evidence for this behavior, studies suggest that the obesity epidemic has plateau in recent years and it may stabilize due to nationwide programs increasing the awareness about obesity prevalence. To address the potential

stabilization in obese population, we added a quadratic term to our regression model of obese population as a function of time.

2.5.2.3 Categorical Obesity Analysis

With the purpose of better understanding the impact of each BMI category on the projected additions to the waitlist, we perform an analysis that stratifies the general obese population by BMI. The obese population by BMI category (30 to 35, 35 to 40, and 40+ kg/m2) is estimated using linear regression with time as the predictor. We develop univariate and multivariate models to examine the relationship of each obese population BMI category and the additions to the waitlist due to NASH.

2.5.3 Projections and Stochastic Simulation

The additions to the waiting list due to NASH for those lagged years with data on obese population available are forecasted using linear regression. We calculate prediction intervals using a standard regression methodology (Faraway, 2009). However, there is no closed form solution to the prediction intervals of NASH additions to the waiting list for the cases where the lagged obese population is estimated by linear regression. Therefore, the point estimates and prediction intervals of NASH additions to the waiting list of these cases are estimated using stochastic simulation.

We simulate the NASH additions to the waitlist using linear regression models. Point estimates for the obese population are first simulated. Let p_i denote the obese population in year y_i in the dataset ($i = 1, 2, \ldots, n_1$). We fit a linear regression to the US obese population to obtain the intercept of the model, $\hat{\beta}_0$, the coefficient for the year of prediction, $\hat{\beta}_1$, and the mean square error of the regression model, $\hat{\sigma}_1^2$. To simulate the obesity population at year $k (k > n_1)$, $\hat{p}_k$, we use the following expression:

$$\hat{p}_k := \hat{\beta}_0 + \hat{\beta}_1 y_k + t_{\nu_1} \hat{\sigma}_1 \sqrt{1 + \frac{1}{n_1} + \frac{(y_k - \bar{y})^2}{\sum_{i=1}^{n_1} (y_i - \bar{y})^2}}, \tag{2.1}$$

where $\hat{\sigma}_1$ is the standard deviation of the obese population model, $\bar{y}$ is the average year in which we fit the regression, n_1 is the number of years used to fit the regression model, and t_{ν_1} is a simulated random variable from the standard student-t distribution with $\nu_1 = n_1 - 2$ degrees of freedom.

The additions to the waiting list due to NASH are simulated in a similar way. Using the obese population at year i, p_i, we forecast NASH additions to the waiting list, with a lag of L years, at year $i + L$ ($i = 1, 2, \ldots, n_2$). We fit a linear regression model to get the

intercept of the model, $\hat{\delta}_0$, the coefficient of obese population, $\hat{\delta}_1$, and the mean square error of the regression model, $\hat{\sigma}_2^2$. The future NASH additions to the waiting list at year $k+L$, $\hat{N}_{k+L}$, are obtained based on the point estimates of the lagged obese population, $\hat{p}_k$. The future additions to the waiting list due to NASH at year $k+L$ are simulated from:

$$\hat{N}_{k+L} \coloneqq \hat{\delta}_0 + \hat{\delta}_1 \hat{p}_k + t_{\nu_2} \hat{\sigma}_2 \sqrt{1 + \frac{1}{n_2} + \frac{(\hat{p}_k - \bar{p})^2}{\sum_{i=1}^{n_2} (p_i - \bar{p})^2}}, \tag{2.2}$$

where $\hat{p}_k$ is obtained from equation 2.1, $\hat{\sigma}_2$ is the standard deviation of the NASH additions to the waiting list model, p_i is the obese population from the ith year used to fit the model, $\bar{p}$ is the average obese population in which we fit the regression, n_2 is the number of years used to fit the regression model, and t_{ν_2} is a simulated value from the standard student-t random variable with $\nu_2 = n_2 - 2$ degrees of freedom.

For the multivariate case, equation 2.2 becomes

$$\hat{N}_{k+L} \coloneqq \boldsymbol{P}_{\boldsymbol{0}}' \hat{\boldsymbol{\delta}} \pm t_{\nu_2} \hat{\sigma}_2 \sqrt{1 + \boldsymbol{P}_{\boldsymbol{0}}' (\boldsymbol{P}'\boldsymbol{P})^{-1} \boldsymbol{P}_{\boldsymbol{0}}}, \tag{2.3}$$

where $\boldsymbol{P}_{\boldsymbol{0}}' \coloneqq [1, p_{01}, p_{02}, \dots]$ is a vector of new observations, $\hat{\boldsymbol{\delta}} \coloneqq [\hat{\delta}_0, \hat{\delta}_1, \hat{\delta}_2, \dots]$ is a vector of regression coefficients, and $\boldsymbol{P}$ is the design matrix of the linear model.

We replicate the simulation using 2.1 and 2.2 or 2.3 10,000 times. We obtain point estimates by averaging the simulated values and 95% prediction intervals by taking the 2.5% and 97.5% quantiles of the replications.

2.5.4 Model Accuracy Measures

We evaluate the predictive accuracy of our models at each time lag using leave-one-out cross-validation (Refaeilzadeh et al., 2009). Our dataset is divided into training and testing sets at each replication. The testing set at each replication is composed of a single observation selected at random from the dataset. The cross-validation procedure is replicated until all the observations are used as the testing set. The predictive performance of our models is assessed by calculating the average rMSE in the testing set across all the cross-validation replications. Additionally, we evaluate how well the models fit the data using the adjusted coefficient of determination R^2 and the average rMSE in the training set across all the cross-validation replications.

2.5.5 Liver Demand Projection Results

2.5.5.1 Lag Time Selection

Using linear regression and leave-one-out cross-validation we find that NASH is best explained and predicted using obese population from 9 years before, e.g., NASH additions to the waiting list at 2009 are associated with obese population at 2000. The relationship between the NASH additions to the waitlist and obese population from 9 years before from 2009 to 2014 is depicted in the scatter plot in Figure 2.5.

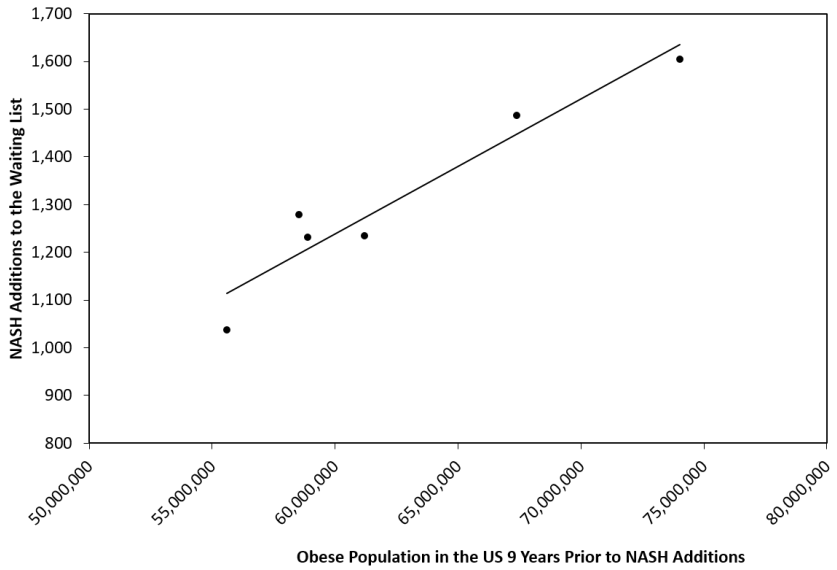


Figure 2.5: Relationship between NASH additions to the waiting list and obese population in the US using a 9-year lag.

Predicting NASH additions to the waiting list using obese population from 9 years before, we obtain an adjusted R^2 value of 0.9 and average rMSE of 50.49 in the training sets and of 86.49 in the testing sets. The performance of the model under all the time lags considered during our analysis (0 to 10 years) is included in Table 2.2.

2.5.5.2 National Projection of NASH Waitlist Additions

The number of additions to the LT waiting list due to NASH is expected to increase by 71.9% (1,354 to 2,327) from 2016 to 2030 (Figure 2.6). The point estimates and 95% confidence intervals from 2016 to 2023 are obtained with linear regression (before dashed line) and from 2024 to 2030 with stochastic simulation (after dashed line).

Table 2.2: Performance under different time lags.

Time Lag (Years)	rMSE Training	rMSE Validation	Adjusted R^2
0	199.56	227.31	0.74
1	177.23	199.27	0.78
2	184.11	223.60	0.75
3	204.32	257.25	0.63
4	182.39	218.48	0.63
5	102.66	125.05	0.85
6	141.98	184.41	0.64
7	179.15	274.42	0.28
8	117.14	168.35	0.62
9	50.49	86.49	0.90
10	52.72	83.78	0.82

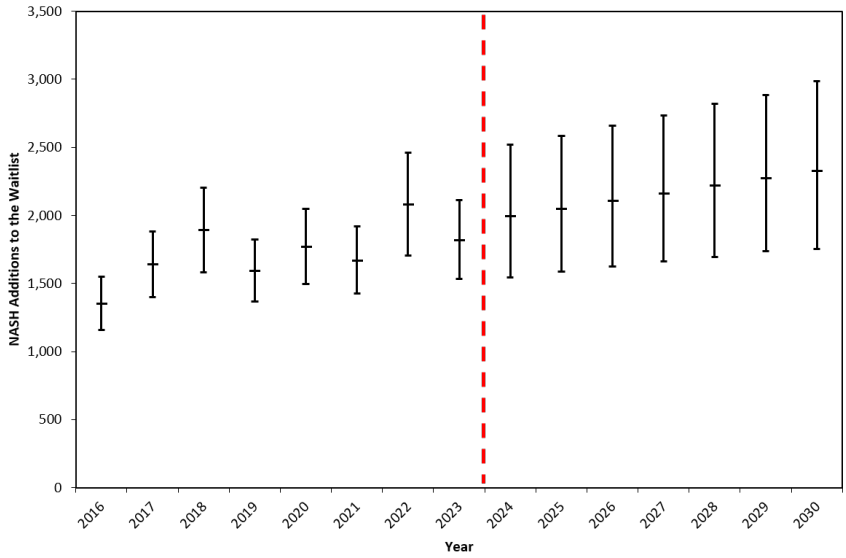


Figure 2.6: Point estimates and prediction intervals of NASH additions to the waiting list from 2016 to 2030.

2.5.5.3 Sensitivity Analysis on Obese Population

Figure 2.7 shows the point estimates and 95% prediction intervals for the additions to the waitlist due to NASH from 2024 to 2030, assuming a stabilization is the obesity prevalence in the US. As in our base case, these estimates are obtained using stochastic simulation. If the obesity prevalence in the US plateaus, we would expect a decrease of 3.2% (from 1,995 to 1,932) in 2024 and a decrease of 9.9% (from 2,327 to 2,104), when compared to the linear growth case.

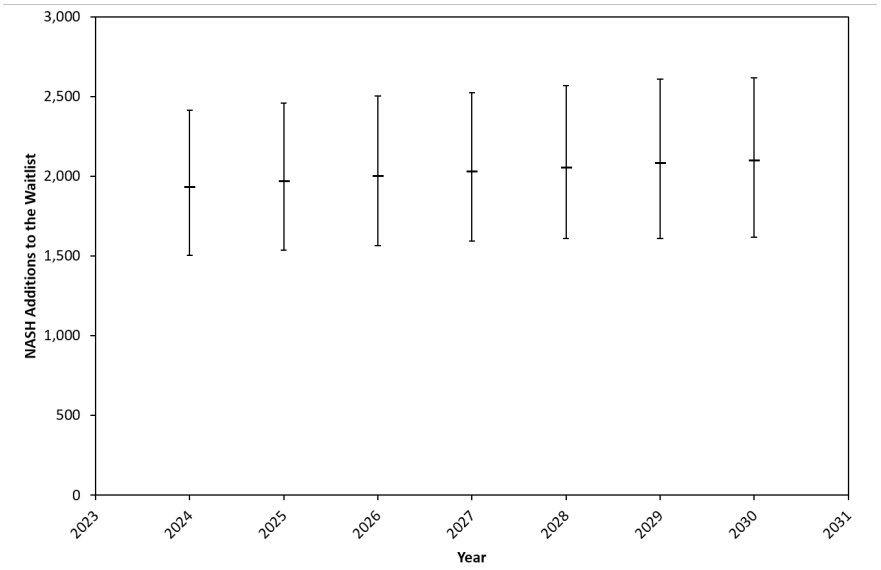


Figure 2.7: Point estimates and prediction intervals of NASH additions to the waiting list from 2024 to 2030 given plateau in obesity prevalence.

2.5.5.4 BMI Categorical Analysis

We find that the number of individuals with BMI 30 to <35 and BMI 35 to <40 are both significant predictors of additions to the waitlist due to NASH (both $P < 0.01$) in our univariate analysis. Nevertheless, the number of individuals in the BMI 40+ category is not a statistically significant predictor of NASH additions to the waitlist ($P = 0.16$). We also find that no single BMI category is a significant predictor of NASH additions to the waitlist with a 9-year lag if they all are included in our multivariate analysis. However, when we perform a stepwise removal of BMI categories from the multivariate model, only the number of individuals in the US population with BMI 30 to <35 is a significant predictor of NASH additions to the waitlist ($P < 0.01$).

2.5.6 Discussion

NASH-related cirrhosis and HCC are the fastest growing indication for liver transplantation in the US and many countries worldwide (Wong et al., 2014b; Doycheva et al., 2017; Goldberg et al., 2017). Our analysis provides epidemiological insight into the relationship between obesity in the general population and the NASH-related additions to transplant waitlists. We demonstrated, an increase in NASH-related liver transplant waitlist additions is expected to occur 9 years after population-level increases in obesity. This has potential future implications for the US and other countries around the world where obesity prevalence are rising (Malik et al., 2013). We have shown the impact on rising obesity with our projections of a 71.9% increase NASH related LT waitlist additions in the US over the next 15 years, likely making NASH the dominant indication for LT in the US in coming years. Given limited donor supply in the US, reductions in obesity on a population-level are particularly important to reduce the burden of NASH and NASH-related complications.

In our BMI categorical analysis, we found obesity classes I and II (BMI 30 to < 35 and BMI 35 to < 40 kg/m2) are predictive of NASH-related waitlist additions, while class III obesity (BMI > 40 kg/m2) is not predictive. This likely reflects the fact that many transplant centers do not consider extremely obese individuals for LT and patients very rarely lose significant amounts of weight (Schlansky et al., 2016; Thuluvath, 2007). Patients who have a BMI > 40 have a low likelihood of dropping in BMI category over a 9-year period, have a higher competing risk of non-liver related mortality, and thus are typically selected out of being added to the transplant waitlist (Marchesini et al., 2016).

This study quantifies the burden of NASH on the future transplant list, which is only a proxy of the overall burden NASH will have on the US healthcare system. Only a minority of patients with cirrhosis will be eligible liver transplantation due age, obesity, other comorbidities, or psychosocial barriers. This may be especially salient in patients with NASH as cirrhosis tends to occur in obese, older patients with cardiovascular comorbidities (VanWagner et al., 2016; Musso et al., 2013). Furthermore, our analysis underlies the need for better diagnostic and screening tools for identifying those with NASH in order to provide earlier interventions, such as weight loss strategies (Ajmera et al., 2017; Noureddin et al., 2016). Unfortunately, most patients who have NASH are unaware of their diagnosis and many only present when they have hepatic decompensation or develop advanced stage HCC, at which time there are few options that can modify the disease course. Thus, the overall impact of the increase in NASH is underestimated by this analysis, however we have provided an important estimation of the impact will have on LT in the US.

Our study has many notable strengths and weaknesses that warrant attention. The correlation between obesity and NASH additions to the waitlist, while biologically plausible, is an association and we have not proven causation. The data used for our historical analysis and development of projections are based on single source data from the US Census Bureau and the OPTN, thus we lack the ability to pool data sources for ranges in sensitivity analyses. Our model had high R^2 values and low rMSE, suggesting high performance, and results were consistent across subanalyses; however, it is well known that model performance is lower in validation cohorts than derivation cohorts. We performed internal cross-validation but could not externally validate our results. Given lack of adequate granular data we were forced to use different methods (i.e. linear regression and stochastic simulation) for point estimates and prediction intervals for the NASH LT addition projections.

Another potential limitation of our study is the possibility of historical misclassification between NASH and alcoholic liver disease. Potential misclassifications could affect the relationship between obesity and NASH. Future studies may explore potential ways of examining whether a substitution effect may be present. Finally, our analyses assume the current state of NASH-related care. Public health efforts to reduce obesity prevalence could affect our projections, as weight loss is associated with decreased steatosis and possible regression of fibrosis (Patel et al., 2015; Promrat et al., 2010). However, such efforts have largely failed thus far as seen in our historical analysis of obesity prevalence. Similarly, several new investigational treatments are being studied for the treatment of NASH which, if effective, could significantly impact our projections (Lazaridis and Tsochatzis, 2017).

How likely a patient is to receive a liver transplant can vary widely depending on where in the country the patient is listed. As a result, a proposal was presented to change how organs are allocated in the US. In the next section, we study the implications of geographical redistricting in liver availability using an extension of the liver supply model introduced in Section 2.4.

2.6 Modeling Liver Supply in Allocation Models

There are currently 11 geographical regions and 58 DSAs used for organ allocation in the US (U.S. Organ Procurement and Transplantation Network, 2019) The median Model of End-stage Liver Disease (MELD) score at transplant can vary from 22 to 32 depending on which of the 11 regions a patient is listed on nationally (Yeh et al., 2011). Previous studies have shown that the 90-day likelihood of being transplanted in the intermediate MELD

score range (18-30) can vary from 15% to 70% between regions (Massie et al., 2011) This variability can be translated to wide variations in risk of death while waiting for a LT (Massie et al., 2011; Rana et al., 2015).

This geographic inequity has led to the proposal of geographically redistricting the regions established by the United Network for Organ Sharing (UNOS). The proposal suggest that the allocation system of the nation should be modified from the current 11-region model into an 8-region model (Gentry et al., 2013). Figure 2.8 shows the current and proposed allocation systems. A 5-year liver simulation allocation model of the regional proposal suggests that the proposed 8-region model could achieve a net decrease in waitlist deaths, when compared to the current 11-region model (Gentry et al., 2013). While this analysis measures the potential impact of the redistricting proposal on waiting list mortality, it does not account for demographic shifts in the US population, such as changes in age, obesity prevalence, and racial distribution. Consequently, we aim to use a population-based approach to understand the impact of redistricting and demographic changes in the US on the number of liver donors available.

2.6.1 Liver Donor Projections by District Model

2.6.1.1 Historical Population

Similar to Section 2.4, we use data from the US Census Bureau to obtain 2002-2014 national population estimates. We also use data from the US Census Bureau to derive the population estimates by state and county. Three datasets are used for this purpose: (1) Intercensal Estimates of the Resident Population by Five-Year Age Groups, Sex, Race and Hispanic Origin: April 1, 2000 to July 1, 2010; (2) Annual State Resident Population Estimates for 6 Race Groups by Age, Sex, and Hispanic Origin: April 1, 2010 to July 1, 2014; and (3) Annual County Resident Population Estimates by Age, Sex, Race, and Hispanic Origin: April 1, 2010 to July 1, 2014. These datasets are used to calculate donation rates by county, state, and region.

We first divide our data by region and group it by regional model (8 or 11). The 8-region and 11-region maps are illustrated in Figure 2.8. The data is then stratified by race, sex and age group (18-34, 35-49, 50-64, 65-84 years old). The 18-34 age group of our models overlaps between two different age groups in the US Census Bureau dataset from 2002 to 2009. We account for this overlap by estimating the 18-34 age group by region based on the proportion of the remaining age groups (35-49, 50-64, 65-84 years old) by region to the total population of the 35-84 age group across all regions. Since data on county population is not available from 2002 to 2009, we assume that the population of

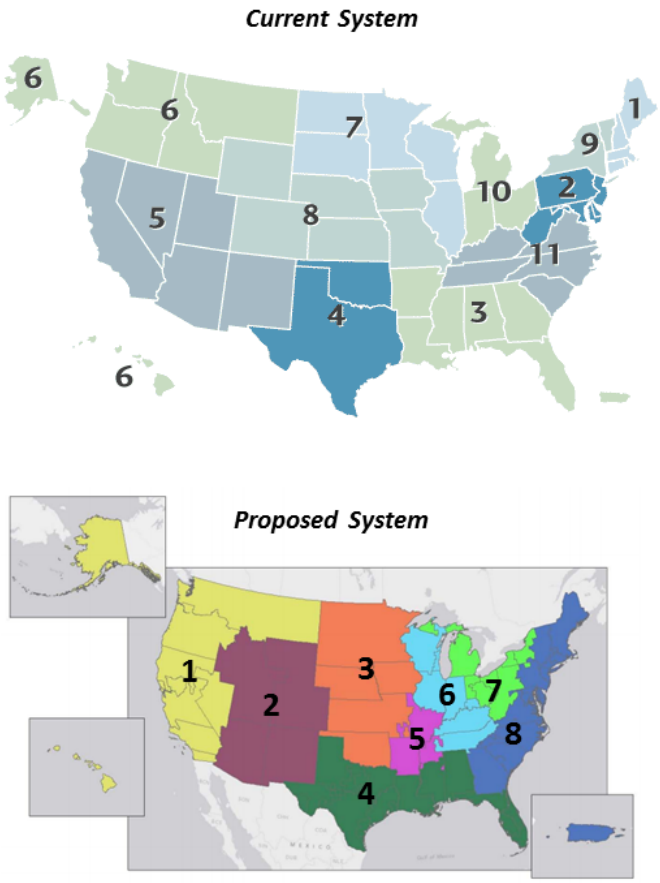


Figure 2.8: Liver allocation systems.

the states that extended along more than one region is divided according to the average proportion of the population (relative to the total state population), which lived in those counties from 2010 to 2014.

2.6.1.2 Population Projections

We use state population projections from 2010 to 2030 from the University of Virginia's Weldon Cooper Center for Public Service to estimate the proportion of the total population projection that corresponds to each OPTN region (Weldon Cooper Center for Public Service, 2013). The population projections are stratified by OPTN region, race, sex and age group. Since the population projections by state only include information for years 2010, 2020 and 2030, we assume linear growth to estimate the population projections from 2014 to 2019 and 2021-2025. The proportions of the population projections by OPTN region are then applied to the US Census Bureau population projections to maintain consistency on the total population projected.

2.6.1.3 Donor Availability

We once again use data from the OPTN database from 2002 to 2014 to determine the utilization rates of whole and split livers for all donors with at least one organ transplanted (see Section 2.4). We calculate donation rates as the number of livers donated per total population stratified by age groups, BMI, sex, and race/ethnicity. Donors are regionally localized based on their registered zip code in the OPTN database. We use donor permanent address rather than site of procurement to maintain consistency in our projections. Only 6.4% of patients have a regional discrepancy between permanent residence and site of procurement in the OPTN database.

2.6.2 Exploratory Analysis of Geographic Inequity

With the purpose of better understanding the leading contributors to geographic inequity, we perform an analysis where we standardize liver donation and utilization rates across the regions in both regional models. We use the average national donation rates, average national utilization rates, and both average national donation and utilization rates from 2010 to 2014. We then use the coefficient of variation to estimate the geographic variation in terms of donors per 100,000 population (D/100K) in each regional allocation model.

2.6.3 Liver Allocation Results

2.6.3.1 Regional Projected Donor Availability

The donor availability forecast using national population growth projections is shown in Figure 2.9. We estimate that the D/100K is going to decrease nationally over time from 2.53 in 2016 to 2.49 in 2025. Nonetheless, due to population growth, the overall number of liver donors is projected to increase from 6,133 in 2016 to 6,507 in 2025.

2.6.3.2 Geographic Inequity in Allocation Models

We project that the regional variation, in terms of D/100K, is going to decrease slightly from 20.3% in 2016 to 20.2% in 2025 in the 11-region model. In contrast, the geographic heterogeneity is projected to increase slightly from 13.2% in 2016 to 13.3% by 2025 in the proposed 8-region model (Figure 2.10).

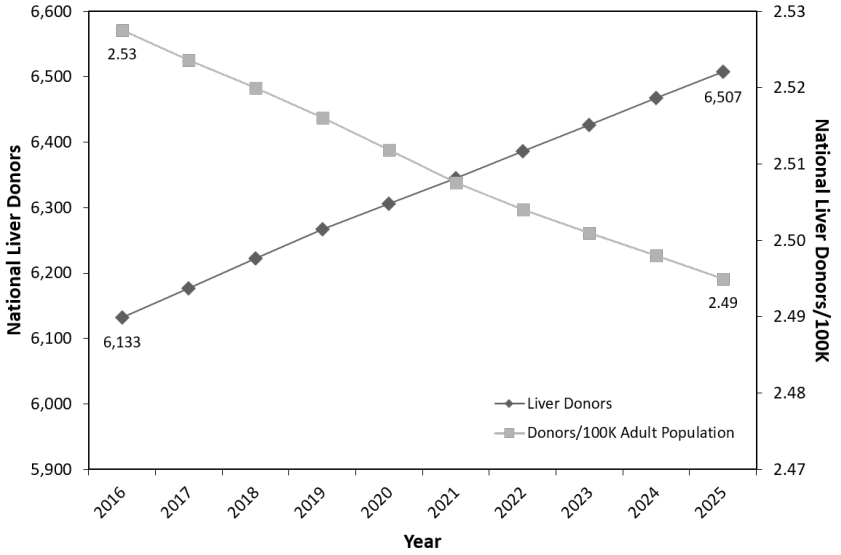


Figure 2.9: Projected National Changes in D/100K US Population and Liver Donors Available from 2016-2025.

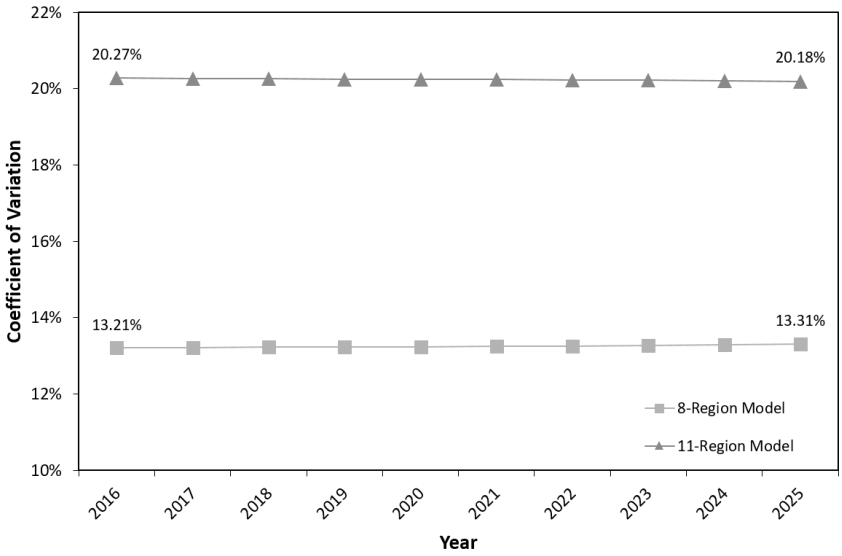


Figure 2.10: Projected Change in Regional Variation in D/100K US Population in each Allocation Model from 2016 to 2025.

2.6.3.3 Exploratory Analysis of Geographic Inequity

We aim to better understand the magnitude of regional disparity in donation and utilization rates. The expected effect of using the 2010-2014 national average rates of donation, utilization, and both on geographic variability in the regional models is summarized in Table 2.3. We estimate that standardizing utilization rates would reduce the geographic heterogeneity in 2016 by 2.2% to 11.0% in the 8-region model and by 3.0% to 17.3% in the current 11-region model. In contrast, standardizing national donation rates would reduce geographic heterogeneity by 5.7% to 7.5% in the 8-region model and by 11.5% to 8.8% in the current 11-region model. Moreover, using both national average donation and utilization rates would have the greatest impact in reducing geographic heterogeneity, reducing geographic variance to 4.9% in the 8-region model and 4.6% in the current model. A similar impact would be expected in both regional models in 2025.

Table 2.3: Effect of standardizing liver utilization rates, liver donation rates, and both on geographic variation in D/100K.

	2016 High-Low D/100K	**% Coefficient of Variation**	**2025 High-Low D/100K**	**% Coefficient of Variation**
National Average Utilization Rates and Regional Donation Rates				
8-Region Model	2.12-2.87	11.0%	2.10-2.83	11.2%
Current Model	1.88-3.16	17.3%	1.85-3.14	17.2%
National Average Donation Rates and Regional Utilization Rates				
8-Region Model	2.24-2.87	7.5%	2.21-2.84	7.6%
Current Model	2.21-2.91	8.8%	2.18-2.89	9.1%
National Average Donation and Utilization Rates				
8-Region Model	2.33-2.72	4.9%	2.26-2.65	4.9%
Current Model	2.33-2.71	4.6%	2.27-2.64	4.7%

2.6.4 Discussion

Regional variation in LT leads to inequity in organ distribution and excess waitlist dropout. The recent national redistricting proposal from UNOS attempts to address this inequity by redrawing the regional lines into an 8-region model. Using a population based approach we have shown that the geographic inequity in donor availability will be drastically decreased with the 8-region model. This finding confirms the notion that larger regions will lead to less geographic inequity in liver graft availability. The geographic inequity in the next decade is projected to fall slightly in the current 11-region model while geographic inequity is projected to increase slightly in the 8-region system. While our modeling con-

firms the desired effects of redistricting on improvement in geographic inequity, there are many barriers to implementing redistricting. These challenges must be weighed against the goal of decreasing waitlist mortality. Some of the same concerns were expressed with the implementation of "Share 35", however early results are encouraging with diminished mortality in the high MELD patients (Rana et al., 2015).

In our exploratory analysis to better understand the root cause of geographic inequity, we found that standardization of donation and utilization rates would lead to a greater reduction in geographic heterogeneity than redistricting. We found that variation in liver donation rates plays a bigger role in contributing to geographic inequity than utilization, suggesting that improving donation rates in lower performing regions may be impactful to improve geographic inequity. Similarly, focused interventions to improve transplant center organ utilization in low performing regions, would also aid in reducing inequity. Although these are not easy metrics to change, interventions to improve organ donation and utilization may be more palatable to many transplant centers and other stakeholders than redrawing the regional allocation maps. Such interventions, if successful, would also lead to an overall increase in number of donors and number of organs used – ultimate goals that would not be achieved with redistricting. In fact, there is some evidence from the liver simulated allocation model (LSAM) analysis of redistricting, there may be a net decrease in the number of transplants conducted (Goldberg et al., 2016). Our historical data shows that even low performing regions have at some point been near or above national averages in liver utilization, indicating that every region can perform at a higher level and measures to consistently improve practices in these regions can be successful. Improving donation rates in the low performing regions may be more difficult, as these regions have consistently lower donation than national averages in our historical data. Systematic evaluation of donor recruitment practices in these regions may be an important first step in improving donor authorization rates. Variations in regional demographics accounts for the remainder of the geographic inequity. However, our data suggest this is a relatively minor factor when compared to variation in liver utilization and liver donation rates.

Several prior studies have explored the challenges and implications surrounding redistricting and geographic inequity in LT. An inevitable consequence of sharing organs within larger regions are increased cold ischemia time and transportation costs for grafts (DuBay et al., 2015). However, a study showed that there would be cost-savings overall to the healthcare system due to shorter wait times for high-MELD patients, though some transplant centers might experience higher costs (Gentry et al., 2016). One concern with redistricting is that organs would flow from poor performing organ procurement

organizations (OPOs) to better performing OPOs. Gentry et al. (2015) used the LSAM to determine whether OPO import volume would change depending on the OPO performance metrics. They found that OPO performance would not impact flow of liver grafts, rather livers would flow from OPOs with lower than expected waitlist deaths to those with higher than expected waitlist deaths. Our study confirms that geographic differences in donor availability would decrease with redistricting. However, better aligning donation and utilization nationally may have more of an impact without increasing cold-ischemia time and possibly the cost of transplantation. None of the studies on the impact of redistricting have addressed the concern that redistricting may potentially decrease donation because organs cannot stay local. This would not be a concern if efforts to decrease geographic inequity focus on bringing low performing OPOs to national average donation and utilization rates.

Our study has many strengths and weaknesses that warrant attention. We compiled unique data elements from many sources to estimate historical and projected donor availability, and thus many of our input variables are dependent on single source data. We were unable to conduct a DSA-level analysis due to lack of historical or projected general population statistics on a DSA-level. We also used the donor zip code to identify their region, which does not account for donor regional sharing (Massie et al., 2015). It is unclear to what extent sharing would occur in the 8-region system or change in coming years, thus we used the donor registration site as a proxy for donor availability. We also did not use eligible deaths as defined by the OPTN as a proxy for potential donor availability in our analysis, given the increasing proportion of donors who may fall out of that definition (i.e., age greater than 70) as the population changes in the coming decade.

Our population-based analysis and projections are based on historical donation trends, thus donor authorization rates (donors/eligible deaths) are implicit in this analysis. The projections of liver availability assume no major changes in donor availability or breakthroughs in procurement technologies that could dramatically increase liver utilization rates. We do not anticipate approval or widespread adoption of new technologies during the relatively short time-horizon of this study. There has been a recent increase in donor availability due to the national epidemic opioid overdoses, a trend which bears monitoring and may impact our projections (Rudd et al., 2016). Finally our measure of equity, D/100K adult population, assumes that LT demand is proportional to the total population; however there are wide variations in demand depending on several sociodemographic factors (Axelrod et al., 2008; Hirose et al., 2016). Balancing donation rates would decrease geographic heterogeneity in donation, but would likely not completely balance geographic inequity in the ratio of eligible deaths to waitlist recipients. Dif-

ferential regional demographics, disease prevalence, urban/rural population, insurance coverage, and density of transplant centers may impact demand that is not captured by our model. However, with the changing landscape of chronic liver disease in the US, and changes in the availability in medical insurance with the adoption of the Affordable Care Act, D/100K may be a complementary metric to consider in evaluating organ allocation policies (Axelrod et al., 2010).

In this section, we found that improving regional donation and utilization rates may have a more profound effect on liver availability, without the logistical challenges of redistricting. Another approach to improve donor potential and reduce the gap between organ supply and demand is to increase the overall organ yield, or the number of organs transplanted per donor. But before improving overall organ yield, models to accurately predict such metric are needed. In the next section, we present a machine learning approach to predicting deceased donor organ yield.

2.7 Improving the Prediction of Deceased Donor Organ Yield

Since 2006, the United States Department of Health and Human Services (HHS) and the Centers for Medicare and Medicaid Services (CMS) have used the overall organ yield as a primary outcome measure to evaluate and recertify the OPOs coordinated by the OPTN (Centers for Medicare & Medicaid Services and United States Department of Health and Human Services, 2006).

Overall organ yield is also assessed internally by the OPTN Membership Professional Standards Committee (MPSC). The MPSC compares the observed and expected overall yield to evaluate the performance of each OPO in the U.S. The MPSC reviews any OPO with an observed yield of less than 10 organs per 100 donors compared to the expected yield, an observed yield more than 10% lower than the expected yield, and an observed yield significantly different from the expected yield (Organ Procurement and Transplantation Network, 2020). The 2019 HHS and CMS revisions to the outcome measure requirements propose to evaluate OPOs according to these measures as well (Centers for Medicare & Medicaid Services and United States Department of Health and Human Services, 2019). Thus, the prediction of expected organ yield is a key element in accurately assessing organ procurement performance.

The first published model designed to predict organ yield used ordinary least squares regression with variables in the OPTN database to predict the number of organs trans-

planted per donor (Selck et al., 2008). This model was then extended with the development of overall and organ-specific yield models (Messersmith et al., 2011). The authors used ordinal logistic regression to predict the overall and kidney yields, and binomial logistic regression to predict lung, pancreas, and liver yield. Most recently, the Scientific Registry of Transplant Recipients (SRTR) developed another set of organ-specific yield models (Scientific Registry of Transplant Recipients, 2019). The SRTR also predicted heart, intestine, lung, pancreas, and liver yields with binomial logistic regression, but predicted kidneys yield using multinomial logistic regression. The overall yield was modeled as the sum of the individual organ-specific yields.

Although these techniques have been shown to have reasonable performance in predicting overall and organ-specific yield, other modeling approaches may improve on the accuracy of prediction. In particular, the overall yield can be modeled as counts, which allows for the consideration of a wide range of machine learning models. In this study, we aim to evaluate different modeling techniques to determine which has the best performance in predicting overall deceased donor organ yield.

2.7.1 Methods

2.7.1.1 Data Source and Study Design

Our analysis is performed using data from the OPTN database from 2000 to 2018 (U.S. Organ Procurement and Transplantation Network, 2019). The inclusion criteria for the study are adult deceased donors between 18 and 84 years of age that have at least one organ procured for transplantation.

The outcome of interest during this analysis is the number of organs transplanted per deceased donor (i.e. overall deceased donor organ yield). To be consistent with the SRTR models, we consider an overall yield ranging from 0 to 7 organs per donor. We count single lung and double lung transplants as a single organ transplant and count kidney transplants separately as 0, 1, or 2 organs per donor.

The initial set of predictors for organ yield is selected from previously published studies (Selck et al., 2008; Messersmith et al., 2011; Marrero et al., 2016; Scientific Registry of Transplant Recipients, 2019). We exclude donors with missing data in any numerical variable, with the exception of warm ischemia time and donor ejection fraction ($n = 26,303$). To avoid the exclusion of 30,866 donors, we categorize the donor ejection fraction into the following groups: less than 20%, from 20% to 30%, from 30% to 40%, from 40% to 50%, from 50% to 60%, from 60% to 70%, from 70% to 77%, greater than 77%, and unknown. Since the warm ischemia time is only available for DCD donors prior to

03/31/2015, we also categorized this variable to avoid the exclusion of donors ($n = 64$). We group the warm ischemia time of DCD donors into the following categories: less than 20 minutes, from 20 to 30 minutes, from 30 to 40 minutes, more than 40 minutes, and unknown.

On the remaining data, we also exclude any donor with missing information on categorical variables with a percentage of missingness below 10% ($n = 19,140$). The purpose of this exclusion is to avoid categorical levels in the random holdouts that are not included during the model fitting process in our cross-validation analysis. We recode the binary variables missing in at least 10% of the donors in our dataset into categorical variables with three levels: yes, no, and unknown. The impact of this exclusion criteria is investigated in a sensitivity analysis that included all donors with missing data.

2.7.1.2 Machine Learning Models

We test several machine learning models to predict overall deceased donor organ yield. The following models are considered: mean-only, Poisson regression, negative binomial (NB) regression, general additive models (GAM), multivariate adaptive regression splines (MARS), artificial neural networks (ANN), classification and regression trees (CART), random forests, tree-based bootstrap aggregation, tree-based gradient boosting, and Bayesian additive regression trees (BART). Since the response variable of interest is the number of organs transplanted per deceased donor, all the models are trained to predict a count response. These models are briefly described in Appendix A.1.

For comparison purposes, we also apply the methodology of previously published organ yield models to our dataset (Selck et al., 2008; Messersmith et al., 2011; Scientific Registry of Transplant Recipients, 2019). These models will be referred to as the adjusted Selck model (Selck et al., 2008), the adjusted Messersmith model (Messersmith et al., 2011), and the adjusted SRTR model (Scientific Registry of Transplant Recipients, 2019). The application of these models is described in Appendix A.1.

2.7.1.3 Model Evaluation

The predictive accuracy of each model is assessed using Monte Carlo cross-validation (Hastie et al., 2009; McLachlan, 1992). We use 80% of the data for derivation in the cross-validation analysis and the remainder of the data as a validation set. The Monte Carlo cross-validation analysis is replicated 30 times and the random holdouts consisted of 20% of the derivation cohort. For the appropriate models, we use the remaining 80% of the derivation cohort to determine the best tuning parameters using 5-fold cross-validation

(Kuhn, 2008; Kapelner and Bleich, 2013; Friedman, 1991).

The predictive out-of-sample performance of the machine learning models is assessed in terms of the mean absolute error (MAE) and the mean squared error (MSE). Both of these metrics are commonly used measures of how different predictions are from the true values (Hastie et al., 2009; Borovicka et al., 2012). These metrics are negatively oriented, which means lower values are better. The MAE is the same units as the variable of interest, whereas the MSE is in squared units of the variable of interest. One key difference among them is that the MAE is less sensitive to outliers, while the MSE penalizes large errors more heavily. When applicable, we present the average and standard deviation (SD) of these performance measures over the cross-validation analysis. The comparison of the MAE of the models also allowed us to quantify the practical impact of accuracy improvements on the prediction of overall organ yield.

2.7.1.4 Statistical Analysis

We aim to find the model with the lowest predictive error. Any model with convergence issues or violating any statistical assumption is discarded as a potential final model. To verify which model has the lowest MAE and MSE, we first assess the normality of these values using the Shapiro-Wilk test. We then use 2-sided two-sample t-tests or Mann-Whitney U tests with the Bonferroni correction to compare these performance measures, as appropriate (Faraway, 2014; Hollander and Wolfe, 1999). Once the best performing model is selected, we use variable importance methods to identify the most influential predictors in the model and partial dependence plots to understand the functional relationships between these predictors and the overall organ yield (Friedman, 2002; Ridgeway, 2007).

The Bonferroni correction method is an approach to control the familywise error rate when multiple statistical tests are performed simultaneously. This correction method compensates for the increase in the chance in observing rare events (and incorrectly rejecting a null hypothesis) by testing each individual hypothesis at a significance level of α/k, where α is the desired overall significance level and k is the number of hypotheses being tested simultaneously. For instance, if the desired overall significance level is $\alpha = 0.05$ and 14 different machine learning models are compared simultaneously, then the Bonferroni correction method would test each individual hypothesis at a significance level of $\alpha/k = 0.05/14 = 0.004$. See Hastie et al. (2009) or Faraway (2014) for additional details.

Descriptive statistics are presented as counts with percentages for categorical variables and as means with SD for numerical variables. An overall significance level of $\alpha = 0.05$ is used during all the analyses. The significance level and confidence interval (CI) of

individual statistical tests are adjusted with the Bonferroni correction method to account for the simultaneous comparison of multiple models (Hastie et al., 2009). All analyses are performed with R (v3.5.3 The R Foundation for Statistical Computing, Vienna AT).

2.7.1.5 Sensitivity Analyses

To incorporate the donors originally excluded in our analysis, we impute their missing data using the multivariate imputation by chained equations (MICE) package in R (Van Buuren and Groothuis-Oudshoorn, 2011). This technique estimates the conditional density of each incomplete variable and specifies an imputation model per variable. Starting from an initial imputation, the chained equations draw imputations by iterating over the conditional densities. Missing values are imputed 10 times for 10 iterations as per suggestions in the literature (Bodner, 2008; White et al., 2011). We then replace the missing values using the mean imputed value for numerical variables and random sampling for categorical variables. This version of our dataset is then used in our cross-validation analysis to evaluate the performance of each machine learning model.

Because OPOs are evaluated in two through four-year cohorts by MPSC and CMS (Zaun et al., 2012; Centers for Medicare & Medicaid Services and United States Department of Health and Human Services, 2019), we also compare the performance of the models over multiple 2-year cohorts. We evaluate the predictive accuracy of the models in our cross-validation framework using the following 2-year cohorts: 2013-2014, 2015-2016, and 2017-2018. No considerable differences were observed in the donors across these cohorts. Since the SRTR model is built upon organ yield models using a 2-year cohort, we also include the direct application of their overall yield model in this sensitivity analysis for comparisons purposes. This model will be referred to as the SRTR model.

2.7.2 Results

2.7.2.1 Cohort Characteristics

In total, 135,277 deceased donors are between 18 and 84 years of age. After applying the exclusion criteria, a total of 89,520 donors are included during the analysis. A summary of all the predictors considered in our study is included in Table 1.

Table 2.4: Summary of all the predictors of overall deceased donor organ yield considered in our study ($n = 89,520$).

Description	No.	%
Donor sex (male)	52,280	58.4%
Donor race		
White	60,029	67.1%
Black	13,981	15.6%
Hispanic	12,523	14.0%
Asian	2,191	2.5%
AIAN	372	0.4%
NHPI	222	0.3%
Multiracial	202	0.2%
Blood type		
O	42,882	47.9%
A	12,698	14.2%
A1	17,652	19.7%
A1B	1,202	1.3%
A2	2,676	3.0%
A2B	418	0.5%
AB	1,404	1.6%
B	10,588	11.8%
HBV core antibody status	4,881	5.5%
HBV surface antigen status	107	0.1%
Hepatitis C status	4,887	5.5%
CMV status	58,570	65.4%
Risk factors for blood-borne disease transmission	13,441	15.0%
RPR-VDRL serology result	666	0.7%
Tattoos	31,751	35.5%
Heavy alcohol use (2+ drinks/day)	16,773	18.7%
History of cigarette use	24,275	27.1%
Cigarette use in the last six months (20+ packs/year)		
No	4,262	4.8%
Unknown	65,354	73.0%
Yes	19,904	22.2%
History of cocaine use	15,521	17.3%
Cocaine use in the last six months		
No	6,506	7.3%

Unknown	75,804	84.7%
Yes	7,210	8.1%
History of drug use (not cocaine)	33,932	37.9%
Drug use (not cocaine) in the last six months		
No	8,962	10.0%
Unknown	58,656	65.5%
Yes	21,902	24.5%
History of diabetes	11,247	12.6%
History of hypertension	33,946	37.9%
History of MI	3,834	4.3%
History of cancer	3,333	3.7%
Clinical infection	53,168	59.4%
Blood infection	8,400	9.4%
Pulmonary infection	43,951	49.1%
Urine infection	11,881	13.3%
Infection from another source		
No	46,533	52.0%
Yes	6,635	7.4%
Unknown	36,352	40.6%
Cause of death		
CVA	35,247	39.4%
Anoxia	24,089	26.9%
Head trauma	27,895	31.2%
CNS tumor	459	0.5%
Other	1,830	2.0%
Circumstance of death		
Natural causes	38,119	42.6%
MVA	13,378	14.9%
Suicide	7,882	8.8%
Homicide	4,703	5.3%
Non-MVA	9,513	10.6%
Other	15,925	17.8%
Mechanism of death		
Natural causes	1,839	2.1%
Drowning	340	0.4%
Seizure	668	0.8%

Drug intoxication	6,726	7.5%
Asphyxiation	2,883	3.2%
Cardiovascular	12,481	13.9%
Electrical	44	0.1%
Gunshot wound	8,134	9.1%
Stab	167	0.2%
Blunt injury	18,546	20.7%
Intracranial hemorrhage/stroke	35,872	40.1%
Other	1,820	2.0%
Declaration of brain death due to cardiac arrest since neurological event	6,063	6.8%
DCD donor	5,848	6.5%
Controlled DCD status		
No	115	0.1%
Unknown	83,687	93.5%
Yes	5,718	6.4%
Insulin dependent		
No	5,990	6.7%
Yes	4,661	5.2%
Unknown	78,869	88.1%
Protein in urine	40,803	45.6%
Donor recovered outside the U.S.	669	0.8%
Year of recovery		
2000	13	0.0%
2001	14	0.0%
2002	10	0.0%
2003	60	0.1%
2004	2,983	3.3%
2005	5,399	6.0%
2006	5,705	6.4%
2007	5,236	5.9%
2008	5,264	5.9%
2009	5,756	6.4%
2010	6,222	7.0%
2011	6,444	7.2%
2012	6,353	7.1%

2013	6,413	7.2%
2014	6,554	7.3%
2015	6,089	6.8%
2016	6,862	7.7%
2017	7,028	7.9%
2018	7,115	8.0%
Description	**Mean**	**SD**
Donor age (years)	44	1540.8%
Squared donor age	2,141	138652.8%
Donor height (cm)	172	1024.8%
Donor weight (kg)	83	2100.0%
Donor BMI (kg/m2)	28	667.8%
Squared BMI	838	44122.5%
Terminal blood urea nitrogen (mg/dL)	21	1717.6%
Terminal lab creatinine (mg/dL)	2	150.9%
Lung pO2 (mmHg)	228	14924.0%
pO2/FiO2 (ratio)	8	265.0%
Terminal lab AST (IU/L)	113	44381.8%
Terminal lab ALT (IU/L)	101	32268.1%
Terminal lab total bilirubin (mg/dL)	1	136.3%
KDRI	1	55.8%
Ejection fraction (%)	57	1312.6%
Warm ischemia time (minutes)	19	1502.3%

2.7.2.2 Model Performance

The adequacy and predictive performance of the machine learning models are evaluated throughout each replication of the Monte Carlo cross-validation framework. The model adequacy checks and reasons for discard are included in Appendix A.2. The MAE and MSE of the models are illustrated in Figures 2.11 and 2.12.

The Bonferroni correction threshold results in 0.004, as 14 models are compared simultaneously. We find that most models are significantly different from each other in terms of their MAE (all $P < 0.001$). The exceptions are the mean-only model with CART (99.6% CI -0.005 to 0.005, $P > 0.99$), Poisson regression with NB regression (99.6% CI -0.002 to 0.003, $P = 0.728$), and GAM with random forest (99.6% CI -0.001 to 0.005, $P = 0.061$). Similar results are found when comparing the MSE of the statistical models. However, the

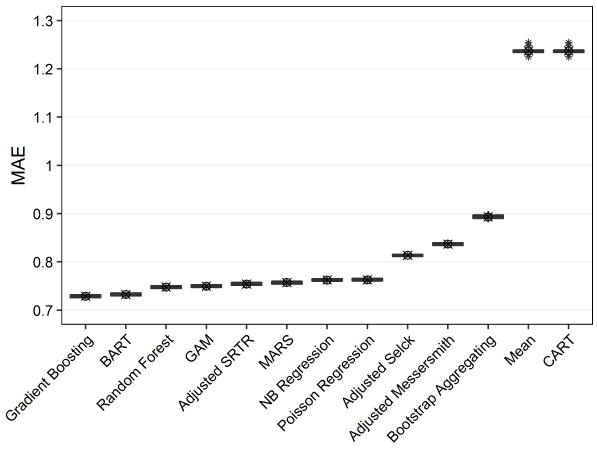


Figure 2.11: MAE of potential models to predict overall deceased donor organ yield. Points in the center of the boxes represent the mean MAE of the statistical models and asterisks represent outliers. The MAE of the ANN models are excluded for illustration purposes. ANN presented a median MAE of 2.766 with an interquartile range of 0.852.

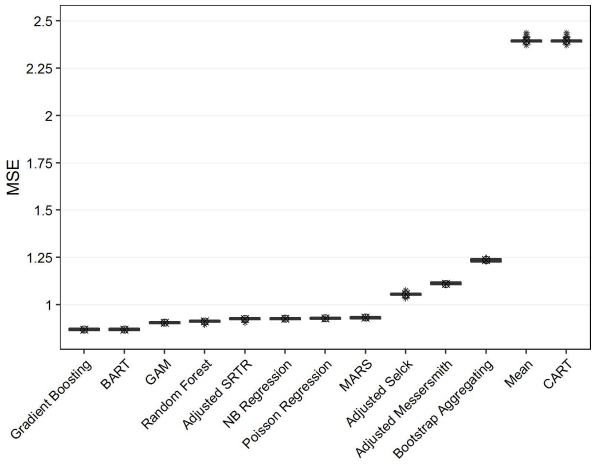


Figure 2.12: MSE of potential models to predict overall deceased donor organ yield. Points in the center of the boxes represent the mean MSE of the statistical models and asterisks represent outliers. The MSE of the ANN models are excluded for illustration purposes. ANN presented a median MSE of 11.449 with an interquartile range of 5.531.

adjusted SRTR, Poisson regression, NB regression, and MARS models are not statistically different from each other in terms of their MSE.

Combining the results from Figure 2.11 with two-sample t-tests, we determine that a tree-based gradient boosting outperforms the remainder of the models. Linking the results from Figure 2.12 with Mann-Whitney U tests we can draw similar conclusions. However, there is not enough evidence to claim that the MSE of gradient boosting is significantly different from the MSE of BART (99.6% CI -0.006 to 0.006, $P = 0.901$). This result is consistent with our previous work, where we find that BART outperformed the rest of the models (Marrero et al., 2018).

Gradient boosting results in the lowest predictive error among the compared models under both performance measures and thus we select it to predict overall deceased donor organ yield. This model presents an average (SD) MAE of 0.729 (0.004) and an MSE of 0.868 (0.004) throughout the Monte Carlo cross-validation analysis. The model is then fit to the derivation set ($n = 62,664$) and evaluated in the validation set ($n = 26,856$), which results in similar performance (MAE of 0.725 and an MSE of 0.863).

2.7.2.3 Relationship Between Predictors and Deceased Donor Organ Yield

Since the final model contained over 110 predictors (all variables and levels of categorical predictors listed in Table 2.4), we use variable importance methods to simplify our results (Friedman, 2002). Although 109 predictors have non-zero influence on deceased donor organ yield, we examine the relationship between deceased donor organ yield and any predictor with a relative importance greater than 1% (out of a relative importance normalized to 100%). The variable importance strategy indicated that donor race, hepatitis C infection, and ejection fraction are the three most important categorical predictors. The partial dependence plots of these predictors are included in Figure 2.13.

The variable importance strategy shows that following numerical predictors are the most important: kidney donor risk index (KDRI), partial pressure of oxygen (pO2), terminal lab creatinine, BMI, age, and terminal blood urea nitrogen, the ratio of pO2 to the fraction of inspired oxygen (pO2/FiO2), weight, aspartate aminotransferase (AST), alanine aminotransferase (ALT), and terminal lab total bilirubin (Figure 2.14).

2.7.2.4 Impact on Organ Yield Prediction

By comparing the MAE of our gradient boosting model to the MAE of the models in the literature, we are able to measure the impact of improving prediction accuracy on overall organ yield in terms of organs per 100 donors. We estimate that gradient boosting would improve prediction by an average (SD) of 8 (0.005) organs per 100 donors (compared to the adjusted Selck model), 11 (0.005) organs per 100 donors (compared to the adjusted Messersmith), or 3 (0.005) organs per 100 donors (compared to the adjusted SRTR model). This translates to an improvement of 10%, 13%, or 3%, in comparison to the models developed by Selck et al. (2008), Messersmith et al. (2011), and the Scientific Registry of Transplant Recipients (2019) after adjustment for the current dataset, respectively.

Comparing the aggregated yield across all the deceased donors in our cross-validation analysis, we are able to further measure the effect of improving the accuracy of predictions. The average (SD) observed yield over the cross-validation analysis is 292 (0.010) organs per 100 donors. The average (SD) expected yield is 292 (0.008), 292 (0.009), 295 (0.008), and 291 (0.007) organs per 100 donors calculated using the gradient boosting model, the adjusted Selck model, the adjusted Messersmith model, and the SRTR model after adjustment for the current dataset, respectively. Therefore, greater predictive accuracy for individual donors could also result in better aggregate yield predictions over the deceased donor population.

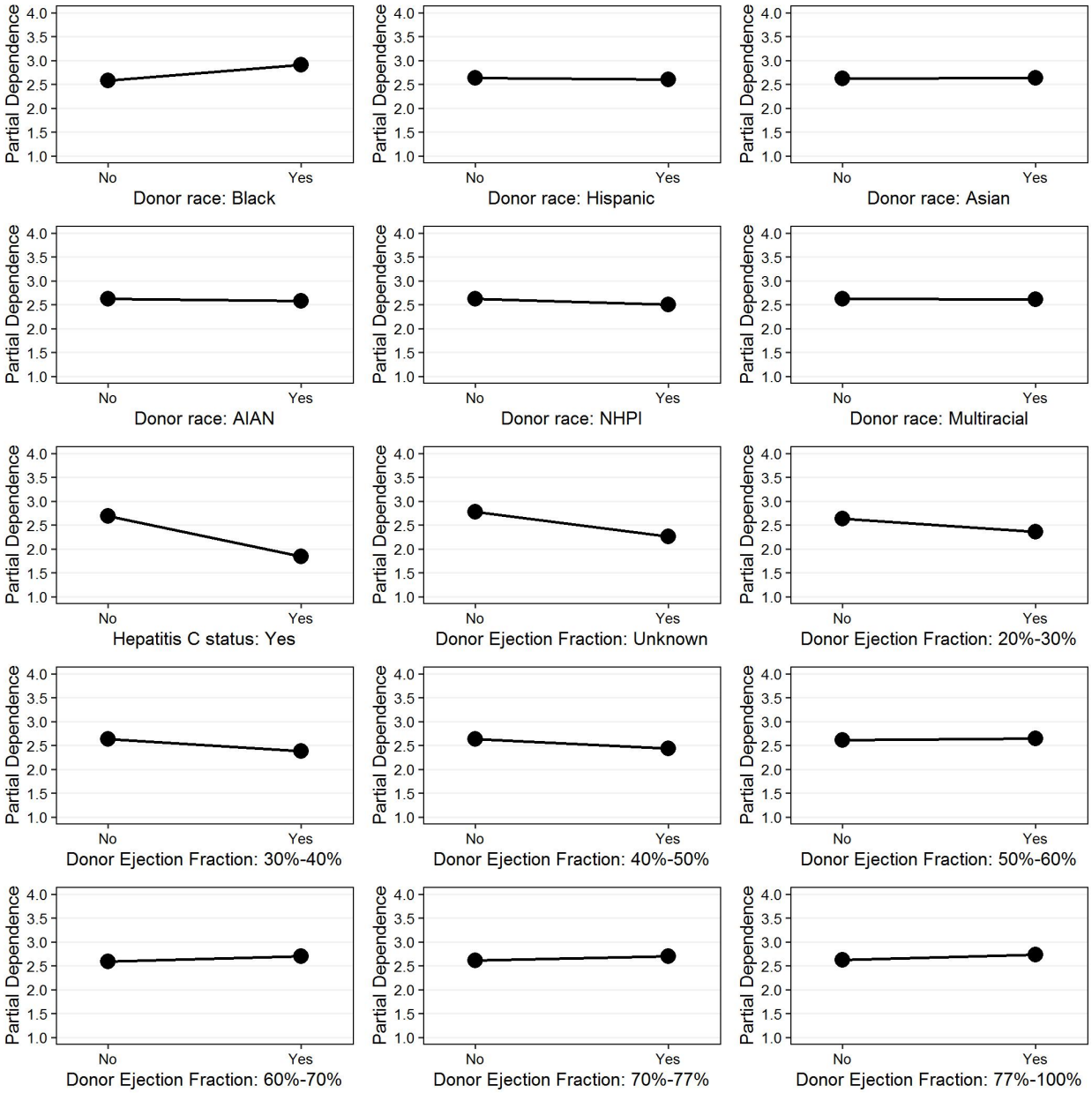


Figure 2.13: Partial dependence plots of the main categorical predictors of overall deceased donor organ yield.

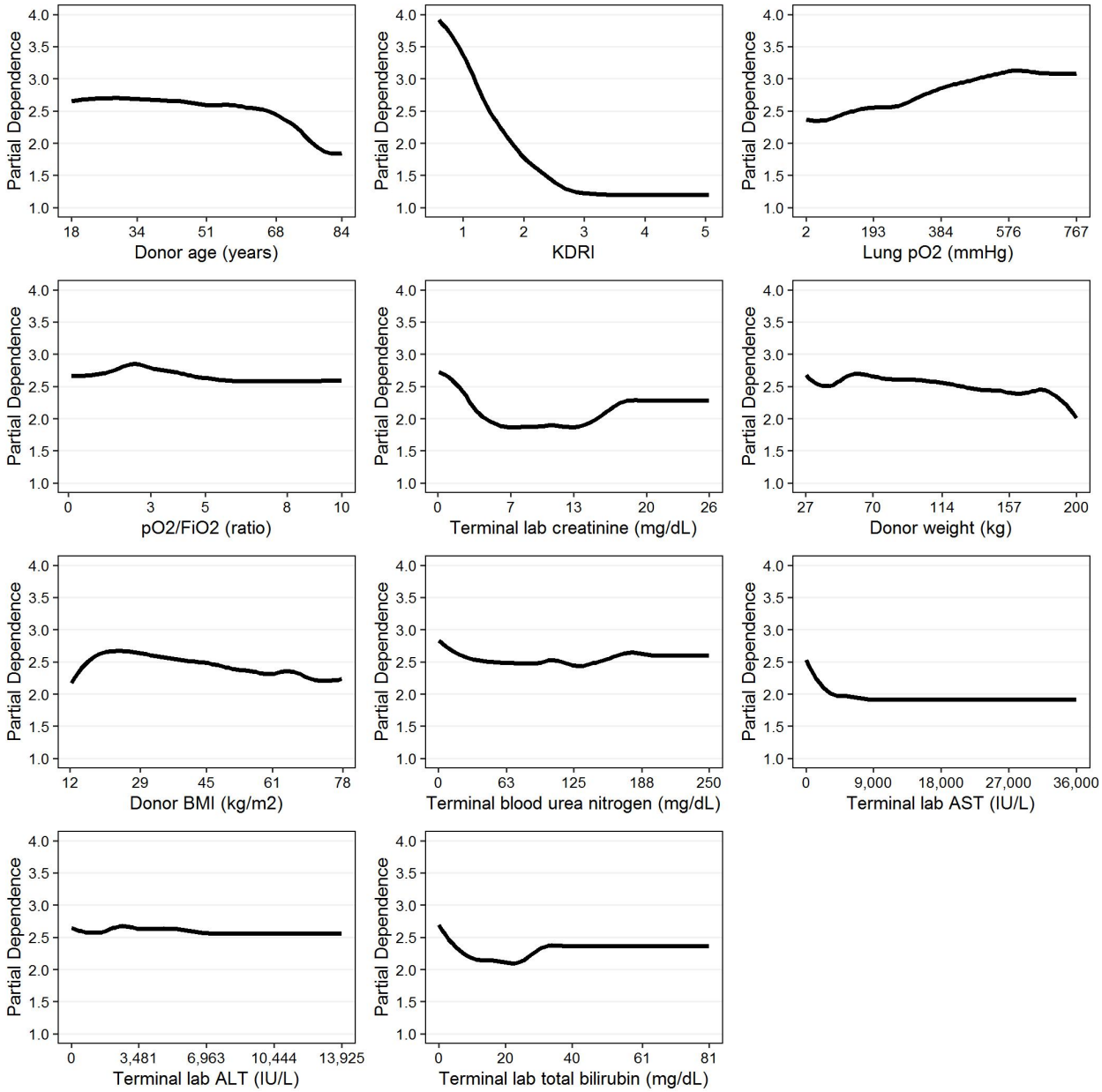


Figure 2.14: Partial dependence plots of the main numerical predictors of deceased donor organ yield. Points in the horizontal axis of the plots represent the empirical quartiles of each predictor.

2.7.2.5 Sensitivity Analyses

The imputation of the missing data of the donors originally excluded in our analysis does not substantially alter our findings (Figures 2.15 and 2.16). All models, except for Poisson regression and NB regression (99.6% CI -0.002 to 0.003, $P = 0.549$), are significantly different from each other in terms of their MAE (all $P < 0.001$). Nonetheless, gradient boosting is not significantly different from BART (99.6% CI -0.007 to 0.004, $P = 0.412$) in terms of the MSE. Combining these results, gradient boosting methodology outperforms the remaining models. Gradient boosting obtains an average (SD) MAE of 0.737 (0.003) and an MSE of 0.892 (0.003) over the Monte Carlo cross-validation analysis, and a MAE of 0.741 and an MSE of 0.902 when fit to the derivation set ($n = 94,694$) and examined in the validation set ($n = 40,583$).

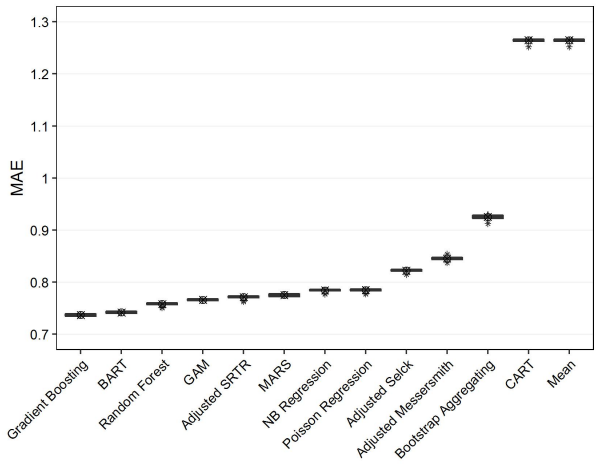


Figure 2.15: MAE of potential models to predict overall deceased donor organ yield using imputed dataset. Points in the center of the boxes represent the mean MAE of the statistical models and asterisks represent outliers. The MAE of the ANN models are excluded for illustration purposes. ANN presented a median MAE of 2.670 with an interquartile range of 1.028.

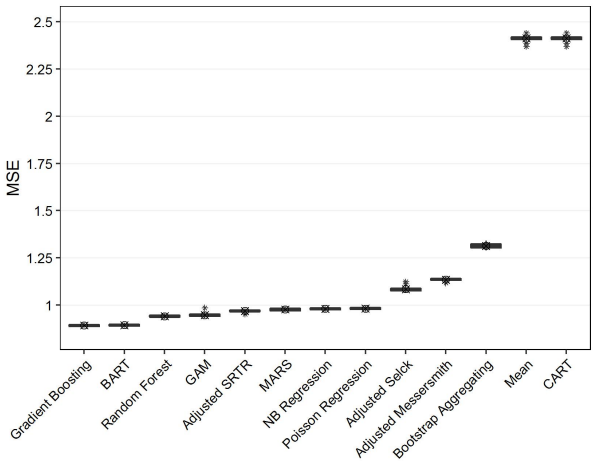


Figure 2.16: MSE of potential models to predict overall deceased donor organ yield using imputed dataset. Points in the center of the boxes represent the mean MSE of the statistical models and asterisks represent outliers. The MSE of the ANN models are excluded for illustration purposes. ANN presented a median MSE of 9.471 with an interquartile range of 8.206.

The results of the 2-year cohort sensitivity analysis can be found in Figures 2.17 and 2.18. We observe that training the models on smaller cohort sizes do not change our conclusions considerably. Since in this sensitivity analysis we are comparing 15 models simultaneously, the Bonferroni correction threshold results in 0.003. We find that gradient boosting has a significantly lower MAE than the rest of the models in cohorts 2013-2014

and 2017-2018 (all $P < 0.001$). In cohort 2015-2016, gradient boosting does not significantly outperform BART (99.7% CI -0.002 to 0.003, $P = 0.003$). The MSE of gradient boosting is not significantly different from BART in any cohort (99.7% CI -0.023 to 0.013, $P = 0.328$; 99.7% CI -0.022 to 0.014, $P = 0.633$; and 99.7% CI -0.013 to 0.012, $P = 0.820$ for cohorts 2013-2014, 2015-2016, and 2017-2018, respectively). We also notice that the MSE of gradient boosting is not significantly different from the MSE of GAM (99.7% CI -0.007 to 0.030, $P = 0.061$) in the 2013-2014 cohort.

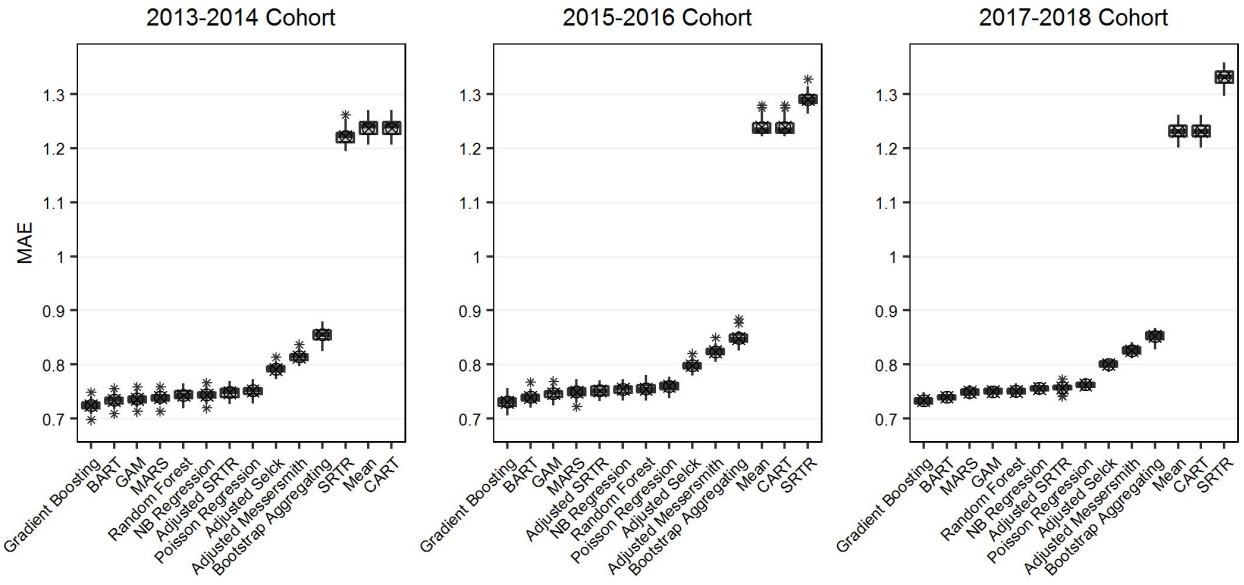


Figure 2.17: MAE of potential models to predict overall deceased donor organ yield using imputed dataset. Points in the center of the boxes represent the mean MAE of the statistical models and asterisks represent outliers. The MAE of the ANN models are excluded for illustration purposes. The median (interquartile range) MAE of the ANN models in cohorts 2013-2014, 2015-2016, and 2017-2018 are: 2.431 (0.615), 2.748 (0.719), and 2.398 (0.431), respectively.

2.7.3 Discussion

A model to accurately predict overall deceased donor organ yield can serve as an aid to assess organ procurement performance. Previous studies in organ yield have focused on modeling the overall organ yield with generalized linear models and splines. While these conventional approaches have good predictive performance, we found that alternative approaches can achieve higher accuracy.

Our study evaluated 15 statistical models, and a tree-based gradient boosting was able to obtain the best overall performance according to the MAE and MSE. Tree-based meth-

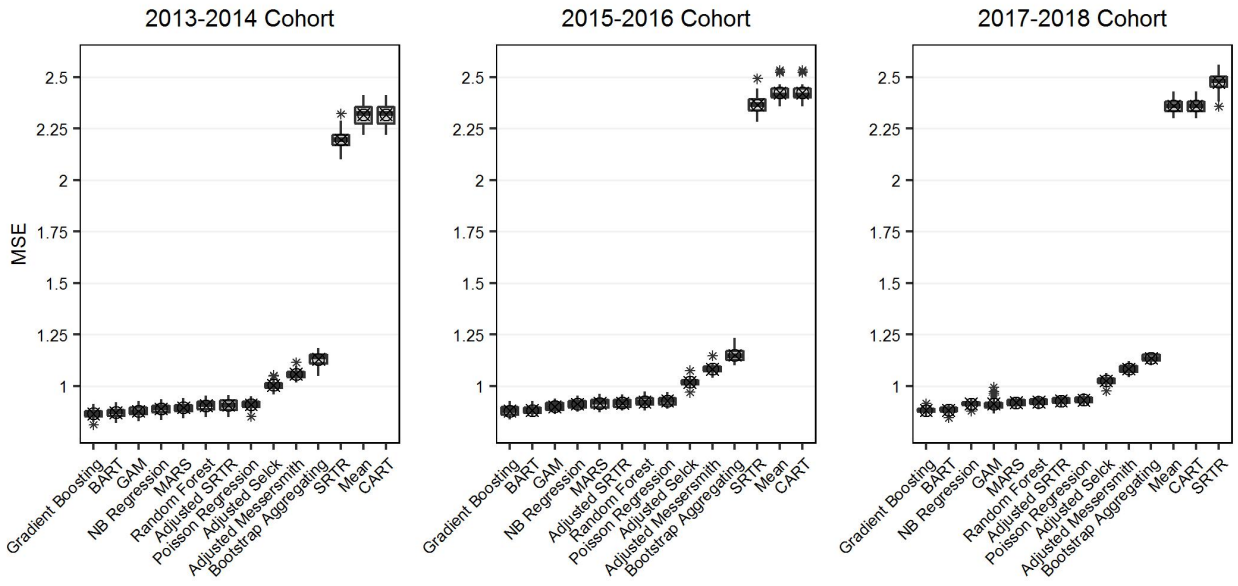


Figure 2.18: MSE of potential models to predict overall deceased donor organ yield using imputed dataset. Points in the center of the boxes represent the mean MSE of the statistical models and asterisks represent outliers. The MSE of the ANN models are excluded for illustration purposes. The median (interquartile range) MSE of the ANN models in cohorts 2013-2014, 2015-2016, and 2017-2018 are: 8.126 (3.371), 9.971 (4.611), and 7.957 (2.483), respectively.

ods model the association between the predictors and overall organ yield without making any assumptions of their underlying relationship. Besides easily integrating non-linear relationships, tree-based methods can implicitly capture interactions among predictors. The tree-based gradient boosting algorithm reduces prediction error by averaging over the predictions of many sequentially built trees. If calibrated properly, this modeling methodology has been previously shown to have excellent performance (Caruana and Niculescu-Mizil, 2006).

We noticed that BART, random forest, and GAM consistently outperformed the organ yield models previously developed in the literature. These models were built using ordinary least squares (Selck et al., 2008), ordinal logistic regression (Messersmith et al., 2011), and a sum of logistic regression models (Scientific Registry of Transplant Recipients, 2019). Beyond gradient boosting, our study suggests that other machine learning algorithms can also achieve better predictive performance than the previous approaches in estimating deceased donor organ yield.

Since the non-linear machine learning algorithms, such as gradient boosting, often exhibit high variability in their predictions and require large amounts of data, our main analysis included data from 2000 to 2018. We observed that the variability of most models increased in our 2-year cohort sensitivity analysis. Nevertheless, gradient boosting still outperformed the remainder of the models. In general, the more data the algorithms use during the learning process the more reliable their predictions become.

The two most impactful factors on organ yield in the gradient boosting model were KDRI, which is used commonly in kidney donor evaluation, and lung pO2, which could be a proxy for cardiopulmonary status of the donor. The remainder of the factors impacted performance to a lesser extent.

We showed that model performance has practical implications. Better model performance achieves higher accuracy on the prediction of organ yield per 100 donors at the individual and aggregated level. Accurate predictive ability is critical when evaluating performance in donor conversion and organ yield. Identifying outliers in deceased donor organ yield and improving achievement of donor management goals, are important steps towards increasing the number of organs available for transplantation (Malinoski et al., 2011). The increasing availability of novel technologies, such as organ machine perfusion, will likely have significant impacts on organ yield models in the future (Moers et al., 2009).

A noteworthy strength of our work is the examination of the predictive ability of conventional and novel machine learning models in a large dataset containing over 89,000 donors across a 19-year span. Another strong point of the present study is the use of an innovative approach to predict overall deceased donor organ yield.

While our study has several strengths, it also contains some limitations. The major shortcoming of this study is the missingness of data inherent to the OPTN database. However, we showed that our results were robust in our sensitivity analysis after imputing missing data. Additionally, we were only able to analyze data collected in the OPTN database and there may be other uncollected donor and transplant-center specific information that may impact overall organ yield.

Another notable limitation of our analysis is that our models were not built to predict organ-specific yield. This was not the objective of our analysis, as the organ-specific yield models developed by the SRTR are updated frequently making contemporary comparisons difficult (Scientific Registry of Transplant Recipients, 2019). The addition of individual organ yield predictions may introduce additional error in the prediction of the overall yield. This is a potential reason why the direct application of the SRTR model was outperformed by multiple parametric and non-parametric models in the 2-year cohort analysis. Our study has highlighted the importance of predicting overall yield in the aggregated scale to avoid unnecessary variability and obtain insights into the relationship with its predictors. Future studies aimed an evaluating novel models in individual organ yield prediction may be warranted based on the results of this analysis.

2.8 Conclusions

This chapter presented data-driven methods to understand how the expected population changes in the US may impact future liver availability and demand. In addition, we introduced a machine learning methodology to improve the prediction of deceased donor organ yield. Improving the accuracy of organ yield prediction may serve as an initial step towards understanding how to increase donor potential and narrow the breach between organ supply and demand.

In Section 2.4, we projected a further exacerbation of the donor shortage for liver transplantation over the next decade, with total population growth outstripping the grown in potential donors. Changes in the proportion of obese US adults over the next decade will significantly impact the number of available donors. Our study can serve as an objective guide so that steps can be taken for future planning to help alleviate the mismatch between liver donors and recipients to prevent waitlist dropout.

Our analysis in Section 2.5 showed that a lag of 9 years best explains the rise in obesity in the US population and the rise in NASH related additions to the waitlist. Using this lag and the anticipated increase of obesity in the US population, we projected a 71.9% increase in NASH related transplant waitlist additions. This has several public health

implications for the transplant community and for the overall burden of NASH-related liver disease in the US (Dulai et al., 2017). It is especially worrisome in the setting of a plateauing donor supply, making receipt of a LT more difficult for those on the waitlist (Young et al., 2016; Orman et al., 2013; Parikh et al., 2015). Continued public health efforts to curb obesity prevalence and improvement in the diagnosis of and treatment of NASH will be important to mitigate the overall impact of our projections.

In Section 2.6, we found that redistricting regional allocation will result in decreased geographic inequity in LT, when measured from a population perspective. However, improving regional donation and utilization rates may have a more impactful effect. A hybrid approach to balancing LT supply and demand would likely yield the best results. That is, some form of redistricting coupled with systematic study and improvements in donation/utilization rates in low performing DSAs.

Finally, we observed that a tree-based gradient boosting outperformed the existing models for deceased donor organ yield in Section 2.7. Improving accuracy of donor yield production could allow for better evaluation of OPO performance in donor conversion. Improving OPO performance is key to mitigating the organ shortage and ultimately reduce morbidity and mortality in patients awaiting organ transplantation.

Chapter 3

Understanding the Role of Genetic Information in Cardiovascular Treatment Planning

In this chapter, we present the development of cholesterol treatment plans using a risk-based threshold policy and policies derived with a MDP. We also introduce a simulation-based framework to estimate the risk of a CHD event due to clinical and genetic factors. Treatment plans using the threshold policies will be obtained using the risk of ASCVD events with and without genetic factors. The treatment plans obtained with the MDP models are derived using, clinical information only as well as clinical and genetic information. Lastly, we propose a framework to obtain simultaneous cholesterol treatment plans and genetic testing strategies by combining dynamic programming with VoI analysis. The implication of the policies at a population level are evaluated using a thoroughly validated simulation framework.

3.1 Background

Over the last few decades, biomedical research has focused its attention in identifying which genes are responsible for trait variation in humans (Kathiresan and Srivastava, 2012). Genetic variants have been associated with several conditions such as cardiovascular disease, celiac disease, type 1 diabetes, Crohn's disease, and cancer, among others (The CARDIoGRAMplusC4D Consortium, 2013; Abraham et al., 2013, 2014; Sun et al., 2007). The National Institutes of Health (NIH) have suggested clinicians to use additional tools to improve their medical assessments and ability to identify patients at risk of certain diseases (Vasan, 2006). Genetic tests are recently developed tools that could be used with such purpose. A GRS could help characterize individuals who might receive the greatest benefit from treatment (Mega et al., 2015). However, a test to obtain a GRS may represent an additional cost to patients and not every patient may benefit from such a test. Under-

standing which patients might benefit the most from a genetic test and when to perform such test are key questions faced by clinicians.

Performing genetic testing too early in the life of individuals or universal testing may not be practical, as its cost might outweigh its benefits. Conversely, not performing genetic testing or conducting it too late in the life of patients may lead to wrong treatment decisions, as clinicians often determine how to manage their patients based on their estimated levels of risk (Bibbins-Domingo et al., 2016; Goff et al., 2014). Therefore, patients should receive genetic testing depending on whether or not their treatment plans could potentially be improved upon the new risk information.

Although it is known that ASCVD has familial and genetic components, understanding the role of genetic testing in the prevention and treatment of cardiovascular diseases has been limited (MacRae and Vasan, 2016; Jarmul et al., 2018). However, the development of the new GRS that helps predict ASCVD and the benefit of treatment could make genetics central to daily cardiovascular care (Mega et al., 2015; Khera et al., 2016; Natarajan et al., 2017).

While genetic risk scores have recently become available, they are not commonly performed yet (Lewis and Vassos, 2020). However, medical practitioners and geneticists estimate that the GRS that helps predict ASCVD will soon be regularly used in clinical practice (Knowles and Ashley, 2018; Connor et al., 2020). Currently, genetic risk scores are mostly available as a mail-in saliva tests from direct-to-consumer genetics companies (Lewis and Vassos, 2020). Patients send a saliva sample to a laboratory to start the genetic test. Once the results are available, patients receive a report for them to share with their primary care physician. Patients are usually provided with consultations to better understand the results of the test and what questions they should discuss with their primary care physician, though this can vary between companies.

The risk for ASCVD events is mainly composed of demographic, behavioral, and clinical factors. One controllable risk factor for ASCVD events is the cholesterol level of patients Goff et al. (2014). Incorporating a GRS to the risk for ASCVD events would improve risk prediction. Improving risk prediction could in turn help identify which individuals may benefit from cholesterol treatment and how patients should be treated (Sussman et al., 2017). But since this information would only benefit a subset of the population, whether and when to gather genetic information and its role in cholesterol treatment planning is unclear.

3.2 Organization of the Chapter

This remainder of this chapter is organized as follows. In Section 3.3, we provide a review of the relevant literature. Section 3.4 introduces a decision analytic model to compare cholesterol treatment strategies with and without genetic information as well as our parameters. Particularly, this section compares the cholesterol treatment plans based on clinical guidelines. In Section 3.5, we build upon the decision analytic model and develop a rigorously validated simulation model to evaluate the impact of genetic information on cholesterol treatment policies. These policies are derived using MDP models informed with clinical information only and with clinical and genetic information. The genetic testing strategies made in Sections 3.4 and 3.5 are based on patients' risk for ASCVD events. Finally, in Section 3.6, we use the simulation model introduced in Section 3.5 to integrate the MDP models with VoI analysis to develop simultaneous cholesterol treatment plans and genetic testing strategies. Figure 3.1 shows the organization of this chapter after the background and literature review.

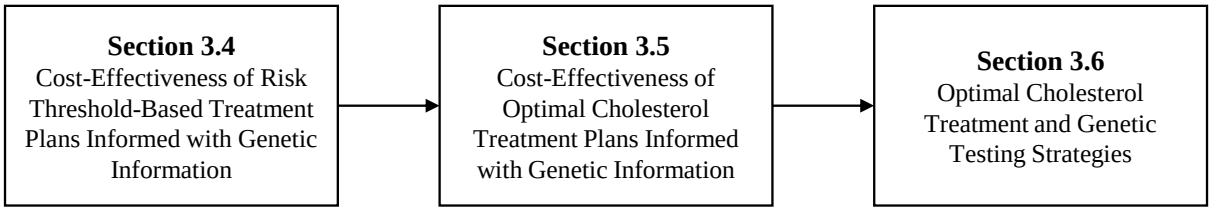


Figure 3.1: Organization and connections between sections in Chapter 3.

3.3 Literature review

The relevant literature to this research lies in the following fields: (1) risk prediction models that incorporate genetic information; (2) models to quantify VoI in healthcare applications; (3) treatment decision models; and (4) models that incorporate testing and treatment decisions simultaneously. In this section, we highlight some prominent papers in each category and briefly describe how our proposed methodology differs from them.

Risk prediction models that incorporate genetic information have been previously developed in the literature. Sun et al. (2007) presented a model to improve the prognosis of breast cancer using both clinical and genetic markers. Abraham et al. (2014) created

a predictive model for Celiac disease based on genome-wide single nucleotide polymorphisms profiles. The authors also presented an extensive analysis of multiple statistical models for the prediction of Celiac disease, type 1 diabetes, and Crohn's disease using genome-wide single nucleotide polymorphisms profiles (Abraham et al., 2013). Although these models use genetic information to predict a risk for adverse events, none of these models have been used to improve decision making.

Given the generality of VoI analysis, it has been applied in a variety of fields including health policy, medical care, clinical trials, environmental health, and toxicology, among others (Cipriano and Weber, 2018; Yokota and Thompson, 2004; Ozcan, 2005). Comprehensive surveys of VoI analysis in healthcare applications include (Steuten et al., 2013) and (Heath et al., 2017). Several solution strategies have been proposed in the literature including simulation, analytical, and discretization methods (Thompson and Yokota, 2004). Among these, simulation has been one of the most commonly used approaches (Yokota and Thompson, 2004; Claxton et al., 2001; Felli and Hazen, 1998, 1999). In this chapter, we have also chosen to use simulation to quantify the VoI of genetic testing. Nonetheless, our method differs from the methods presented in these papers in that we quantify VoI in a sequential framework.

The value of sequential information has also been widely studied (Miller, 1975; Pozzi and Der Kiureghian, 2011; Eckermann and Willan, 2008; Memarzadeh and Pozzi, 2016). Similar to our work, previous research have formulated a system of interest as a Markov model and calculated the expected value of perfect information of interventions using Monte Carlo simulation (Dong et al., 2007; Martikainen et al., 2005; Ginnelly et al., 2005). Dong et al. (2007) modeled the total knee replacement procedure as a Markov model and calculated expected value of perfect information of a computer-assisted total knee replacement procedure using Monte Carlo simulation. Martikainen et al. (2005) formulated the progression of recurrent glioblastoma multiforme as a Markov model. The authors evaluated the value of new information on reducing uncertainty related to the choice of treatment using the expected value of perfect information and a second-order Monte Carlo simulation. Ginnelly et al. (2005) modeled the frequency of recurrent urinary tract infections as a Markov process. In their study, Ginnelly et al. (2005) used VoI analysis to quantify the cost of uncertainty associated with the choices of therapy for urinary tract infections. While these studies used VoI analysis to quantify the benefit of new interventions, we extended this concept by designing strategies of when to perform such interventions based on VoI analysis.

Existing treatment decision models include Long et al. (2008); Lee et al. (2008); Chen et al. (2018); Chan et al. (2013); Long et al. (2018); Liu et al. (2017) and Negoescu et al.

(2017). Overviews of disease management decision models can be found at (Denton et al., 2011), (Capan et al., 2017), and (Saville et al., 2018). Researchers have also developed treatment decisions models for patients at risk of cardiovascular diseases (Cooper et al., 2006; Stanford R.E., 2004; Zargoush et al., 2018). Hauskrecht and Fraser (2000) modeled ischemic heart disease treatment plans with a partially observable Markov decision process partially observable Markov decision process (POMDP). Denton et al. (2009); Kurt et al. (2011) constructed an MDP model to determine the optimal time for starting cholesterol medications in patients with diabetes. Mason et al. (2014) extended this work by determining the optimal timing of blood pressure and cholesterol medications in patients with diabetes. Schell et al. (2016) developed a method to approximate optimal hypertension treatment policies derived with an MDP. Nevertheless, none of these models incorporate genetic information to guide treatment plans.

Several studies have focused on finding the optimal time to gather additional information, perform a test, or screening procedure including Hicklin et al. (2018); Chhatwal et al. (2010); Suen et al. (2018); Agnihothri et al. (2018); Onen et al. (2018); Ayer et al. (2016); Lee et al. (2018); Deo et al. (2015); Helm et al. (2015); Maillart et al. (2008); Zhang et al. (2012); Erenay et al. (2014); Skandari et al. (2015); Sabouri et al. (2017) and Lin et al. (2018). Among these, Hicklin and coauthors developed Bayesian decision models to determine when it is suitable to gather additional information before deciding to end natural labor and perform a C-section. The authors then used VoI analysis to quantify the benefit gained by having access to observations to inform decision making. Even though the models presented in these papers aim to find the best time to gather additional information, perform a test, or screening procedure, none of them incorporate VoI in a sequential decision making framework. Other studies have estimated the cost-effectiveness of testing and treatment interventions (Hutton et al., 2007; Hassmiller Lich et al., 2017; Leshno et al., 2003; Chirikos, 2003; Lin et al., 2019).

Models that incorporate testing and treatment decisions are also available in the literature (Robins et al., 2008; Kirkizlar et al., 2010; Kazemian et al., 2018; Yang et al., 2013; Ghamat et al., 2017; Harper and Jones, 2005). Robins et al. (2008) presented a dynamic marginal structural model to jointly develop optimal testing and treatment regimes. Kazemian et al. (2018) provide a linear quadratic Gaussian state space model formulation to determine how to monitor and control chronic diseases. The authors applied their modeling framework for the management of glaucoma. In addition, Yang et al. (2013) developed a dynamic programming framework to find screening and treatment policies for childhood obesity. Another study relevant to our research was presented by Negoescu et al. (2017). The authors used a continuous-time, multi-armed bandit setting to determine

optimal treatment policies and treatment discontinuation thresholds in chronic diseases. These studies differ from ours in that none of them use VoI analysis in their decision making formulation. By incorporating VoI analysis in our formulation, we determine which patients should receive genetic testing and when they should receive such a test.

3.4 Cost-Effectiveness of Risk Threshold-Based Treatment Plans Informed with Genetic Information

Risk prediction is playing a growing role in the prevention of ASCVD. Since 2013, multiple clinical practice guidelines have recommended that the use of cholesterol-lowering statin drugs should be guided by estimates of a patient's risk of developing ASCVD (Goff et al., 2014; Bibbins-Domingo et al., 2016; Downs and O'Malley, 2015). Research has found that the benefit of statin drugs is nearly constant for all levels and causes of risk for ASCVD events Collins et al. (2016). Any tool that improves those estimates could also improve statin use (Sussman et al., 2017). In this way, genetic risk could be used as an independent risk factor, much like smoking status (Abraham et al., 2016). But to date it remains unknown if the effects are large enough and the costs low enough for using risk scores that use GRS in clinical practice to be practical.

The objective of this study is to determine the potential public health impact and cost-effectiveness of large-scale genetic testing to inform the use of statin drugs, based on clinical guidelines. We also aim to discover in which patients genetic testing is most cost-effective, and to examine how much improvements in genetic testing might make testing more practical.

3.4.1 Methods

We develop a sample that is representative of Americans with no history of ASCVD who would be eligible for the USPSTF statin guidelines. We then estimate ASCVD event rates using the pooled cohort equations (PCE) risk scores (the current standard) and new PCE that are informed by genetic risk. (The PCE and GenePCE strategies, respectively.) To find a group for whom genetic testing is most likely to influence their care, we propose testing anyone whose care recommendation would be changed if their GRS is outside one standard deviation from the norm. Following the USPSTF guidelines, we assume people who have a 10-year estimated risk above 10% are treated by a moderate-intensity statin and those who have a history of heart attack or stroke are treated with a high-intensity statin. A decision analytic model is developed to compare guiding care with the PCE risk

vs. the GenePCE risk. Finally, we estimate the lifetime effects of following either of the treatment strategies for 10 years on every patient in the population.

3.4.1.1 Population

To understand the clinical and policy implications of genetic testing, we aim to create a simulated population that resembles all Americans who would meet our criteria for genetic testing (summarized in Figure 3.2). We used NHANES from 2009 to 2014, which is a nationally representative study sample of the US population (Centers for Disease Control and Prevention, 2020). Our primary sample is composed of adult White or African-American patients from 40 to 75 years old with no history of myocardial infarction, stroke, or congestive heart failure and LDL of at most 190 mg/dL.

To do this, we impute missing data with the MissForest package in R (Stekhoven and Buhlmann, 2012). We then create a synthetic dataset based on the NHANES sampling weights that represent the guideline-eligible population. We use nonparametric classification and regression trees and the synthpop package in R. This process samples from the dataset's probability distributions so that the main statistical features of the original data are preserved (Nowok et al., 2016). Individuals with LDL greater than 190 mg/dL are excluded.

Since many risk factors change over time, we estimate the progression in risk factors over the 10-year time horizon. To do this, we used linear regression with systolic blood pressure (SBP), high-density lipoprotein (HDL), LDL, and total cholesterol (TC) as outcome variables (Table B.1 in Appendix B.1). We predict this change using the variables in the PCE risk score and cholesterol treatment status. The intercept term of each regression model is adjusted by applying the difference between the linear regression fitted value and the observed value in the Continuous NHANES data. We select the models using stepwise variable selection. If an individual is being treated with either atorvastatin or rosuvastatin, we code this variable as a high intensity statin, all other statins are coded as moderate intensity.

3.4.1.2 Treatment Strategies

Our treatment strategies are based on the 2016 USPSTF statin guidelines (Bibbins-Domingo et al., 2016). In these guidelines, with a risk of at least 10% is recommended at least a medium strength statin.

In the PCE strategy, patients are started on moderate-intensity statins when their 10-year PCE risk reached 10% (Goff et al., 2014). The GenePCE strategy mirrored the PCE

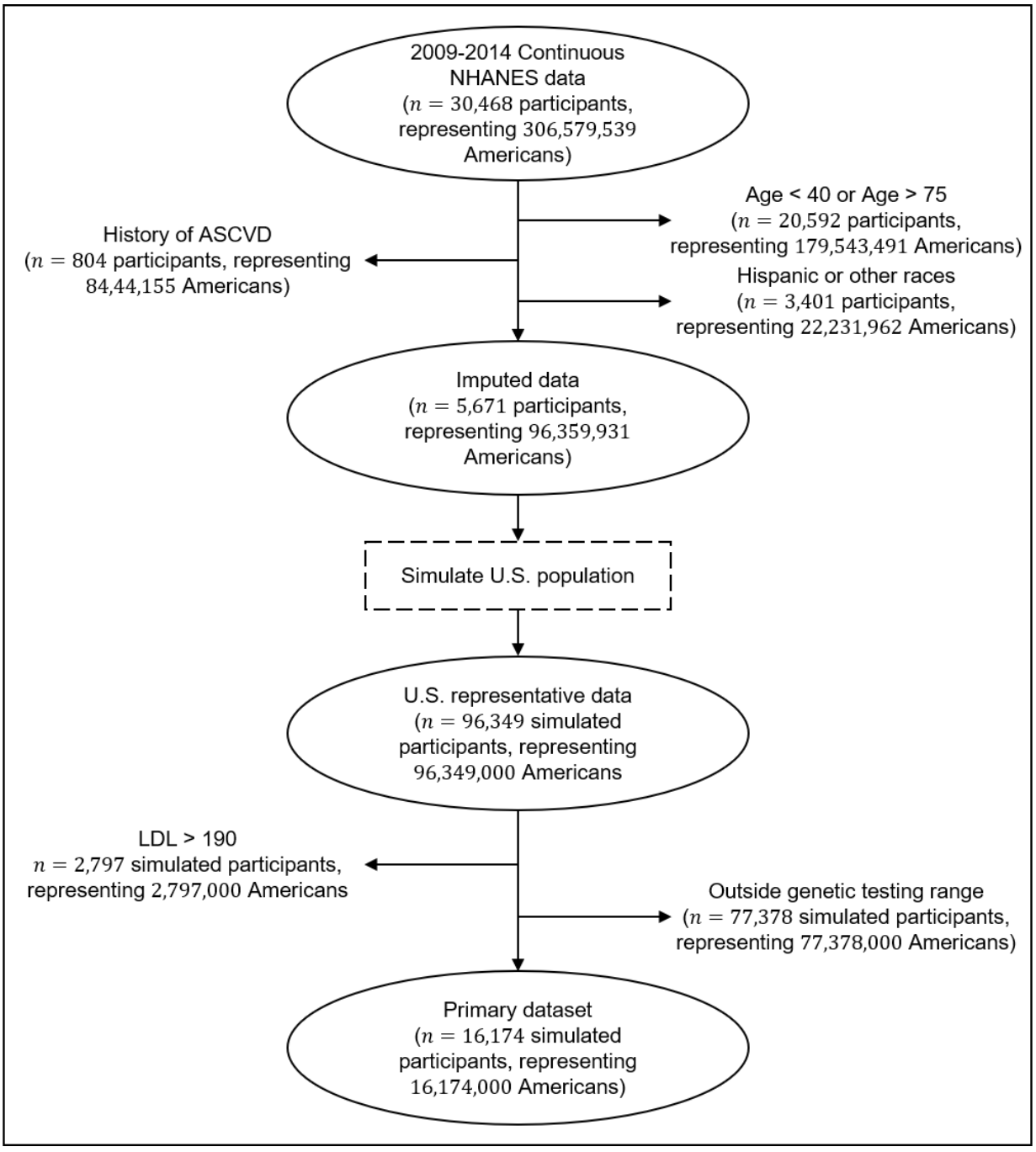


Figure 3.2: Flow diagram of the study dataset.

strategy, except the treatment is guided by the patient's GenePCE risk, which incorporates the patient's genetic risk with the PCE risk. We define genetic risk with a GRS, a combination of individual single-nucleotide polymorphisms that are independently predictive of CHD, each weighted by the strength of their effect. ASCVD risk increases by an odds ratio of 1.67 per standard deviation of genetic risk (Khera et al., 2018; Abraham et al., 2016; The CARDIoGRAMplusC4D Consortium, 2013). Genetic risk is nearly independent of the PCE risk, so we can accurately incorporate the GRS into the PCE (Khera et al., 2016, 2018; Natarajan et al., 2017). The GRS only alters the CHD component of the PCE risk prediction. See Section 3.6.1.2 for details on the development of the GenePCE score.

Patients in the GenePCE evaluation would be recommended genetic testing if their genetic risk has a reasonable chance of changing their treatment at the beginning of the study period. We define this as anyone for whom a GRS outside 1 standard deviation from the mean would alter their care. This proves to be anyone with a PCE risk score between 7.68% and 13.63% at the first year of the simulation.

However, when applied directly, a different number of patients is recommended treatment using the GenePCE score than the PCE score. Since we want to isolate the effects of genetic testing, not the effects of statin use, we equalize the amount of statin treatment between arms. We discover a treatment cut-point for the GenePCE score of 10-year GenePCE risk above 8.47% did this. See Appendix B.2 for details on the identification of the GenePCE treatment threshold. Anyone with history of CHD or stroke is continued to be treated with a high-intensity statin.

3.4.1.3 States of the Model

At the beginning of each year, we assume that each patient could be in one of ten health states throughout the 10-year planning horizon. The ten health states are: (1) healthy (no history of CHD or stroke); (2) history of CHD but no adverse event this period; (3) history of stroke but no adverse event this period; (4) history of CHD and stroke but no adverse event this period; (5) survived a CHD event this period; (6) survived a stroke this period; (7) death from a non-cardiovascular disease related cause; (8) death from CHD event this period; (9) death from stroke this period; and (10) dead. All patients are in the healthy state at the beginning of the first year. Each patient within each state has different transition possibilities, quality of life weights, and costs based on their risk information.

3.4.1.4 Transition Probabilities

Every patient has a 1-year event risk for every possible event at each year of the 10-year planning horizon (Goff et al., 2014; Mega et al., 2015). We assume that 60% of the PCE risk is from CHD events and 40% is from stroke events (Benjamin et al., 2017; Brønnnum-Hansen et al., 2001; Burn et al., 1994).

As in previous research, we assume that second events are more common than would be predicted by the PCE score alone. We multiply the patient's CHD odds by 3 if the patient has history of CHD, multiplied the stroke odds by 2 if the patient is at least 60 years old and has history of stroke, and multiplied the stroke odds by 3 if the patient has history of stroke and is less than 60 years old (Brønnnum-Hansen et al., 2001; Burn et al., 1994). The probability of having a new event is altered by statin use.

For people who have an event in a given year, we calculate the probability that the event would be fatal by applying fatality likelihoods for CHD and stroke events to the post-treatment risk of ASCVD events (Lloyd-Jones et al., 2009; Miniño et al., 2007; Sussman et al., 2011, 2013). These are developed from the National Center for Health Statistics rates of known fatal event compared to the overall event rates predicted by the PCE score, adjusted for age and gender (NCHS, 2017). Mortality from second cardiovascular events and known overdiagnosis of cardiovascular diseases is accounted for by decreasing the fatal event rates reported by the National Center for Health Statistics by 50% (Govindan et al., 2014). In addition, we calculate the probability of non-ASCVD mortality using life-tables (Arias et al., 2017).

3.4.1.5 Treatment Benefit

The estimates of the effects of statins on ASCVD events are derived from Collins et al. (2016). This study finds that the average LDL reduction from a moderate-intensity statin would be 37% and from a high-intensity statin would be 53%. For every 40% reduction in LDL the risk of ASCVD would decrease by 21%. We investigate the impact of statin benefit misestimation by allowing discontinuation and restarting of treatment in a sensitivity analysis.

3.4.1.6 Input variables

The quality of life for each health state is defined in Table 3.1. We assume the disutility of statin use incorporates all side effects and any increased rate of diabetes (Fryback et al., 1993; Pignone et al., 2006; Pignone, 2007). We calculate the lifetime impacts of these choices. To do this, we use the patient's estimated lifespan from life tables (Arias et al.,

2017), and adjust the survival estimates for patients who have history of ASCVD events (Pandya et al., 2015; Smolina et al., 2012; Dennis et al., 1993).

Table 3.1: Input variables.

Health State	Value	Source
Utility Weights		
Healthy	1	Fryback et al. (1993); Pignone et al. (2006); Pignone (2007)
History of CHD or stroke but none this year	0.90	Fryback et al. (1993); Pignone et al. (2006); Pignone (2007)
History of CHD and stroke but none this year	0.81	Fryback et al. (1993); Pignone et al. (2006); Pignone (2007)
Non-fatal CHD event	0.88	Fryback et al. (1993); Pignone et al. (2006); Pignone (2007)
Non-fatal stroke	0.67	Fryback et al. (1993); Pignone et al. (2006); Pignone (2007)
Dead	0	Fryback et al. (1993); Pignone et al. (2006); Pignone (2007)
Moderate intensity statin	0.001	Pandya et al. (2015)
High intensity statin	0.002	Pandya et al. (2015)
Costs ($/year)		
Non-fatal CHD event		
Year of the event	$67,155	O'Sullivan et al. (2011)
Subsequent years	$4,499	Medical Expenditure Panel Survey (2015)
Non-fatal stroke		
Year of the event	$22,143	O'Sullivan et al. (2011)
Subsequent years	$7,100	Medical Expenditure Panel Survey (2015)
Fatal CHD event	$18,634	O'Sullivan et al. (2011)
Fatal stroke	$11,495	O'Sullivan et al. (2011)
Genetic testing (one time only)	$200	Personal communication
Moderate intensity statin	$144/year	GoodRx (2017)
High intensity statin	$450/year	GoodRx (2017)

All cost parameters are adjusted for inflation from the original citations and all expenses and quality of life weights are discounted by 3% (O'Sullivan et al., 2011; Medical Expenditure Panel Survey, 2015).

3.4.1.7 Analysis

The number of times we run our simulation model is derived empirically. We initially run the base case of our simulation 2,000 times. Then, we take random samples of 1 to 1,500 replications and calculate the mean of the incremental cost-effectiveness ratio (ICER) of the GenePCE strategy compared to the PCE strategy. Summary statistics, survival rates, and survival years of the population are then calculated. Afterwards, we examine how genetic testing changes the treatment strategies of the patients during the first year of our study. We also estimate the lifetime amount of treatment, ASCVD events averted, QALYs saved, costs incurred, and cost per QALY saved following both, the PCE and GenePCE treatment strategies using a 10-year planning horizon. These results are presented for the whole population that received genetic testing and by treatment aggressiveness. All results are estimated by averaging over all simulation replicates.

3.4.1.8 Sensitivity Analyses

We perform sensitivity analyses on many of the parameters under study. To assess the robustness of the GenePCE strategy, we consider the potential impact on health and spending if the GenePCE event rate is incorrect. One non-traditional analysis is the development of the AdjustedGenePCE strategy. For this strategy, we incorporate a genetic adjustment to the relative benefit of statins, in which patients with high genetic risk have an additional improvement in relative statin benefit (Natarajan et al., 2017; Mega et al., 2015).

We also perform a sensitivity analysis on the treatment threshold used for the GenePCE treatment strategy. In our base-case, we establish cut-points that ensure that the PCE and GenePCE provide the same amount of statin use over the 10 years of the study. In sensitivity analyses, we match the amount of treatment during the first year of our study, called equal initial treatment. We test a treatment threshold of 10% for both treatment strategies. We also vary the odds ratio per standard deviation of the genetic score, one based on Abraham et al. (2016) and one that hypothesizes continued improvement in GRS effectiveness. In addition, we examine the case that decisions in the GenePCE strategy are made according to the GRS effectiveness as estimated by Khera et al. (2018). While the true GenePCE event rate followed the effectiveness of GRS as per Abraham et al. (2016), namely the wrong GRS effectiveness scenario.

We vary statin cost, treatment related disutility of statins use, and incorporate the effect of statin discontinuation and restarting in our sensitivity analyses based on Vinogradova et al. (2016). Finally, we execute a one-way sensitivity analysis on the effect of ASCVD

events on future mortality. A two-way sensitivity analysis is performed to study the effect of the genetic testing range and cost of genetic testing on the ICER of GenePCE compared to the PCE strategy. All analyses are performed with R (v3.3.0 The R Foundation for Statistical Computing, Vienna AT) (R Core Team, 2016).

3.4.2 Results

3.4.2.1 Selection of Number of Replications for Analyses

The convergence of the mean ICER of the GenePCE strategy compared to the PCE strategy is shown in Figure 3.3. We observe that the mean ICER starting at 500 replications is close to the ICER of the GenePCE strategy compared to the PCE strategy at 2,000 replications ($63,722/QALY saved - represented with the red horizontal line in plot), with a moderately small variation. For the main analysis, we choose to run the simulation 1,000 times.

The sensitivity and subgroup analyses are performed using 500 replications. This number of replications is enough to achieve a standard error below $1,000/QALY saved with as little as 10 random samples (with replacement) from the initial simulation of 2,000 replications. The standard error of 10 samples of 500 and 1,000 replications are $851/QALY saved and $830/QALY saved, respectively. Note that 10 random samples with replacement from the initial simulation of 2,000 replications is equivalent to running 10 independent simulations with a smaller number of replications.

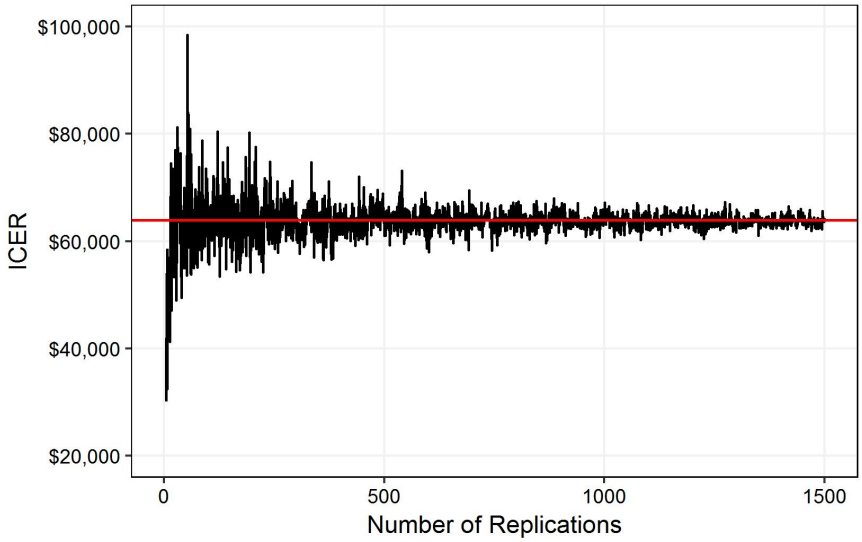


Figure 3.3: ICER of GenePCE (compared to the PCE strategy) under different number of replications at a genetic testing costs of $200 per test. Points represent the mean ICER using from 1 to 1,500 replications.

3.4.2.2 Study Population

We propose testing all patients for whom a moderately extreme GRS (defined as outside one standard deviation) would alter treatment. This is found to be a 10-year PCE risk between 7.68% and 13.63% during the first year of observation would receive genetic testing. A total of 16 million patients would be recommended a genetic test during the first year of our study (Figure 3.2). Table 3.2 describes the baseline characteristics of our simulated cohort of patients according to the PCE risk threshold (below or above the PCE treatment cut-off and within the genetic testing range). After correcting for asymmetry, the average GRS in the 1 SD genetic testing range is equal to 1. This implies that overall event rate in the population remains unchanged, as the average GenePCE risk is equal to the average PCE risk. In a baseline simulation in which the study population is untreated, we estimate 1.06 million CHD events and 640,039 stroke events. Of these, 481,056 are fatal CHD events and 74,850 are fatal stroke events. Our event rates are well calibrated with national data.

Table 3.2: Description of patients at first year of study.

PCE risk	**Within one SD below 10% 10-year risk[a]**	**Within one SD above 10% 10-year risk[a]**
Patients, in millions (%)	8.00 (49.48%)	8.17 (50.52%)
Women, in millions (%)	3.00 (37.46%)	2.63 (32.14%)
African-Americans, in millions (%)	1.31 (16.42%)	1.35 (16.49%)
Patients with LDL >160 mg/dL (%)	1.05 (13.12%)	0.88 (10.77%)
Tobacco users (%)	2.41 (30.13%)	2.61 (31.95%)
Patients with diabetes, in millions (%)	1.12 (13.94%)	1.09 (13.32%)
Average Age, years (SD)	59.36 (7.40)	62.07 (7.37)
Average SBP, mmHg (SD)	128.70 (16.96)	131.4 (16.94)
Expected 10-year survival rate if untreated, % (SD)	82.1% (0.93%)	77.73% (1.32%)
Expected years survival if untreated, years (SD)	9.12 (0.06)	8.9 (0.07)

[a] Patients are recommended genetic testing if their risk is within 1 SD of having a treatment change. For the group below 10%, this includes all patients with a PCE score from 7.68% to 10%. For the group above 10%, this included all patients with a PCE score from 10% to 13.63%.

3.4.2.3 Effect of Genetic Testing

Although genetic information does not change the overall rate of events in the population, it may change the individual risk scores of patients. Patients with a PCE risk score within

the 1 SD genetic testing range during the first year of the study receive genetic testing in our base case. The GRS of patients alters their risk for ASCVD events and may change their treatment plans. Genetic testing changes the first year of treatment of 4.84 million (29.91%) of patients (Figure 3.4). This effect could have been larger if both treatment strategies have had the same treatment cut-point, but this would have been an unfair comparison among the strategies. A summary of the difference between of the PCE and GenePCE risks using the PCE treatment threshold (10% PCE risk) is included in Table 3.3. We also study the impact of genetic testing on the distribution of the PCE risk over time. The distribution of the risk scores before and after genetic testing is illustrated in Figure B.1.

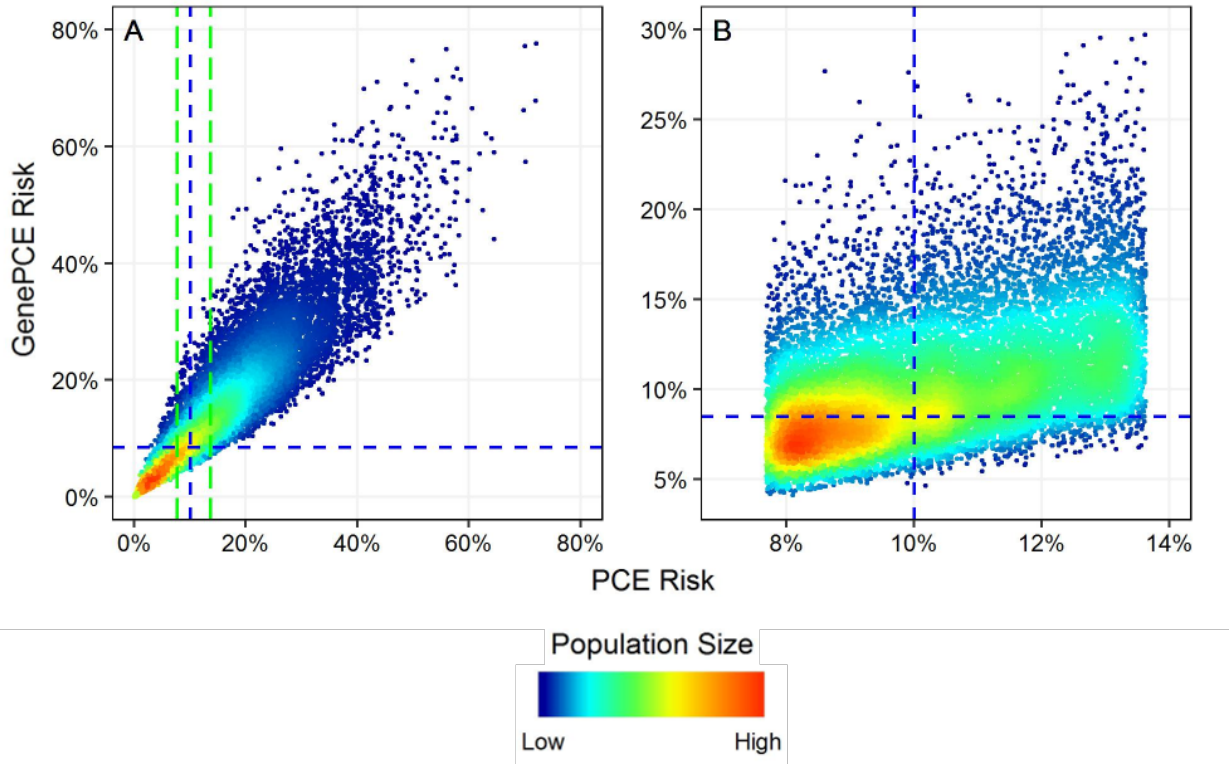


Figure 3.4: Comparison of PCE and GenePCE risks. The population size is represented with the color gradient (dark blue indicates the smallest population sizes and dark orange indicates the biggest population sizes). The green long dashed vertical lines indicate the population within the 1 SD genetic testing range. The blue short dashed lines represent the treatment thresholds. People with GenePCE risk scores above the horizontal blue short dashed line are treated according to the GenePCE strategy. People with PCE risk scores to the right of the vertical blue short dashed line are treated according to the PCE strategy. A shows the risk scores for the whole population of 93.6 million people (after LDL exclusions). B shows the risk scores of the people that received a genetic test during the first year of our study (16.17 million people).

Table 3.3: Treatment reclassification of patients at first year of study. Counts are presented in millions with percentages in terms of the population that received genetic testing.

PCE risk	GenePCE Risk	
	7.68%-10%	10%-13.63%
7.68%-10%	6.10 (37.73%)	1.90 (11.75%)
10%-13.63%	3.04 (18.8%)	5 .13 (31.72%)

3.4.2.4 Benefit of Genetic Testing

Once a genetic test is performed on a patient, the GenePCE risk score becomes available. This risk score may then change the treatment plans over the planning horizon. By performing genetic testing on 16.17 million people, the GenePCE strategy averts 4,928 ASCVD events and saves 42,052 more QALYs than the PCE strategy. This benefit is mainly due to the GenePCE strategy treating more patients during the initial years of our study. The amount of high intensity in both treatment strategies is a consequence of primary ASCVD events. Thus, the GenePCE strategy is more effective in preventing primary events, as it suggests approximately 30,000 less high intensity statins than the PCE strategy.

Segregating the outcomes of our simulated cohort of patients by treatment aggressiveness (treatment less/more intense for at least 3 years of the study prior an ASCVD event), we are able to identify the patients that received the most benefit from genetic testing (Table 3.4). If patients are treated less aggressively in the GenePCE strategy, the PCE strategy averts 3,865 more ASCVD events and saves 21,190 more QALYs than the GenePCE strategy. On the other hand, the GenePCE strategy prevents 3,946 more ASCVD events and saves 24,336 more QALYs than the PCE strategy by treating patients more intensively for at least three years of our study.

3.4.2.5 Cost-Effectiveness of Genetic Testing

While under our selection of parameters the PCE strategy results cost-saving, the GenePCE does not. By initiating treatment for more patients earlier in the study, the cost of treatment increases from $16.75 billion in the PCE strategy to $16.79 billion in the GenePCE strategy ($144/year for moderate intensity statin and $450/year for high intensity statin). On the other hand, initiating treatment earlier decreases the cost due to ASCVD events from $83.26 billion in the PCE strategy to $82.85 billion in the GenePCE strategy. Additionally, the GenePCE strategy has an overall cost of $3.14 billion due to genetic testing, which leads to the 2.77 billion 2018 USD increase in cost and an increase of $1,358 per QALY saved. The ICER of the GenePCE strategy compared to the PCE strategy is $66,298.

Table 3.4: Population health outcomes over 10-year time horizon. Includes only people who receive genetic testing during the first year of the study with 10-year PCE risk between 7.68% and 13.63% (16.17 million patients).

Policy	PCE	GenePCE
Recommended moderate statin, million patient-years	127.51	127.51
Recommended high intensity statin, million patient-years	3.58	3.55
CHD events averted by treatment	181,158	187,777
Stroke events averted by treatment	127,644	125,535
ASCVD events averted per 100 patients-years	1.91	1.94
Estimated number of patients treated more aggressively, millions[b]	1.33	1.11
CHD events averted by more aggressive treatment[b]	1,771	3,329
Stroke events averted by more aggressive treatment[b]	2,094	617
QALYs saved per 100 patient-years	12.39	12.65
QALYs saved, million	2	2.05
Total Cost, billion 2018 USD	-0.22	2.56
Cost per QALY saved, 2018 USD	-108	1,250
Cost per QALY saved, compared to PCE	-	66,298

[b] Treatment aggressiveness is defined as less/more intense treatment for at least 3 years of the study prior an ASCVD event.

3.4.2.6 Sensitivity Analyses

Table 3.5 illustrates the results of our one-way sensitivity analyses. By assuming that the GRS is incorrect, we find the extent of possible costs and harms of inappropriate adoption. Next, the AdjustedGenePCE strategy, in which people with high GRS have a greater relative risk reduction in addition to added predictive value, has substantially greater benefit than the base case. In the treatment threshold analyses we find that using the GRS without equalizing statin treatments dramatically changes the cost-effectiveness and the benefits (Table B.2). Changing the predictive quality of the GRS (by changing the odds ratio increase per standard deviation and, consequently, the treatment threshold) also has fairly large effects (Table B.2). If PCE risk increases by an odds ratio of 1.38 per standard deviation of the genetic score, the benefit GenePCE strategy is smaller than in our base case. The GenePCE strategy results in 21,985 less QALYs saved while costing 20 billion 2018 USD more than in the base case, when compared to the PCE strategy. An odds ratio increase of 2 per standard deviation of the genetic score has the opposite effect. In this case, the GenePCE strategy leads to more QALYs saved (55,176 vs 42,636 QALYs saved) and lower costs (2.63 vs 2.74 billion 2018 USD) than the base case, when compared to the PCE strategy. If treatment decisions in the GenePCE strategy are made according to the wrong GRS effectiveness, the GenePCE strategy results less cost-effective than in the

base case.

Table 3.5: Sensitivity analyses. Compares the results of the GenePCE strategy to the PCE approach.

Sensitivity analysis scenario	**ASCVD events averted**	**Cost/QALY saved**
Baseline	4,508	64,310
PCE event rate correct[c]	-2,676	-125,184
AdjustedGenePCE policy	19,780	22,792
GenePCE treatment threshold[d]		
Equal initial treatment	-11,196	-48,632
10% GenePCE threshold	-22,564	-20,382
Odds ratio increase per SD of genetic score		
1.38 (Abraham et al., 2016)	1,904	142,478
2	6,130	47,618
Wrong GRS effectiveness	1,458	167,400
Statin cost		
Both $50/year	4,508	63,701
Both decreased by half	4,508	63,796
Population ASCVD event rate		
Declines 30% due to unrelated factors	4,004	27,815
Treatment related disutility		
High intensity 0.004 QALY/statin-year	4,508	64,238
Double if age >70	4,508	63,615
Statin benefit misestimation		
Treatment discontinuation and restarting (Vinogradova et al., 2016)	3,798	85,048
Survival estimates		
Decreases by 25% due to ASCVD mortality	4,508	76,259
Decreases by 50% due to ASCVD mortality	4,508	93,581

[c] In this scenario we consider the potential impact on health and spending if the GenePCE event rate is incorrect and the PCE event rate is correct.
[d] The primary analysis set cut-points so that each treatment arm has the same amount of 10-year statin use. The "equal initial treatment" set cut-points to the same amount of initial statin use. The "10% GenePCE threshold" treats everyone once they reach a 10% event rate (Table B.2).

Reducing the cost of statins does not have a considerable effects on our base results. If the population ASCVD event rate falls by 30%, the GenePCE strategy becomes more cost-effective. We find that if statin treatment is discontinued and restarted fewer events are prevented in the GenePCE strategy at a higher cost than in the base case (2.93 vs 2.74 billion 2018 USD), when compared to the PCE strategy. Lastly, decreasing the survival estimates by 25% due to ASCVD mortality results in 13,227 less QALYs than in the base case, when compared to the PCE strategy. Decreasing the lifetime survival by 50% reduces

the benefit of the GenePCE strategy over the PCE strategy in 24,968 QALYs saved.

Genetic Testing Range and Cost: We perform a two-way sensitivity analysis on the cost and the recipients of genetic testing (Figure 3.5). We examine only testing people whose risk is under 10% (so genetic testing will only increase treatment) and different cut-points. In the "2 SD" ("0.5 SD") group, we test anyone whose PCE risk could cross the 10% threshold if their GRS is greater than 2 (0.5) standard deviations from the median, instead of the 1 SD group baseline. See Figure 3.5 for the results of the two-way sensitivity analysis.

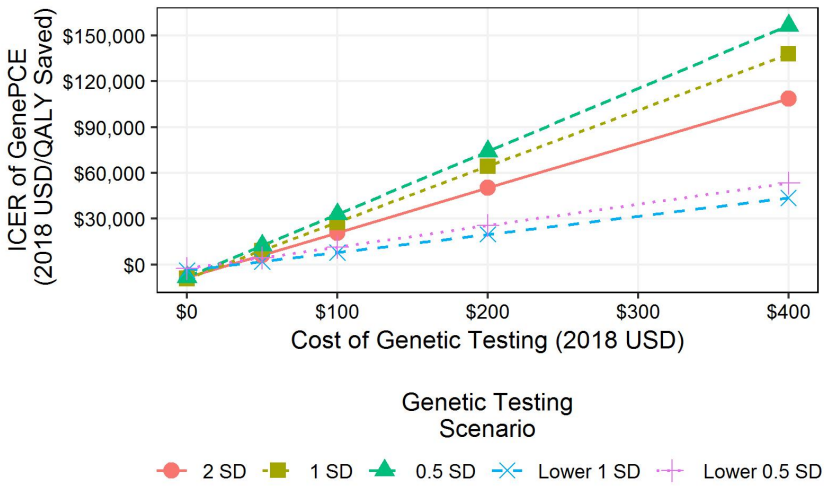


Figure 3.5: ICER of GenePCE (compared to the PCE strategy) using multiple genetic testing scenarios and costs of genetic testing. Points indicate ICER of GenePCE at genetic testing costs of $0, $50, $100, $200, $400, $800, and $2,000 per test.

The results are very sensitive to the cost of genetic testing, with all groups becoming cost-effective if genetic testing costs less than $100. The results are also much more cost-effective when only testing those with risk < 10%. This is because increasing statin use is itself cost-effective (Pandya et al., 2015). Performing genetic testing on more patients also improves cost-effectiveness, contrary to our hypothesis.

3.4.3 Discussion

We found that incorporating genetic risk scores into primary prevention of cardiovascular disease may already be cost-effective for over 60 million Americans if appropriately incorporated into treatment decisions. However, the overall public health impact would be

relatively small, likely preventing approximately 490 ASCVD events per year, improving the benefit of statins by roughly 4,200 QALYs.

Our work aligns well with the limited previous work on the subject. We found that the combination of the improving quality of GRS together with the decreasing price of genetic testing and the tremendous expense of ASCVD in American makes genetic testing relatively practical. Our study is quite different from the only other existing cost-effectiveness study that we are aware of, since our study was designed to estimate population-level effects of policies, instead of specific case examples. Our results are consonant with other cost-effectiveness analyses of novel risk markers, like coronary artery calcium screening (Roberts et al., 2015).

One unique analytic choice was to normalize treatment intensity between the arm that received genetic testing to one that didn't. We think this isolates the clinical question to being about genetics, without adding in the impact of increasing statin use.

As with any cost-effectiveness analysis, our findings are limited by the available data. Genetic risk is a rapidly advancing field; new findings that change our assumptions will happen regularly. The GRS could prove less effective in other populations. Healthcare costs also advance rapidly and vary widely. There is some question about the accuracy of the underlying PCE risk calculator. That said, these are the best available assumptions at the time and we performed multiple sensitivity analyses. Also, we decided a cut-point a priori of what we hypothesized would be a population that was relatively likely to benefit. Our analyses found that a larger population would potentially have been more cost-effective and had a larger public health benefit.

While we have modeled cholesterol treatment decisions based on the USPSTF statin guidelines, a crucial extension of this work is to model these decisions using MDP models. We address this extension in the subsequent section.

3.5 Cost-Effectiveness of Optimal Cholesterol Treatment Plans Informed with Genetic Information

In this section, we determine the potential public health impact of large-scale genetic testing to inform the use of optimal cholesterol treatment plans. We also evaluate how the cost-effectiveness of genetic testing changes with respect to the population tested. Cholesterol treatment plans are modeled using finite-horizon MDP models. The treatment plans are compared in a simulation framework. This allows us to asses different aspects of our model with great flexibility (Glover et al., 2018).

3.5.1 Simulation Framework

We use a large representative sample of the US population to understand the implications of genetic testing. To find a group for whom genetic testing is most likely to influence their care, we use the USPSTF statin guidelines (Bibbins-Domingo et al., 2016). Please refer to Subsection 3.4.1.1 for details on our study population.

We model the health status of each patient as a Markov chain and simulate their state trajectory over a 10-year planning horizon. The trajectory of a single patient in our modeling framework is summarized in Figure 3.6.

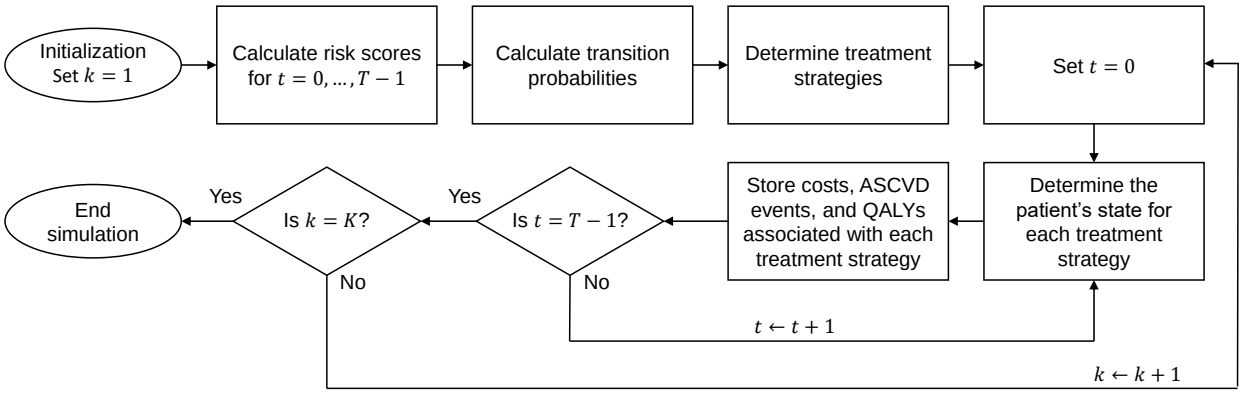


Figure 3.6: Summary of simulation framework for a single patient.

Before the start of the simulation, we calculate each patient's 1-year risk for ASCVD events at each year $t = 0, 1, \ldots, T-1$ using risk scores that are and are not informed by genetic information. We then estimate transition probabilities based on each risk score. The transition probabilities are used to determine treatment strategies using clinical information only and clinical and genetic information.

We compare the outcomes of every patient under each treatment strategy at every replication $k = 1, 2, \ldots, K$ and year t. Since genetic variants play a role in each patient's health (independent of whether or not they have received genetic testing), the health trajectory of each patient progresses from year t to year $t+1$ according to the risk score informed by clinical and genetic information. Finally, we estimate the lifetime effects of the treatment strategies on every patient, represented by the terminal time period T. The simulation ends once the planning horizon of each patient is simulated K times.

We describe our risk scores, and treatment strategies in more detail in the following subsections. A summary of the parameters used during our study can be found in Table 3.1 in Section 3.4.

3.5.1.1 Risk Scores

The risk for ASCVD events due to clinical factors of each patient per year is measured using an existing PCE risk score (Goff et al., 2014). This risk score is based on a proportional hazards model with age, gender, race, smoking status, diabetes status, SBP, HDL, and TC as explanatory variables. Based on clinical expertise, we assume independence among the ASCVD events. Moreover, we assume that 60% of the ASCVD risk is due to CHD events (Benjamin et al., 2018).

While no information of the risk for stroke events due to genetic factors is available, based on communications with clinical collaborators we assume that the risk for CHD events due to genetic factors can be measured using a GRS. We assume the GRS of each patient does not change over time. Previous studies have shown that the risk for CHD due to genetic factors is nearly independent from the risk for CHD due to clinical factors (Khera et al., 2016, 2018; Natarajan et al., 2017). Past research has also found that the distribution of genetic scores is near normal (Khera et al., 2018; Mega et al., 2015). This allows us to model the genetic score of each patient in our population as a standard Gaussian random variable. We then estimate the GRS as the odds ratio for CHD events per standard deviation of the genetic score, as estimated by Khera et al. (2018). The PCE risk increases by an odds ratio of 1.67 per standard deviation of the genetic score (Khera et al., 2018; Abraham et al., 2016).

We assume the event rate predicted by the PCE risk score as true at a population level (Goff et al., 2014). Hence, the average risk for ASCVD events in the population is expected to remain unchanged after the addition of genetic information. Since the distribution of odds ratios is asymmetric, the GRS would likely change the average risk for CHD events at a population level. To account for this, we develop a correction factor for the genetic risk scores. We define this correction factor as the ratio of the mean odds for CHD due to clinical factors to the mean odds for CHD due to clinical and genetic factors at the first year of our study. Combining the two components of the risk for CHD events, we can then estimate the risk for ASCVD events due to clinical and genetic factors, or the GenePCE risk score.

3.5.1.2 State Space

Our simulation model uses a state-space representation to fully describe the characteristics of every patient at each year t. A state $s_t \in \mathcal{S}$ consists of a patient's demographic information, laboratory measurements, health status, and GRS at year t. Each patient's demographic information and laboratory measurements are incorporated into our state in

the form of the PCE risk score. The demographic information encompasses the patient's age, race, and sex. The laboratory measurements include measurements of the patient's untreated SBP, HDL, TC, smoking status, and diabetes status. We assume each patient's health status can be classified at the beginning of each year into one of the following: (1) healthy (no history of CHD or stroke); (2) history of CHD but no adverse event in the current period; (3) history of stroke but no adverse event in the current period; (4) history of CHD and stroke but no adverse event in the current period; (5) survived a CHD event in the current period; (6) survived a stroke in the current period; (7) death from a non-cardiovascular disease related cause; (8) death from a CHD event in the current period; (9) death from a stroke in the current period; and (10) dead. We also assume that all patients have a healthy status at the beginning of our simulation.

3.5.1.3 Action Space

We limit our action space $\mathcal{A}$ to three treatment options: no treatment, moderate intensity statins, or high intensity statins. The treatment that patients receive each year depends on their current and future states. Once a treatment choice $a_t \in \mathcal{A}$ is made at the beginning of year t, we assume it has a near-immediate effect on each patient's health. We incorporate this effect by estimating the relative risk reduction due to treatment choice a_t. The estimates of the effect of cholesterol-lowering statins drugs on ASCVD events are derived from Collins et al. (2016).

3.5.1.4 Transition Probabilities

We use $p_t(\tilde{s}_t|s_t, a_t)$ to denote the transition probability from state s_t to state $\tilde{s}_t$, after taking action a_t at the beginning of year t. To calculate the transition probabilities, we first estimate the PCE risk score (risk for ASCVD events due to clinical factors) and GenePCE risk score (risk for ASCVD events due to clinical and genetic factors). As in previous studies, we assume that if patients have a history of CHD or stroke, they are more likely to have additional ASCVD events (Schell et al., 2016). To account for this, we multiply the patient's CHD odds by 3 if the patient has a history of CHD, multiply the stroke odds by 2 if the patient is at least 60 years old and has a history of stroke, and multiply the stroke odds by 3 if the patient has a history of stroke and is less than 60 years old (Brønnum-Hansen et al., 2001; Burn et al., 1994). The risks for CHD and stroke are adjusted if the patient receives treatment. For people who have an event in a given year, we calculate the probability for this event to be fatal by applying fatality likelihoods to the post-treatment risk for CHD and stroke (Lloyd-Jones et al., 2009; Sussman et al., 2013). We calculate the

fatality likelihoods as the ratio of known fatal event rates from the National Center for Health Statistics to the overall event rates predicted by the PCE risk, adjusted for age and gender (NCHS, 2017). In addition, we incorporate the probability of non-ASCVD mortality using life-tables (Arias et al., 2017).

3.5.1.5 Rewards and Costs

We define the reward $r_t(s_t, a_t)$ as the quality of life (QoL) associated with the patient's health status at state s_t minus the burden from treatment choice a_t at year t. During our study, we assume that the QoL weight depends only on the patient's health status. We obtain the QoL weights and burden from the treatment choices used during our analyses from previous studies (Fryback et al., 1993; Pignone et al., 2006; Pignone, 2007; Pandya et al., 2015). To evaluate the lifetime effect of the treatment choices, we assume the terminal condition of a patient can be computed as the product of the patient's expected lifetime, a mortality factor that accounts for the effect of ASCVD events on future mortality, and a terminal QoL weight (Fryback et al., 1993; Pignone et al., 2006; Pignone, 2007; Arias et al., 2017; Pandya et al., 2015).

The cost of fatal and non-fatal ASCVD events, history of ASCVD events, medications, and genetic testing are also obtained from existing literature (O'Sullivan et al., 2011; Medical Expenditure Panel Survey, 2015; GoodRx, 2017; Color Genomics, 2018). All the cost parameters in our analyses are adjusted for inflation from the original citations. All QoL weights and costs are discounted by 3% (Neumann et al., 2016).

3.5.1.6 Treatment Strategies

To evaluate the impact of genetic testing, we model the process of sequentially determining cholesterol treatment medications over the 10-year planning horizon as a finite-horizon discrete-time MDP (Puterman, 2014). We develop two MDP policies for each patient in our simulation, one based on clinical information only and another based on clinical and genetic information.

The objective of each MDP is to determine the treatment strategy that maximizes the expected discounted QALYs, based on the available information. We obtain optimal cholesterol treatment plans by solving the set of dynamic programming equations:

$$v_t(s_t) = \max_{a_t \in \mathcal{A}} \left\{ \sum_{\tilde{s}_t \in \mathcal{S}} p_t(\tilde{s}_t | s_t, a_t) \Big[r_t(\tilde{s}_t, a_t) + \lambda v_{t+1}(\tilde{s}_t) \Big] \right\},$$

for $s_t \in \mathcal{S}$, and $t = 0, 1, \ldots, T-1$, where $\lambda = 0.97$ is the discount factor of the model. Given

a terminal condition $v_T(s_T)$, we can find the optimal treatment strategy by recursively computing the value functions $v_t(s_t)$ for $t = 0, 1, \ldots, T-1$ and $s_t \in \mathcal{S}$.

Both MDP models share the same states, transition probabilities, and rewards as our simulation model. However, the MDP model developed with clinical information only does not include genetic information when treatment decisions are made.

3.5.1.7 Model Evaluation

Our simulation framework allows for an estimation of the benefit of treatment plans with and without genetic information in a flexible and realistic way. Given the patient's initial state s_0, our simulation framework enables us to approximate the total expected QALYs obtained from each treatment plan as:

$$v_0(s_0) \approx \frac{1}{K} \sum_{k=1}^{K} \sum_{t=0}^{T-1} \lambda^t r_{t,k}(s_{t,k}, a_{t,k}) + \lambda^T v_T(s_{T,k}),$$

where $r_{t,k}(s_{t,k}, a_{t,k})$ is the QoL associated with the patient's health status minus the treatment harm from $a_{t,k}$ at state $s_{t,k}$, year t, and replication k, $v_T(s_{T,k})$ is the terminal condition as defined in the MDP formulation at state $s_{T,k}$, $\lambda = 0.97$ is the discount factor of the model, and K is the total number of replications in our simulation model.

3.5.1.8 Calibration and Validation

To ensure the number of fatal and non-fatal CHD and stroke events in our simulation match those of the national statistics, we calibrate the amount of events predicted by the risk scores. Mortality from second cardiovascular events and known overdiagnosis of cardiovascular diseases is accounted for by decreasing the fatal event rates reported by the National Center for Health Statistics by 50% (Govindan et al., 2014). We estimate the overall event rates predicted by the risk scores by simulating the first year of the 10-year planning horizon of every patient following the USPSTF statin guidelines 50 times. We also run a baseline simulation in which the study population is untreated and the event rates are calibrated with national data (NCHS, 2017). The calibration of our model was verified by a clinical researcher at the University of Michigan Medical School.

Our simulation was built with high face validity by discussing the parameters and logic with experts in the field. Our co-author, a practicing clinician at the University of Michigan Hospital and researcher at the Veterans Affairs Ann Arbor Healthcare System, helped to validate our model. Additionally, our estimates of the effect of genetic information on

the risk for CHD events were discussed with geneticists at the Massachusetts General Hospital and the Department of Medicine at Harvard Medical School.

All analyses are performed with R (v3.5.0 The R Foundation for Statistical Computing, Vienna AT) (R Core Team, 2016). The computations are made using 50 Intel Xeon CPUs and 256GB of RAM. A single replicate of our simulation requires approximately 5 seconds of computing time.

3.5.2 Results

The number of replications needed to capture the heterogeneity in our population and to observe low-probability events is first evaluated. We study how the average QALYs saved and cost incurred per patient change by incorporating genetic information into treatment plans as we increase the number of replications. The effect of genetic information on each patient's PCE risk at the first year of our study is then investigated. Subsequently, we estimate the number of CHD events averted, QALYs saved, costs incurred, and cost per QALY saved following both treatment strategies, compared to no treatment. Lastly, we evaluate the impact of performing genetic testing in patients below and above 50 years old and in male and female patients while varying the cost of genetic testing from \$0 to \$400.

3.5.2.1 Selection of Number of Replications

Before investigating the impact of genetic testing in a representative sample of the US population, we derive the number of replications to run our simulation model empirically. We observe that the average QALYs saved and cost incurred after 750 replications are close to the average QALYs saved and cost incurred at 2,000 replications with small variation (Figure 3.7). Based on this, we choose to run the simulation $K = 750$ times for all analyses.

3.5.2.2 Comparison of Risk Scores

Out of a population of 93.55 million patients, 16.17 million (17.29%) patients receive a genetic test at the first year of our study. Even though genetic information does not change the overall rate of events in the population, it may change the individual risk scores of patients. The genetic information of patients alters their risk for ASCVD events and may change their treatment plans. Figure 3.8 shows an overall comparison of the PCE and GenePCE risk scores at the first year of our study. We observe that genetic information

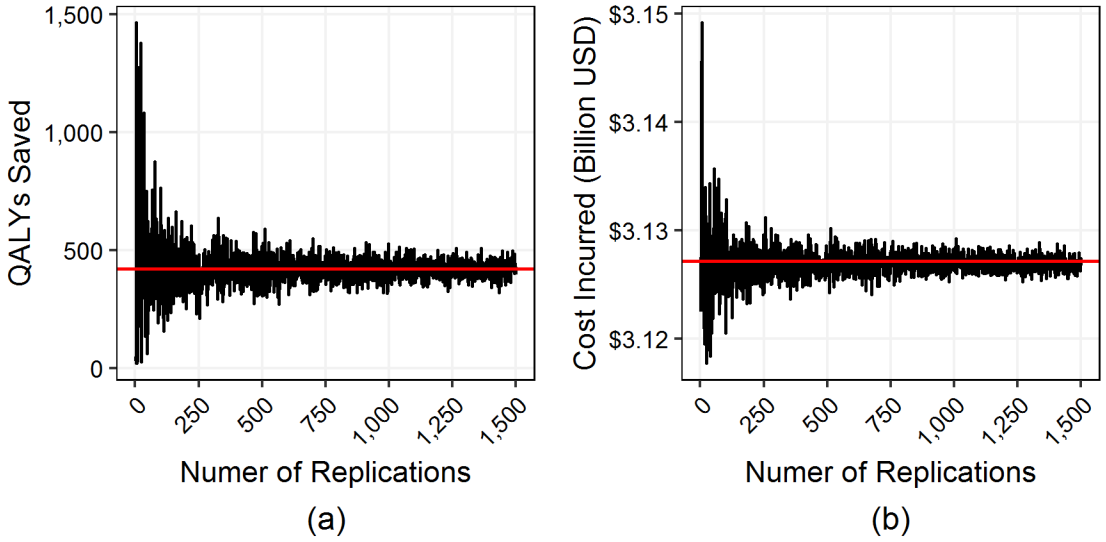


Figure 3.7: Convergence of QALYs saved and cost incurred due to genetic testing in simulation. Figure 3.7(a) illustrates the average QALYs saved due to genetic testing using from 1 to 1,500 replications. Points in Figure 3.7(b) represent the average cost incurred due to genetic testing using from 1 to 1,500 replications. The red horizontal lines in plots show the average QALYs saved and cost incurred at 2,000 replications.

helps identify individuals at increased risk for ASCVD events. However, it can be noticed that most of the population have GenePCE risk scores slightly lower than PCE risk scores.

3.5.2.3 Impact of Genetic Testing

Table 3.6 summarizes the health outcomes and costs of both treatment strategies compared to no treatment. We observe that the policies informed with clinical and genetic information averted 86 more CHD events and saved 390 more QALYs than the policies informed with clinical information only. Patients with GenePCE scores higher than PCE scores may require a more intense treatment than they would have received without genetic information. Conversely, patients for whom the GenePCE score is lower than the PCE score may require less treatment under the GenePCE score than given without genetic information.

After genetic testing, the cost of treatment decreases by $9.42 million. Since genetic testing provides information about which patients should receive higher treatment intensity, the cost due to ASCVD events decreases by $2.59 million using clinical and genetic information. Nonetheless, the policies derived using clinical and genetic information have an additional cost of $3.14 billion due to genetic testing ($200 per test in our base case). These cost estimates lead to an increase of nearly 3.13 billion USD in the total cost of the policies informed with clinical and genetic information. Understanding which

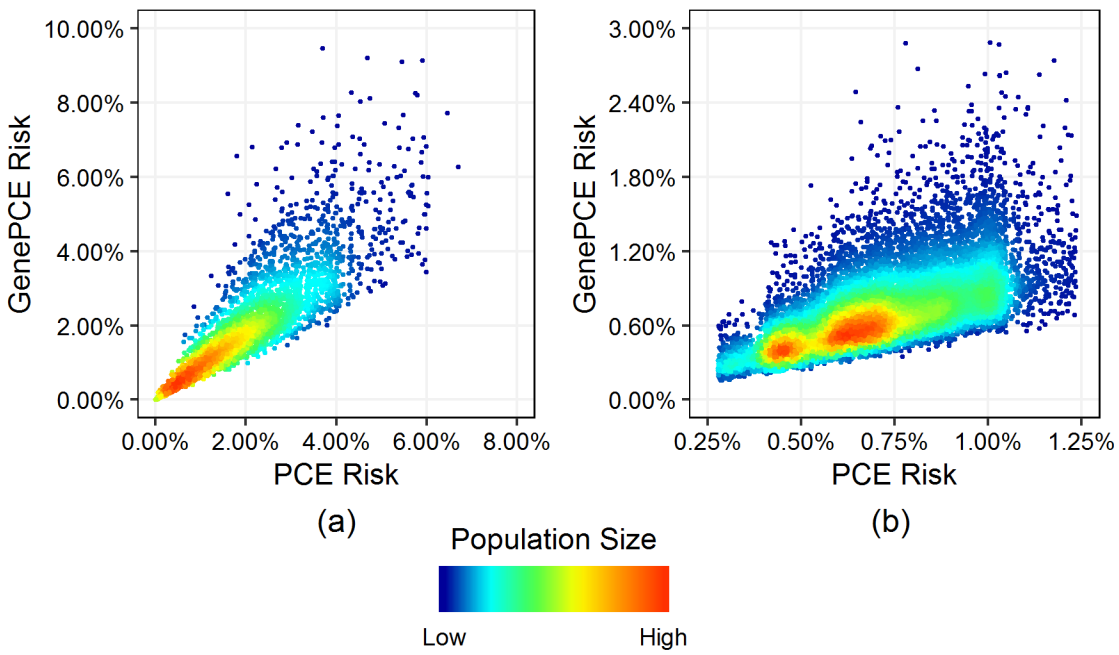


Figure 3.8: Comparison of PCE and GenePCE risk scores. The population size is represented with the color gradient (dark blue indicates the smallest population sizes and dark orange indicates the biggest population sizes). Figure 3.8(a) shows the risk scores of the whole population (93.55 million people). Figure 3.8(b) shows the risk scores of the patients that receive a genetic test during the first year of our study (16.17 million people).

population benefits the most from genetic testing may result in better health outcomes and lower costs.

3.5.2.4 Population and Genetic Testing Cost Analysis

By performing genetic testing on different populations, we are able to further identify which individuals would benefit the most from a genetic test. Both treatment strategies are well below the commonly used cost-effectiveness thresholds for all populations and testing costs, when compared to no treatment (Neumann et al., 2014). We find that the policies informed with clinical and genetic information are cost-saving compared to the policies informed with clinical information only if there is no cost associated with genetic testing. Genetic testing is most cost-effective if performed on people who are less than 50 years old. It is least cost-effective if performed on female individuals only. However, the ICER of the policies informed with clinical and genetic information compared to the policies derived with clinical information only is considerably higher than the regularly used cost-effectiveness thresholds in all scenarios (Neumann et al., 2014). The results of our population and genetic testing cost analysis are included in Figure 3.9.

Table 3.6: Population health outcomes at the end of the 10-year time horizon. The policies informed with clinical information only and with clinical and genetic information are compared to no treatment. Results are averages over 750 replications. Values within parenthesis indicate the standard deviation across replications.

Information Considered	**Clinical**	**Clinical and Genetic**
Moderate intensity treatment, patient-years	121,875 (25,979)	155,512 (30,085)
High intensity treatment, patient-years	149,188,644 (276,853)	149,153,543 (277,348)
CHD events averted by treatment	275,038 (30,272)	275,124 (30,298)
QALYs saved	4,286,972 (241,950)	4,287,362 (242,010)
Total cost, billion USD	\$27.81 (\$2.49)	\$30.94 (\$2.49)
Cost/QALY saved, USD	\$6,488 (\$840)	\$7,216 (\$874)

3.5.3 Discussion

In this section, we presented a simulation framework to evaluate the implications of genetic testing on optimal cholesterol treatment plans using a 10-year planning horizon. We applied this framework to a large sample representative of an adult population in the US. Although our simulation required longer running times than simpler models (such as Markov cohort models), it enabled us to asses different aspects of our model with great flexibility.

This work could be extended by incorporating other cholesterol-lowering drugs (such as fibrates). We chose statins because they are the most commonly used cholesterol-lowering drugs and there is reliable data about their benefit. Another extension of this work could be to incorporate the effect of genetic testing on hypertension treatment or aspirin use. The same genetic test could be used to guide decisions of multiple conditions at once. However, risk prediction plays a smaller role in hypertension treatment decisions and the clinical benefit of aspirin is now unclear (McNeil et al., 2018).

Our simulation framework may also be used to evaluate other treatment strategies such as the USPSTF guidelines or the ACC/AHA statin guidelines in practice (Bibbins-Domingo et al., 2016; Goff et al., 2014). We sought to find treatment strategies that maximized the expected discounted QALYs due to their popularity in quantifying the health-related benefits gained from clinical interventions (Fonarow et al., 2017; Glover et al., 2018; Jarmul et al., 2018; Mason et al., 2014). Other objective functions (such as cost alone, or more general utility functions) may be used to find treatment plans.

Two additional key extensions of this work are to identify the population that would

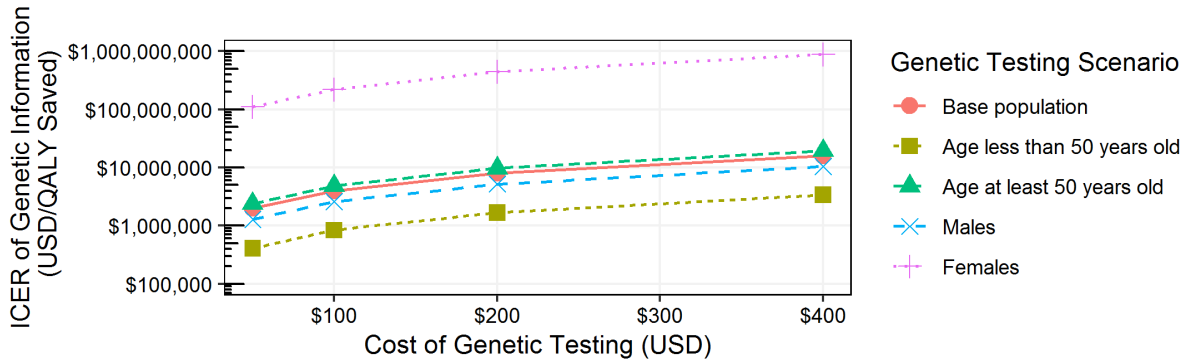


Figure 3.9: ICER of genetic information in different populations and costs of genetic testing. Points indicate the ICER of genetic information at genetic testing costs of $50, $100, $200, and $400 per test. The base population represents performing genetic testing according to clinical expertise and the USPSTF statin guidelines. A base-10 logarithmic scale is used in the vertical axis for illustration purposes. Minor tick marks in the vertical axis represent 10% increases between the major tick marks.

benefit the most from genetic information and when to perform such a test. We address these questions in the next section.

3.6 Optimal Cholesterol Treatment and Genetic Testing Strategies

In this section, we model cholesterol treatment plans as a finite-horizon MDP (Puterman, 2014). We determine which patients should receive genetic testing and the time of testing using VoI analysis (Raiffa and Schlaifer, 1961). That is, we perform a genetic test if there is evidence that it might benefit a patient. The treatment and testing decisions are incorporated in a simulation-based framework. We chose a simulation-based framework to evaluate treatment plans and decide when to perform genetic tests because it allows us to assess different aspects of our model with great flexibility (Glover et al., 2018). The objective of our work is to determine when to perform genetic testing and to understand how optimal treatment plans change with the addition of genetic information. Our research bridges a gap between the adoption of genetic information in patient care and cholesterol treatment plans by modeling the health trajectory of patients stochastically and making treatment and testing decisions simultaneously.

3.6.1 Modeling Framework

We consider a large representative sample of the US population composed of adult Caucasian or African-American patients from 40 to 60 years old. Given the information of each patient, the following decisions must be made at each year $t = 0, 1, \ldots, T-1$ of the planning horizon: (1) if and how the patient's cholesterol should be treated; and (2) whether or not the patient should receive a genetic test. We incorporate both decisions by simulating the health trajectory of every patient over the planning horizon.

The trajectory of a single patient in our modeling framework is summarized in Fig. 3.10. Before the start of the simulation, we calculate the risk for ASCVD events of the patient at each year using a risk score that is only informed by clinical information. Subsequently, we estimate transition probabilities based on clinical information only. Because the effects of genetic information are not known prior to a genetic test, we simulate potential realizations of genetic information from a known probability distribution. At each realization $g = 1, \ldots, G$, we adjust the risk for ASCVD events with the potential genetic information and estimate transition probabilities based on clinical and genetic information. The transition probabilities are then used to determine the treatment policies derived with clinical information only and with clinical and genetic information.

Since genetic variants play a role on patients' health (independent of whether or not they have received genetic testing), the health trajectory of patients evolves from year t to year $t+1$ according to clinical and genetic information. The health trajectory of patients is adjusted based on each treatment plan. We compare the outcomes of patients for each potential realization of genetic information g at every health trajectory replication $k = 1, 2, \ldots, K$ and year t. Lastly, we estimate the lifetime effects of the treatment decisions on each patient, represented by the terminal time period T.

The simulation ends once the planning horizon of every patient is replicated K times under G realizations of their potential genetic information. A genetic test is performed if there is evidence that it would provide some benefit based on the VoI. Specifically, a genetic test is performed if the VoI is above a predetermined threshold θ at year t. We describe our main clinical assumptions as well as the formulation of the risk scores, treatment decisions, and testing decisions in more detail in the following subsections.

3.6.1.1 Clinical Assumptions

To answer the clinically complex question of how to perform cholesterol treatment and genetic testing decisions simultaneously, we make several assumptions associated with the risk for ASCVD events or with the benefits and costs of treatment. All our assumptions

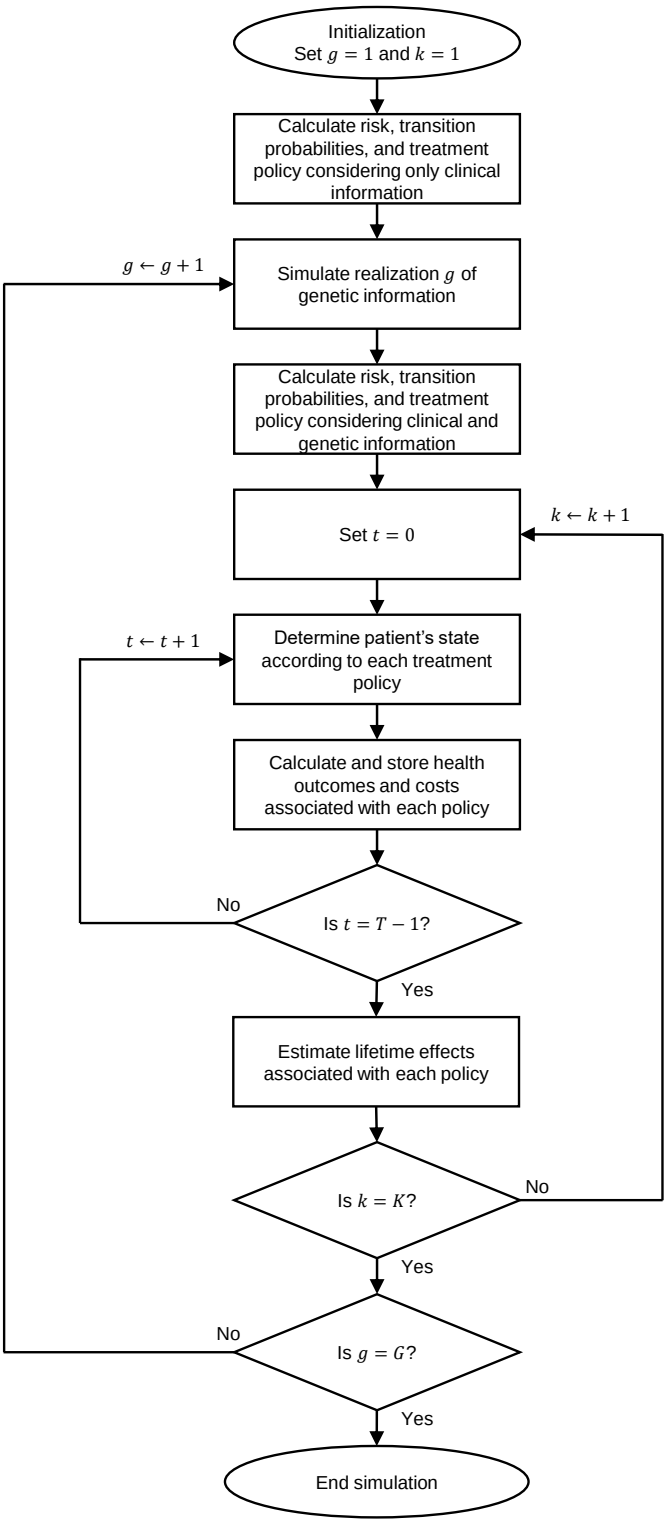


Figure 3.10: Summary of modeling framework for a single patient. The index g represents a realization of genetic information, t represents the year in the planning horizon, and k represents a realization of the health trajectory of the patient.

have clear citations and align with other high-profile cost-effectiveness studies. We begin by describing our risk-related assumptions.

We measure the risk for ASCVD events due to clinical factors using the PCE score from the American College of Cardiology (ACC) and the American Heart Association (AHA), the primary predictive model in multiple clinical guidelines (Bibbins-Domingo et al., 2016; Goff et al., 2014; Stone et al., 2014). The PCE score predicts ASCVD, which consists of CHD and stroke. Since both diseases have the same basic risk factors (and because there are currently no separate PCE scores for CHD and stroke) we make the following assumption:

Assumption 3.1 *The risk for ASCVD events can be decomposed into risk for CHD events and risk for stroke events.*

While not specifically linked to the PCE score, previous studies have found that if patients have a history of CHD or stroke, they are more likely to have additional ASCVD events (Brønnnum-Hansen et al., 2001; Burn et al., 1994). As a result, we assume the following:

Assumption 3.2 *Secondary ASCVD events are more common than would be predicted by the PCE score alone.*

At the time of this work, there is no information of the risk for stroke events due to genetic factors, but the risk for CHD events due to genetic factors has been quantified using a GRS (Mega et al., 2015; Khera et al., 2016; Natarajan et al., 2017). Although this quantity is unknown before a genetic test, previous studies have reported that genetic scores appear to be normally distributed (Khera et al., 2018). Furthermore, past research has shown that the risk for CHD due to genetic factors is nearly independent from the risk for CHD due to clinical factors (Khera et al., 2016, 2018; Natarajan et al., 2017). These findings, along with the understanding that genetic risk scores are not commonly performed in practice yet (Lewis and Vassos, 2020), lead us to our next assumption:

Assumption 3.3 *Genetic scores can be sampled from a normal distribution. The risk for ASCVD events due to genetic factors is independent from the risk for ASCVD events due to clinical factors. Moreover, no patient has received genetic testing prior to the beginning of the planning horizon.*

Since the PCE risk score is widely used in practice, and it has been validated across the populations considered in our study, we make the following assumption:

Assumption 3.4 *The event rate predicted by the PCE risk score is true at a population level.*

A direct consequence of this assumption is that the average risk for ASCVD events in the population is expected to remain unchanged after the addition of genetic information.

Besides risk-related assumptions, we make assumptions on the effect of treatment on patients' health. Based on communications with clinical collaborators, every cholesterol-lowering drug is coded as a moderate intensity statin except for atorvastatin and rosuvastatin. We focus on statin drugs because they are the most commonly used cholesterol-lowering drugs, and clinical guidelines already recommend that the risk for ASCVD events should be used to guide their usage (Bibbins-Domingo et al., 2016; Stone et al., 2014). As statin drugs have relatively fast effects on patients' cholesterol, we make the following assumption:

Assumption 3.5 *Cholesterol treatment has a near-immediate effect on patients' health.*

This assumption allows us to incorporate the impact of treatment in patients' rewards.

After patients receive cholesterol treatment, they may remain healthy, survive an ASCVD event, die from an ASCVD event, or die from a non-cardiovascular disease related cause. Each of these health conditions has a corresponding QoL weight, which has been elicited through patient surveys (Gold et al., 2002). Adding over the amount of time patients are expected to spend having each health condition enables us to calculate QALYs. In a similar way, there is a cost associated with each health condition and each statin drug. This leads to our last main assumption:

Assumption 3.6 *Patients' rewards and costs only depend on their health condition and current treatment. Furthermore, patients' terminal rewards and costs only depend on their health condition at the end of the planning horizon and their life expectancy afterwards.*

The terminal rewards and costs in Assumption 3.6 represent patients' expected QALYs and costs after the end of the planning horizon.

3.6.1.2 Risk Scores

The PCE risk score was developed using a proportional hazards model with age, sex, race (Caucasian or African-American), smoking status, diabetes status, SBP, HDL, and TC as explanatory variables (Kleinbaum and Klein, 2005). For a single patient, the risk for ASCVD events due to clinical factors is given by the PCE risk score:

$$R(\zeta_t) = 1 - S_0(t)^{e^{\vec{\beta}(\vec{x}_t - \vec{u})}}, \tag{3.1}$$

where $S_0(t)$ is the baseline survival function at time t, $\vec{x}_t$ is a vector of the clinical characteristics of the patient (i.e. age, sex, race, smoking status, diabetes status, SBP, HDL,

and TC) at year t, $\vec{\mu}$ is a vector of the race- and sex-specific population means of the clinical characteristics, and $\vec{\beta}$ is a vector containing coefficients for each clinical characteristic. For simplicity of notation, we let $\zeta_t = \{\vec{x}_t, \vec{\mu}, \vec{\beta}\}$ represent all clinical information of the patient at time t. In line with Assumption 3.1, α of the ASCVD risk is due to CHD events and $1 - \alpha$ is due to stroke events. For a single patient, the risk for CHD and stroke events due to clinical factors is obtained from the PCE risk score as $R^{CHD}(\zeta_t) = \alpha R(\zeta_t)$ and $R^{Stroke}(\zeta_t) = (1 - \alpha)R(\zeta_t)$, respectively.

Combining the clinical component of the risk for CHD events with a realization of the GRS γ_g, we can estimate the risk for ASCVD events due to clinical and genetic factors. This risk score will be referred to as the GenePCE risk score. We obtain the GenePCE risk score of a single patient as:

$$R(\zeta_t, \gamma_g) = R^{CHD}(\zeta_t, \gamma_g) + R^{Stroke}(\zeta_t). \tag{3.2}$$

3.6.1.3 Treatment Decisions

The process of sequentially determining cholesterol treatment medications over a planning horizon is modeled as a finite-horizon discrete-time MDP (Puterman, 2014). Each year, we model two treatment plans for each patient in our population: one using the PCE risk score (clinical information only) and another using the GenePCE risk score (clinical and genetic information). The objective of each MDP is to determine the treatment strategy that maximizes the expected discounted QALYs. We use the following notation for treatment decisions:

t	index of years in our planning horizon; $t = 0, 1, \ldots, T-1$. Decisions are made at the beginning of each year t.
s_t	patient's state consisting of clinical and genetic information at time t; $s_t \in \mathcal{S}$, where $\mathcal{S} = \{1, \ldots, S\}$ is a finite set of states.
h_t	health condition of a patient at time t; $h_t = 1, \ldots, 10$.
a_t	cholesterol treatment choice; $a_t \in \mathcal{A}$, where $\mathcal{A} = \{1, \ldots, A\}$ is a finite set of treatment choices.
$RRR(a_t)$	relative risk reduction in the risk for ASCVD events due to treatment choice a_t.
$p_t(\bar{s}_t \mid s_t, a_t)$	transition probability from state s_t to state $\bar{s}_t$, after taking action a_t at year t.
$\phi(h_t)$	scaling factor to account for history of ASCVD events when a patient has health condition h_t.

$\rho(h_t)$	probability that an ASCVD event is fatal, or fatality likelihood, when a patient has health condition h_t.
$\psi(s_t)$	probability of a non-ASCVD death when a patient has state s_t.
$r_t(s_t, a_t)$	patient's reward defined as the QoL weight associated with state s_t and treatment choice a_t.
$d(a_t)$	disutility, or treatment harm, associated with treatment choice a_t.
λ	discount factor of the model; $\lambda \in (0, 1]$.
$v_t(s_t)$	value function for state s_t.
$L(s_T)$	patient's expected lifetime after year $T-1$ at state s_T.
$M(s_T)$	mortality factor that accounts for the effect of ASCVD events on future mortality after year $T-1$ at state s_T.
$q(h_T)$	terminal QoL weight.

We now describe $\mathcal{S}$, $\mathcal{A}$, $p_t(\bar{s}_t|s_t, a_t)$, $r_t(s_t, a_t)$, and the optimality equations in more detail.

State space: A state $s_t \in \mathcal{S}$ consists of a patient's demographic information, laboratory measurements, health condition, and GRS. The demographic information encompasses the patient's age, race, and sex. The laboratory measurements include measurements of the patient's untreated SBP, HDL, TC, smoking status, and diabetes status. Each patient also has one of ten mutually-exclusive health conditions $h_t = 1, 2, \ldots, 10$:

(1) Healthy (no history of CHD or stroke)

(2) History of CHD but no adverse event in the current year

(3) History of stroke but no adverse event in the current year

(4) History of CHD and stroke but no adverse event in the current year

(5) Survived a CHD event

(6) Survived a stroke

(7) Death from a non-cardiovascular disease related cause

(8) Death from CHD event

(9) Death from stroke

(10) Dead

Action space: We limit our action space $\mathcal{A}$ to three treatment options: no treatment, moderate intensity statin drugs, or high intensity statin drugs. As stated in Assumption 3.5, once a treatment decision is made at the beginning of each year, it has a near-immediate effect on a patient's health. We incorporate this effect by estimating the relative risk reduction (RRR) due to treatment choice a_t, denoted by $RRR(a_t)$.

Transition probabilities: To calculate $p_t(\bar{s}_t|s_t, a_t)$, we first estimate the dynamics of patients' laboratory measurements over the planning horizon. These estimates are used to calculate the risk for ASCVD events due to clinical factors or due to clinical and genetic factors using equation (3.1) or equation (3.2), respectively. In accordance with Assumption 3.2, if patients have history of CHD or stroke, they are more likely to have additional ASCVD events. This is incorporated into our transition probabilities using a scaling factor $\phi(h_t)$ that depends only on the health condition of the patient h_t. The risk for CHD and stroke are altered if the patient receives treatment. For patients who have an event in a given year, we calculate the probability that the event is fatal by applying fatality likelihoods $\rho(h_t)$ to the post-treatment risk of CHD and stroke events. These fatality likelihoods depend only on the health condition of the patient h_t. Finally, we consider the probability of non-ASCVD mortality $\psi(s_t)$ in our transition probabilities to account for other risks competing with the risk for ASCVD events in our model.

Rewards: As per Assumption 3.6, the rewards $r_t(s_t, a_t)$ of a patient only depend on their health condition h_t and current treatment a_t. The rewards are calculated as the QoL weight $q(h_t)$ associated with each health condition h_t minus the disutility $d(a_t)$ from a cholesterol treatment decision a_t. In line with Assumption 3.6, the terminal condition $v_T(s_T)$ is computed as the product of the patient's expected lifetime $L(s_T)$, a mortality factor that accounts for the effect of ASCVD events on future mortality $M(s_T)$, and their terminal QoL weight $q(h_T)$.

Optimality equations: The optimal cholesterol treatment plans are obtained by solving the set of dynamic programming equations:

$$v_t^*(s_t) = \max_{a_t \in \mathcal{A}} \sum_{\bar{s}_t \in \mathcal{S}} p_t(\bar{s}_t|s_t, a_t)\Big[r_t(\bar{s}_t, a_t) + \lambda v_{t+1}(\bar{s}_t)\Big], \tag{3.3}$$

for $s_t \in \mathcal{S}$ and $t = 0, 1, \ldots, T-1$. Given the terminal condition $v_T(s_T) = L(s_T)M(s_T)q(h_T)$, we can find the optimal treatment strategy by recursively computing the value functions $v_t(s_t)$.

3.6.1.4 Genetic Testing Decisions

For the genetic testing decisions, we consider two health trajectories per patient over the planning horizon $t = 0, 1 \ldots, T-1$: one under a treatment plan using clinical information only, and another under a treatment plan derived with clinical and genetic information. We compare these two health trajectories and express the opportunity losses using VoI analysis (Raiffa and Schlaifer, 1961). While there is reliable data on the side effects of cholesterol-lowering drugs, limited research has been done on the burden of undergoing a genetic test, such as anxiety of receiving information about the probability of a future disease (Hoel et al., 2006). Since there is no QoL weight or disutility associated with genetic testing, we quantified the VoI in terms of costs. We use the following notation for testing decisions:

π^*	treatment policy derived with an MDP using clinical information only; $\pi^* = (\pi_1^*, \ldots, \pi_{T-1}^*)$, where $\pi_t^* = (\pi_t^*(1), \ldots, \pi_t^*(S))$.
π^{**}	treatment policy derived with an MDP using clinical and genetic information.
s_t'	state of a patient at year t under treatment decisions using clinical information only; $s_t' \in \mathcal{S}$.
s_t''	state of a patient at year t under treatment decisions using clinical and genetic information; $s_t'' \in \mathcal{S}$.
a_t^*	(sub)optimal treatment decision at time t given by policy π^*; $\pi_t^*(s_t') = a_t^*$ for $s_t' \in \mathcal{S}$ and $a_t^* \in \mathcal{A}$.
a_t^{**}	optimal treatment decision at year t following policy π^{**}.
$c_t(s_t, a_t)$	cost of treatment $c_t^{treatment}(a_t)$, plus cost of ASCVD events $c_t^{events}(s_t)$, plus cost of genetic testing c_t^{test}, as appropriate, associated with state $s_t \in \mathcal{S}$ and treatment choice $a_t \in \mathcal{A}$ at year t.

Functions of states or actions with no superscript are applied to both policies, clinical information only and clinical and genetic information. Notation not included above is defined in Section 3.6.1.3.

Before a genetic test is performed, the true benefit of genetic information is unknown. Furthermore, the true health progression of patients over time is also unknown. However, given the states of a patient s_t'' and s_t' at year t, it is possible to compute a hypothetical VoI by taking expectations over the effect of genetic information and the health trajectory of a

patient as:

$$\text{VoI}_t(s'_t, s''_t) = \mathbb{E}_{\mathcal{G}}\Big[\mathbb{E}_{\mathcal{H}}[c_t(s'_t, a^*_t) - c_t(s''_t, a^{**}_t)|\mathcal{G}]\Big], \tag{3.4}$$

where $\mathbb{E}_{\mathcal{G}}[\cdot]$ denotes the expectation with respect to the GRS and $\mathbb{E}_{\mathcal{H}}[\cdot|\mathcal{G}]$ denotes the expectation with respect to the health trajectory of the patient given a realization of the GRS $\mathcal{G}$.

In equation (3.4), both $c_t(s'_t, a^*_t)$ and $c_t(s''_t, a^{**}_t)$ include the cost of treatment and the costs associated with ASCVD events. However, $c_t(s''_t, a^{**}_t)$ also includes the cost of genetic testing, if it has not been performed in an earlier year. This concept can be extended to multiple decision periods by letting equation (3.4) account for the patient's trajectory until year $\tau = 0, 1, \ldots, T-1$ as follows:

$$\text{VoI}_\tau(s'_{0:\tau}, s''_{0:\tau}) = \mathbb{E}_{\mathcal{G}}\Big[\mathbb{E}_{\mathcal{H}}\Big[\sum_{t=0}^{\tau} \lambda^t\Big(c_t(s'_t, a^*_t) - c_t(s''_t, a^{**}_t)\Big)\Big|\mathcal{G}\Big]\Big], \tag{3.5}$$

where $\lambda \in (0, 1]$ is the same discount factor as in the MDP models and $s_{0:\tau}$ represents the trajectory of a patient from year 0 to year τ (i.e. $s_0, s_1, \ldots, s_\tau$).

Accounting for the whole trajectory of the patient, equation (3.5) becomes:

$$\text{VoI}_T(s'_{0:T}, s''_{0:T}) = \mathbb{E}_{\mathcal{G}}\Big[\mathbb{E}_{\mathcal{H}}\Big[\sum_{t=0}^{T-1} \lambda^t\Big(c_t(s'_t, a^*_t) - c_t(s''_t, a^{**}_t)\Big) + \lambda^T\Big(c_T(s'_T) - c_T(s''_T)\Big)\Big|\mathcal{G}\Big]\Big]. \tag{3.6}$$

where $c_T(s_T)$ is a terminal condition that accounts for the lifetime costs of ASCVD events when a patient has state s_T at the end of the planning horizon. In accordance with Assumption 3.6, the terminal condition $c_T(s_T)$ can be computed as the discounted cost of history of ASCVD events over the expected lifetime of the patient $L(s_T)$, adjusted by a mortality scaling factor $M(s_T)$, i.e.

$$c_T(s_T) = c_T^{events}(s_T) \sum_{t=0}^{L(s_T)M(s_T)} \lambda^t.$$

Given the initial state of a patient s_0, our proposed genetic testing strategy is to identify the first year at which $\text{VoI}_\tau(s'_{0:\tau}, s''_{0:\tau}) > \theta$ for $\tau = 0, 1, \ldots, T-1$ and a genetic testing threshold $\theta \in \mathbb{R}_+ = \{y \in \mathbb{R} | y \geq 0\}$. This genetic testing threshold represents the magnitude of the loss a clinician is willing to accept by not performing a genetic test on a patient. Examples of this threshold are zero, for the case that the clinician is not willing to accept any opportunity loss, or the expected cost of an ASCVD event, for the case that the clinician

would only perform a genetic test if it would prevent an ASCVD event, in expectation. Our genetic testing strategy aims to find the first period at which genetic testing would provide an expected cost-saving greater than θ.

3.6.2 Simulation Framework

We develop a simulation model to evaluate cholesterol treatment plans derived with clinical information only and with clinical and genetic information. Using this framework, we determine when genetic testing should be performed. Our simulation model is based on the Markov chain embedded in the MDP derived with clinical and genetic information. Please refer to subsection 3.5.1 for a detailed description of our simulation framework. The base case of our analysis is performed using a genetic testing threshold of $\theta = 0$. As specified in Assumption 3.3, no patient has received genetic testing prior to the beginning of the simulation. We estimate the lifetime effects of following the treatment plans over the planning horizon in a sample of patients representative of a US population. A summary of the parameters used during our study can be found in Table 3.1 in Section 3.4.

3.6.2.1 Selection of Number of Replications

Prior to evaluating the impact of genetic testing in optimal treatment plans, we derive the number of replications to run our simulation empirically. Since the treatment policies aim to maximize QALYs and the testing strategies seek to minimize cost, we use the total QALYs saved and cost saved across the population to evaluate the convergence of our simulation. We first obtain the number of health trajectory replications K necessary to observe low probability events under a single GRS per patient. Once we obtain K, we fix it and examine how many GRS realizations G per patient are needed to represent the potential effects of genetic information. Lastly, we assess if K and G are sufficient to capture the variation in the population's health outcomes under uncertainty in the effect of genetic information.

3.6.2.2 Data

As in the previous sections, we use NHANES from 2009 to 2014 to parameterize our model (Centers for Disease Control and Prevention, 2020). However, in this section we restrict our population to adult Caucasian or African American patients from 40 to 60 years old with no history of heart attack, stroke, or congestive heart failure and LDL less than 190 mg/dL (in contrast to 40 to 75 years old). We chose this population based on conversations

with our clinical collaborators. Patients in this age range are more likely to experience differences in treatment after the incorporation of genetic information than older patients. Please refer to Subsection 3.4.1.1 for details on the pre-processing performed in our input data.

3.6.2.3 Model Evaluation

Similar to Section 3.5, we use our simulation framework to estimate the benefit of treatment plans with and without genetic information. Our simulation enables us to compute and store the costs, number of ASCVD events, and QALYs associated with each treatment strategy at every year t. Furthermore, we are able to identify when clinical and genetic information provides lower costs and/or better health outcomes than clinical information only.

Given the initial state of a patient s_0, we approximate equation (3.3) for each treatment plan as:

$$v_t^*(s_t) \approx \frac{1}{GK} \sum_{g=1}^{G} \sum_{k=1}^{K} \sum_{\bar{t}=t}^{T-1} \lambda^{\bar{t}} r_{\bar{t},k,g}(s_{\bar{t},k,g}, a_{\bar{t},k,g}) + \lambda^T v_T(s_{T,k,g}),$$

where $s_{t,k,g}$ is the state of the patient at year t, replication k and GRS realization g, K is the total number of health trajectory replications in our simulation model, G is the total number of GRS realizations, and $v_T(s_{T,k,g})$ is the terminal condition as defined in the MDP formulation. Furthermore, we approximate equation (3.5) as:

$$\text{VoI}_\tau(s'_{0:\tau}, s''_{0:\tau}) \approx \frac{1}{GK} \sum_{g=1}^{G} \sum_{k=1}^{K} \sum_{t=0}^{\tau} \lambda^t \Big(c_{t,k,g}(s'_{t,k,g}, a^*_{t,k,g}) - c_{t,k,g}(s''_{t,k,g}, a^{**}_{t,k,g}) \Big).$$

Equation (3.7) can be extended to approximate equation (3.6) by letting $\tau = T - 1$ and adding the terminal condition at each replication and GRS realization. This approximation allows us to estimate the first period such that $\text{VoI}_\tau(s'_{0:\tau}, s''_{0:\tau}) > \theta$ for $\theta \in \mathbb{R}_+$.

3.6.2.4 Sensitivity analyses

We perform sensitivity analysis on the genetic testing strategies by varying the model parameters and assumptions. The sensitivity analyses on model parameters are described in Table 3.9. These parameters and their sensitivity analysis values are selected based on existing literature and clinical expertise (Pletcher et al., 2014; Hayward et al., 2010; Pandya et al., 2015; O'Sullivan et al., 2011).

Table 3.9: Sensitivity analysis parameter values.

Parameter (units)	**Base case (sensitivity analysis values)**
Genetic testing threshold (2019 USD)	\$0[e]$\Big(\mathbb{E}[c_t^{CHD}(s_t') - c_t^{CHD}(s_t'')]$[f]$\Big)$
Cost of genetic testing (2019 USD)	\$200 (\$0-\$400)
Cost of moderate intensity treatment (2019 USD/year)	\$144 (\$12- \$288)
Cost of high intensity treatment (2019 USD/year)	\$450 (\$119- \$2,454)
Moderate intensity treatment disutility (QoL weight)	0.001 (0.0005-0.01)
High intensity treatment disutility (QoL weight)	0.002 (0.004-0.02)

[e] A genetic threshold of \$0 represents any cost savings.
[f] A genetic threshold of $\mathbb{E}[c_t^{CHD}(s_t') - c_t^{CHD}(s_t'')]$ represents saving at least the expected cost of a CHD event. We approximate this threshold as $\mathbb{E}[c_t^{CHD}(s_t') - c_t^{CHD}(s_t'')] \approx \frac{1}{GK}\sum_{g=1}^{G}\sum_{k=1}^{K}\lambda^t\Big(c_t^{CHD}(s_t') - c_t^{CHD}(s_t'')\Big)$.

We also perform a sensitivity analysis on how the treatment policies and testing strategies are achieved. To do this, we use a utility function that incorporates costs and QALYs to derive the treatment plans and determine when genetic testing should be performed. This utility function includes the health benefits and harms from treatment as well as the cost of fatal and non-fatal ASCVD events, the cost of treatment, and the cost of genetic testing. We use willingness to pay thresholds of \$50,000/QALY and \$100,000/QALY, to convert from QALYs to costs and include all QoL weights and costs as the rewards of the model.

3.6.3 Results

We first evaluate the number of replications needed in our simulation to capture the heterogeneity in our population. To understand how treatment plans change after the addition of genetic information, we examine the number of people that receive a genetic test over the 10-year horizon. We then study the impact of age and risk in the VoI. We investigate how likely patients are to receive a genetic test at each year of our study. Then, we inspect the difference of treatment plans derived with clinical information only and with clinical and genetic information. Lastly, we evaluate the lifetime effects of performing genetic testing. We show the overall trend in our results using local regression and provide 95% confidence intervals assuming that the errors are normally distributed.

To study the impact of genetic testing, we divide our population by sex, race, and age groups. We create groups on the basis of the quartiles of the empirical distribution of the age of the patients at the first year of our study. Age groups are defined as the ages between two quartiles of the empirical distribution. We obtain the following age groups:

40-45, 46-50, 51-55, and 56-60 years old.

3.6.3.1 Selection of Number of Replications

The results of our convergence analysis are included in Appendix B.4. By simulating the health trajectory of every patient in the population under a single GRS per patient, we find that $K = 500$ health trajectory replications may be sufficient to represent the variation in our population (Fig. B.2 in Appendix B.4). Fixing the $K = 500$, we observe that $G = 100$ GRS realizations may be enough to capture the heterogeneity in the potential effect of genetic information (Fig. B.3 in Appendix B.4). Finally, fixing $G = 100$, we observe that the average QALYs saved and cost saved start to converge around 500 health trajectory replications with small variation (Fig. B.4 in Appendix B.4). Based on these observations, we choose to run the simulation $K = 500$ with $G = 100$ times for all analyses.

3.6.3.2 Genetic Tests Performed

Out of a population of 64.81 million patients, 6.73 million patients receive a genetic test throughout a 10-year horizon. Fig. 3.11 shows the distribution of the initial risk for ASCVD events of patients that receive genetic testing by age group (40-45, 46-50, 51-55, and 56-60). A total of 2.28 out of 17.86 (12.73%), 1.77 out of 15.73 (11.25%), 1.35 out of 16.02 (8.40%), and 1.34 out of 15.20 (8.79%) million patients receive genetic testing in age groups 40-45, 46-50, 51-55, and 56-60 years old, respectively. We observe that no patient with initial risk for ASCVD events below 0.0164% or above 0.1690% undergoes genetic testing. In general, younger patients are more likely to receive genetic testing at lower initial risk than older patients. A reason for this may be that the benefit gained by avoiding ASCVD events (due to the changes in treatment after genetic testing) in younger patients is greater than in older patients.

In additional analyses, we find that racial and sex differences play a significant role on the genetic testing strategies. For example, African American male patients are substantially less likely to receive genetic testing according to the VoI than any other sub-population. This may be due to the fact that African American male patients tend to have considerably higher risk scores than African American female or Caucasian patients. These elevated risk scores lead to high treatment intensity with or without genetic information, and there is no benefit from performing a genetic test.

From this point forward we focus our analyses on the sub-population that undergoes genetic testing. We begin by studying the effect of age and risk on the VoI of genetic testing (Fig. 3.12). The total VoI of genetic testing is 1.18 billion 2019 USD across our

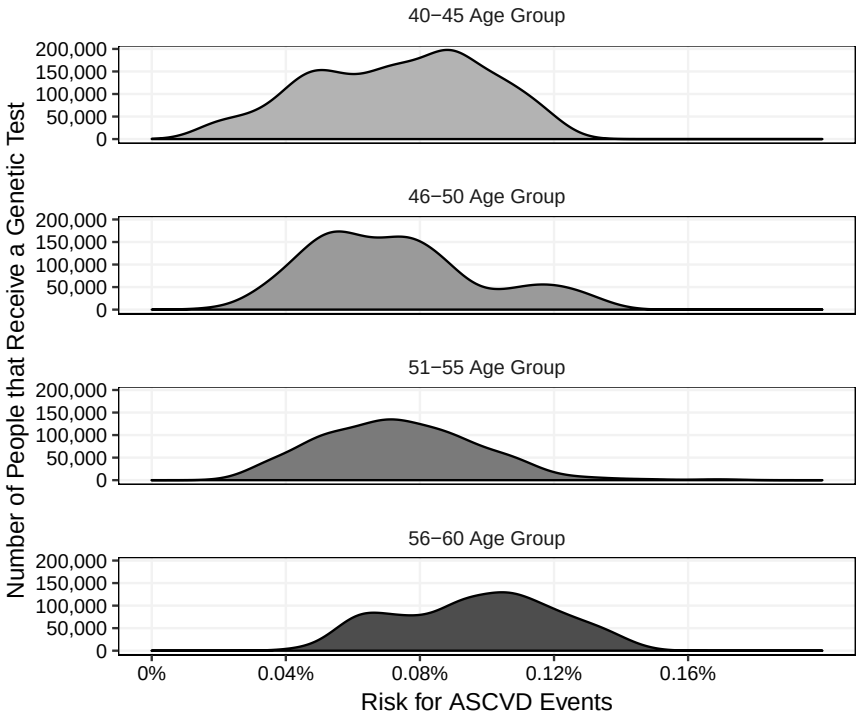


Figure 3.11: Distribution of the initial risk for ASCVD events of patients that receive genetic testing by age group.

population. This results in cost savings of 325.22, 321.82, 268.36, and 261.59 million 2019 USD among patients that are 40-45, 46-50, 51-55, and 56-60 years old, respectively. We observe that patients with moderate risk for ASCVD events due to clinical factors have the highest chance of changing treatment plans after the addition of genetic information. These patients receive the largest benefit from genetic testing. In addition, we notice that the distribution of the VoI peaks at higher initial risk for ASCVD events as age increases. This may be a consequence of the positive correlation between age and the risk for ASCVD events.

We develop our testing strategies based on the VoI of genetic information. Each panel in Fig. 3.13 shows the year when patients receive genetic testing by age group. Patients with the earliest time of testing in Fig. 3.13 generally achieve the largest VoI in Fig. 3.12. We also observe that valleys in Fig. 3.13 correspond to peaks in Fig. 3.12.

Fig. 3.13 shows an overall decreasing trend in the year that patients receive genetic testing as their risk for ASCVD events increases for patients with initial risks below 0.10%. A reason for this is that patients with lower initial risk for ASCVD events are typically more likely to receive genetic testing to avoid potential ASCVD events. Genetic information in these patients often leads to higher treatment intensity at earlier years. The testing year

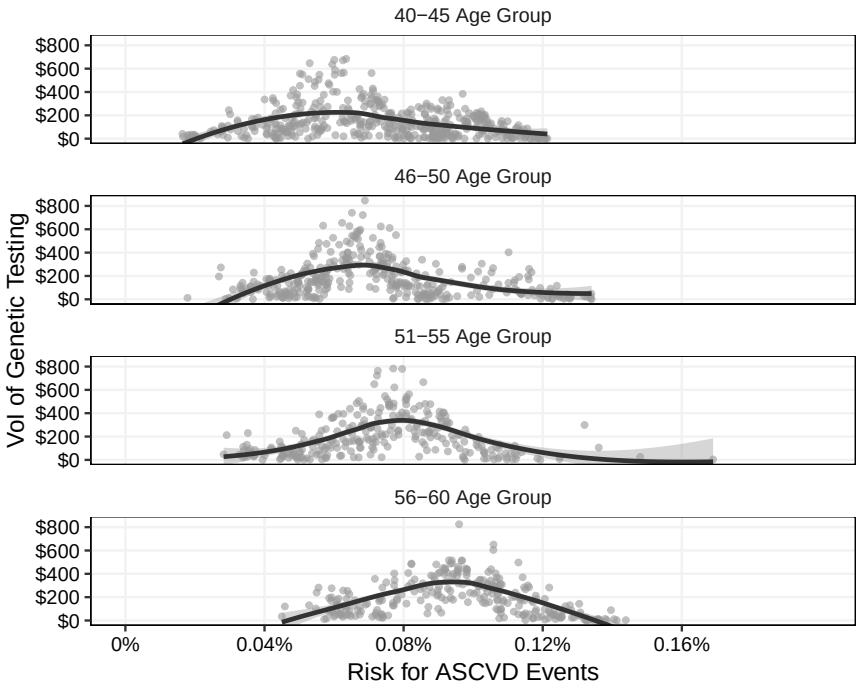


Figure 3.12: Distribution of VoI of genetic testing with respect to the initial risk for ASCVD events by age group in the population that received genetic testing. Each point represents 5,000 patients. Smoothed line are obtained using second degree local regression with a span of 90%. Shaded areas around the smoothed lines represent 95% confidence intervals around the mean.

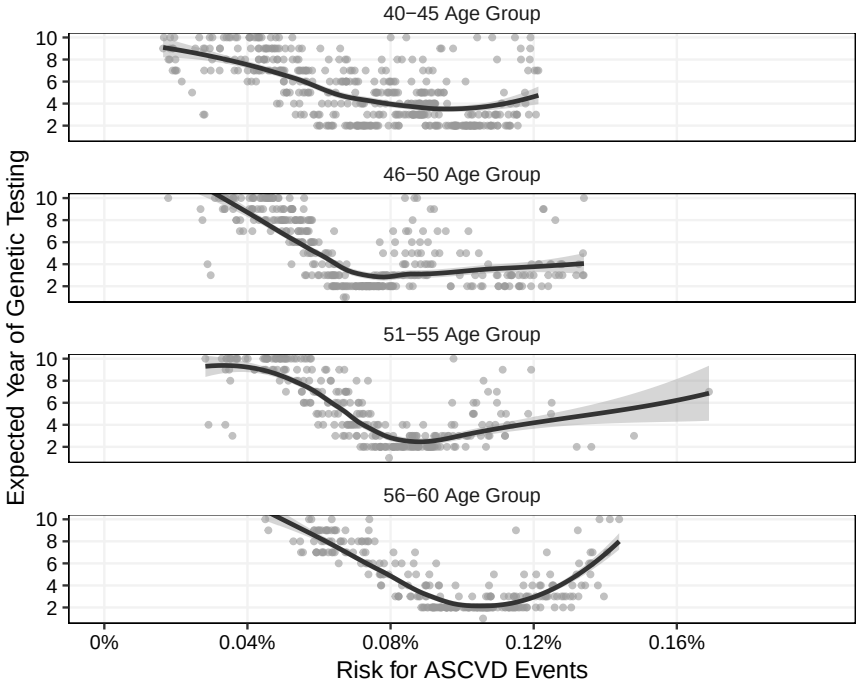


Figure 3.13: Distribution of the expected year of genetic testing with respect to the initial risk for ASCVD events by age group. Each point represents 5,000 patients. Smoothed line are obtained using second degree local regression with a span of 90%. Shaded areas around the smoothed lines represent 95% confidence intervals around the mean.

of patients with initial risks above 0.10% tends to increase as their initial risk for ASCVD events increases. Patients with higher initial risk for ASCVD events tend to receive the same amount of treatment by both policies until genetic testing results in lowering the treatment intensity. This normally results in later testing years.

3.6.3.3 Effect of Genetic Testing

We now proceed to study the difference in treatment plans derived with clinical information only and with clinical and genetic information. Fig. 3.14 shows an overall comparison of the PCE and GenePCE risks at the first year of our study. Recall that the policies derived with clinical information only are informed by the PCE risk score and the policies derived with clinical and genetic information are informed by the GenePCE risk score. While the overall event rate in the population remains unchanged after genetic testing, we observe that nearly 60% of the PCE risk scores are higher than the GenePCE risk scores. This causes the policies derived with clinical information only to be unnecessarily more intense than the policies informed with clinical and genetic information. In other words, the policies derived with clinical information only are over-treating patients. In the re-

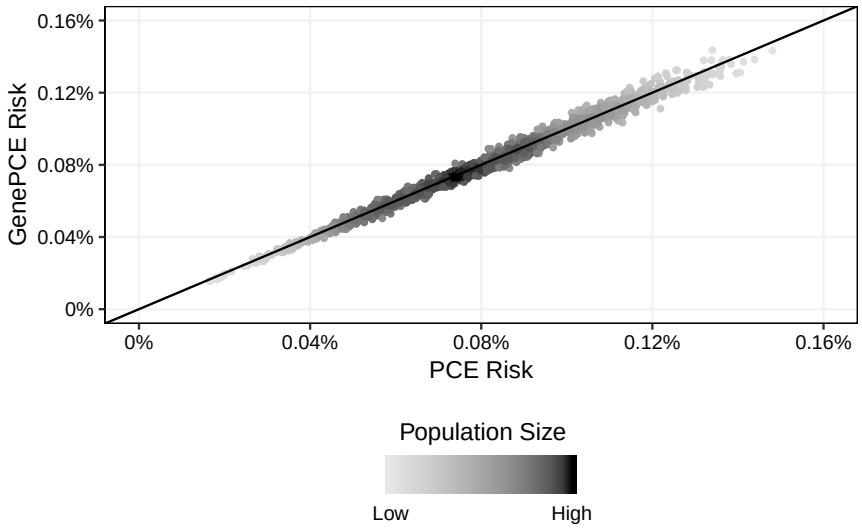


Figure 3.14: Comparison of PCE and GenePCE risk scores. The population size is represented with the color gradient (the darker the color the bigger the population). Equal risk scores up to the fifth decimal point are compressed into single point.

maining 40% of patients, genetic information helps identify individuals at increased risk for ASCVD events. This results in more aggressive treatment policies after genetic testing, which translates in fewer ASCVD events. Genetic testing helps identify which patients can receive lower treatment intensity without experiencing any additional ASCVD events.

The impact of acquiring genetic information on the treatment policies over the planning horizon is depicted in Fig. 3.15. We find that the policies informed with clinical information only treat more intensively than the policies derived with clinical and genetic information. This trend is also observed when patients are segregated by age, race, or sex. In additional analyses, we notice that both treatment policies are increasingly more intense as the risk for ASCVD events increases. Furthermore, most patients with a risk for ASCVD events above 0.11% are treated with high intensity statin drugs, regardless of whether or not they receive genetic testing. In contrast, patients with a risk for ASCVD events below 0.0255% do not receive any treatment.

3.6.3.4 Benefit of Genetic Testing

Once a genetic test is performed on a patient, clinical and genetic information becomes available. Genetic information may then change the treatment plans over the planning horizon. By performing genetic testing on 6.73 million people, the policies derived with clinical and genetic information save 5,487 more QALYs than the policies derived with clinical information only. While genetic information provides benefits in terms of health

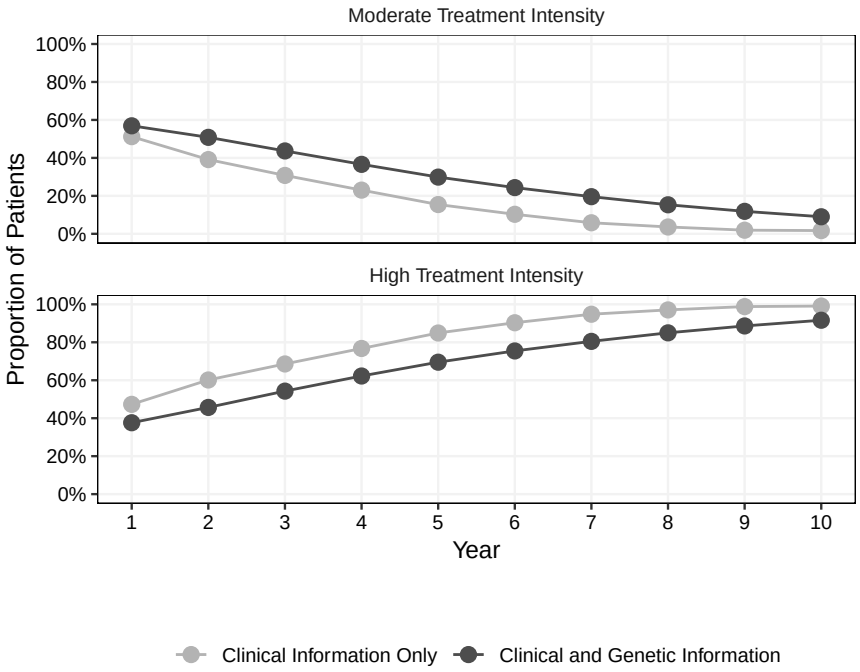


Figure 3.15: Proportion of patients receiving each treatment intensity by treatment policy over the planning horizon. Points represent the ratio of average number of patients treated with each statin intensity and the total population that receives genetic testing.

outcomes, the biggest benefit of obtaining genetic information are in terms of cost-savings.

The cumulative cost savings over time due to genetic information is illustrated in Fig. 3.16. Even though genetic testing has an overall cost of $1.18 billion, the policies informed with clinical and genetic information result in savings related to the cost of treatment and the cost of ASCVD events. By reducing the amount of treatment that patients receive, the cost of treatment decreases from $22.00 billion in the policies derived with clinical information to $19.67 billion in the policies derived with clinical and genetic information. On the other hand, since genetic testing provides information about which patients should receive higher treatment intensity, the cost due to ASCVD events decreases from $3.80 billion to $3.77 billion using clinical and genetic information. This leads to cost savings of 1.18 billion 2019 USD by the end of the 10-year planning horizon. Similar cost savings are observed when we include the lifetime effects of each policy.

Segregating our results by age group, we find that patients with ages above 51 years old avoid the biggest number of ASCVD events after genetic testing. Patients in the 40-45 age group get the highest number of additional QALYs saved and cost-savings. Dividing the population by race and sex, we discover that Caucasian females receive the largest amount of ASCVD events prevented, QALYs saved, and cost-savings.

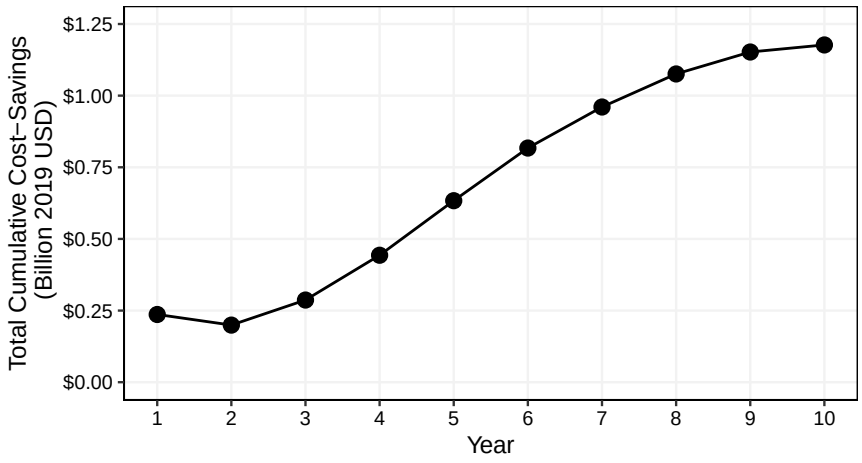


Figure 3.16: Total cumulative cost-savings due to genetic testing. Points represent the total cumulative cost savings by each year.

3.6.3.5 Sensitivity Analyses

We proceed to study how the genetic testing strategies are affected by changing model parameters and assumptions. Our results are summarized in Table 3.10. We segregate the genetic testing year results by age group and include them in Appendix B.5.

If the testing threshold is changed from zero to the expected cost-savings of preventing a CHD event, 3.36 million patients receive genetic testing instead of the 6.73 million patients in the base case. This is equivalent to the number of people that could potentially avoid an ASCVD event (2.53 million patients) plus the number of people receiving savings in the cost of treatment greater than their expected cost of an CHD event (830 thousand patients).

Changing the cost of genetic testing has considerable effects on the cost incurred by the policies derived with clinical and genetic information and, in turn, in the VoI of genetic testing. This causes moderate changes in the genetic testing strategies. We notice that genetic testing is performed generally earlier than in the base case if the cost of testing is lower than in the base case. The opposite effect is observed if the cost of testing is $400 instead of $200. The impact of the cost of genetic testing in the testing strategies by age group is included as Fig. B.5 in Appendix B.5.

We find that changes in treatment costs also have a significant impact in the testing strategies. For instance, if the cost of statin drugs is reduced in half, fewer patients receive genetic testing and typically in later years than in the base case. If the cost of statin drugs is doubled, genetic testing occurs for expected cost-savings greater than any treatment difference or to potentially prevent an ASCVD event. As a consequence, more

Table 3.10: Summary of sensitivity analyses.

Sensitivity analysis scenario	Number of people tested[g]	Genetic testing cost[h]	Overall cost savings[h]	Genetic testing year[i]			
				Q1	Q2	Q3	Avg.
Base case	6.73	$1.18	$1.18	2.00	4.00	7.00	4.97
Genetic testing threshold							
Expected cost of CHD savings	3.36	$0.58	$0.63	5.00	6.00	8.00	6.46
Genetic testing cost							
$0	36.09	$0.00	$2.88	1.00	2.00	5.00	3.58
$50	11.45	$0.53	$2.25	1.00	2.00	6.00	3.56
$100	9.36	$0.85	$1.81	2.00	3.00	6.00	4.11
$400	3.56	$1.21	$0.45	4.00	6.00	8.00	5.99
Treatment cost: moderate intensity (high intensity)							
$72 ($225)	3.62	$0.62	$0.23	4.00	6.00	8.00	6.00
$12 ($119)	1.89	$0.31	$0.06	5.00	7.00	9.00	6.92
$12 ($2454)	10.78	$2.00	$18.24	1.00	2.00	5.00	3.22
$288 ($900)	9.27	$1.68	$3.59	2.00	3.00	6.00	4.08
Treatment disutility: moderate intensity (high intensity)							
0.0005 (0.002)	8.21	$1.44	$1.59	2.00	4.00	7.00	4.97
0.001 (0.004)	7.14	$1.24	$1.43	3.00	5.00	8.00	5.40
0.00384 (0.0075)	6.28	$1.09	$1.45	3.00	5.00	8.00	5.48
0.01 (0.02)	2.80	$0.47	$0.34	4.00	7.00	9.00	6.53
Cost-QALY utility function: willingness to pay threshold							
$50,000/QALY	10.34	$1.94	$16.30	1.00	2.00	4.00	2.87
$100,000/QALY	13.17	$2.49	$15.15	1.00	1.00	3.00	2.64

[g] The number of people that received a genetic test are presented in millions.
[h] The cost associated with genetic testing and overall cost savings are cumulative estimates over the 10-year planning horizon (not including the expected lifetime after the planning horizon) and are presented in billion 2019 USD.
[i] Q1, Q2, and Q3 represent the first, second (median), and third quartiles of the empirical distribution of the genetic testing years, respectively. Avg. stands for the empirical average.

patients receive genetic testing and generally earlier in the planning horizon. The effect of treatment cost in the testing strategies segregated by age group is shown in Fig. B.6 in Appendix B.5.

In general, we observe that the testing strategies are sensitive to changes in treatment disutility. For example, decreasing the treatment disutility for moderate intensity statin drugs leads to more genetic tests performed at approximately the same time as in the base case. Increasing the treatment disutility for high intensity statin drugs has a similar effect. Nonetheless, genetic testing is performed typically later in the planning horizon. Dividing our population into age groups, we note that there is no substantial difference in the effect of treatment disutility in the testing strategies by patients' age (Fig. B.7 in Appendix B.5).

Using a utility function that incorporates cost and QALYs to derive the treatment policies and testing strategies results in a noticeably different number of people that receives genetic testing. This is mainly due to the difference in the risk scores that informed the treatment policies. Since the policies derived with clinical and genetic information are informed by risk scores that are generally lower than the risk scores that inform the policies with clinical information only, these policies observe much lower ASCVD events rates. Lower event rates translate into higher QALYs and lower costs in expectation. Combining these two factors into a single objective function leads the policies derived with clinical and genetic information to be less intense than in the base case. Therefore, more people receive genetic testing due to lower treatment intensity. We observe that, while a willingness to pay threshold of \$50,000/QALY achieves greater cost savings over the 10-year planning horizon, a willingness to pay threshold of \$100,000/QALY results in more genetic tests performed, and those tests are performed earlier in the planning horizon.

3.6.4 Discussion

The development of the new GRS that serves as an aid in the prediction of CHD could make genetic information vital for the management of ASCVD. Cholesterol treatment plans should incorporate this information to ensure that patients receive the best care available. A framework to determine when a genetic test should be performed and how a patient should be treated before and after genetic information could provide clinicians with essential information.

In this section, we presented a framework to incorporate cholesterol treatment plans and genetic testing strategies. We applied this framework to a large sample representative

of an adult population in the US to quantify the effect of adding genetic information into optimal treatment plans. This simulation framework allowed us to evaluate treatment plans and decide when to perform genetic tests.

Several conclusions can be made from this study. First, we can conclude that not everyone benefits from genetic testing. Through VoI analysis, we were able to identify patients that would benefit from genetic information. This benefit translates to better health outcomes and cost savings. Second, the reasons to perform genetic testing vary by age, race, sex, and risk. As a consequence, the gains obtained from genetic information also vary by age, race, sex, and risk. Third, who receives genetic testing and when changes according to the genetic testing threshold θ. Different values of θ can provide clinicians with a wide range of testing strategies, depending on the magnitude of losses they are willing to accept before recommending a test to gather genetic information. Lastly, the population that receives genetic testing and the time they undergo testing changes with the cost of genetic testing, the cost of treatment, and treatment disutility. As genetic testing becomes more affordable and statin drugs decrease in price and side effects, a genetic test to inform cholesterol treatment plans may benefit a greater population.

We acknowledge that simulation models are generally more complex and require longer running times than other models, such as Markov cohort models. Nonetheless, our simulation allowed us to represent the variability in our population and provided us with great flexibility to assess different aspects of our model. Furthermore, we cannot guarantee the optimality of our testing strategies. The results of simulation models can change as the number replications in the simulation increases. By assessing the convergence of the main outcomes of the simulation, we ensured that the heterogeneity in our population was modeled appropriately.

From a methodological perspective, this work could be extended by acknowledging the uncertainty inherent in the risk estimates of our modeling framework. As the PCE risk score was derived with a proportional hazards model, this risk could be modeled from a Bayesian perspective. The coefficients obtained by Goff and coauthors could serve as prior information for a Bayesian proportional hazards model (Ibrahim et al., 2001; Goff et al., 2014). A robust MDP formulation with an scenario uncertainty model could be used to find robust cholesterol treatment plans (Ben-Tal et al., 2009).

From a clinical perspective, this work could be extended by incorporating other cholesterol-lowering drugs, such as PCSK9 inhibitors or fibrates. We chose statin drugs because they are the most used cholesterol-lowering drugs with the greatest public health impact. Our work could also be extended by incorporating the effect of genetic testing on aspirin use or hypertension treatment. However, the clinical benefit of aspirin is now

unclear and risk prediction plays a smaller role in hypertension treatment decisions (McNeil et al., 2018; Williams et al., 2018; James et al., 2014). Based on clinical expertise, we decided to develop genetic testing strategies to inform cholesterol treatment plans as a starting point. Incorporating the effect of genetic information on aspirin use or hypertension treatment would likely result in a greater number of people undergoing genetic testing. In this sense, our results provide a lower bound on the benefits of performing genetic testing according to the VoI.

Another clinical extension of this work could be a probabilistic sensitivity analysis of the input parameters. Other studies have used different values to parameterize models for cholesterol management (Pandya et al., 2015; Kazi et al., 2016). Particularly, the costs and QoL weights parameters can be highly variable from study to study. Nevertheless, this analysis would require assumptions on the distribution of the input parameters, which can be highly subjective.

3.7 Conclusions

We conclude with clinical and public policy recommendations derived from this work. In practice, clinicians may prefer to recommend genetic testing after the initial medical appointment of patients (represented by the beginning of year 0 in our simulation). Since our framework serves to identify which patients could benefit from genetic information, physicians can use this information to determine who to test. Practically, patients that would benefit from genetic information can receive the test at any time before the recommended year.

Outside of the direct adoption of the treatment policies and testing strategies, doctors and health insurance policy makers can use our results to gain a better understanding of the potential benefits of performing genetic testing in different populations. According to the US Department of Health and Human Services, health insurance plans will typically cover the costs of genetic testing if "the test result will directly influence the disease treatment management of the covered member", among other criteria (United States Department of Health and Human Services, 2006). Our models could serve as an aid for health insurance policies by identifying patients whose cholesterol treatment will likely change after acquiring genetic information.

As genetic testing to obtain a GRS for CHD is not commonly performed in practice yet (Lewis and Vassos, 2020), our simulation results provide an estimate of the policy implications of different cholesterol treatment policies and genetic testing strategies in clinical practice. The role of genetics in many applications remains promising, but somewhat

unclear. As precision medicine becomes increasingly important, the understanding of the potential impact of genetics on disease management becomes crucial.

Chapter 4

Data-Driven Ranges of Near-Optimal Actions for Finite Markov Decision Processes

In this chapter, we propose a new framework for identifying sets of near-optimal actions for finite MDP models. We first present a SBDP algorithm that can be executed using parallel computing and show that it converges to the optimal solutions exponentially fast and that it conserves the structural properties of standard dynamic programming asymptotically. Then, we introduce a simulation-based MCC method to identify actions that are not statistically different from an approximately optimal action. By analyzing the structure of the sets, we characterize their behavior with respect to the modeling data and identify when they can be ordered as a range. Finally, we show the scalability of our approach by finding ranges of near-optimal antihypertensive treatment choices for 16.72 million adults in the US.

4.1 Background

MDP models have been used to inform decisions in a wide variety of applications including medicine, scheduling, transportation, finance, and energy (Boucherie and van Dijk, 2017). In many areas of application, such as the management of chronic conditions or perishable inventory, the dynamics of the system of interest change over time. This type of problems generally have a finite set of periods when decisions must be made. If all the parameters of a non-stationary MDP are known with certainty, and there are a finite number of states and actions, the backwards induction algorithm can be used to find an optimal decision rule (Puterman, 2014; Chang et al., 2013). However, there may be other decision strategies that we cannot statistically differentiate from the optimal. For situations where an MDP is used as a decision support tool, it is indispensable to allow for human judgment to transform the decision rules into practice.

When a human being is responsible for controlling a system, a single decision rule may not be enough, as each person has their own decision process (Fard and Pineau, 2011). It may be appropriate to assume that some aspect of the decision making process will be influenced by the decision maker (DM). Moreover, the difference between the performance of an optimal decision rule and other strategies may be negligible. The DM could choose between an optimal action and another action with similar performance based on their expertise, preference, or other factors. In addition, models are typically estimated from observational data and multiple external sources. This may lead to optimal decision rules that do not perform well in the true system (Mannor et al., 2007). The performance of an optimal decision rule may not be statistically different from other strategies. To test for statistical significance before observing the implications of each action in practice, our proposed strategy is to simulate the effect of each action based on the estimated model of the system of interest. We can then provide DMs with a set of actions, each of which might be the optimal, but we do not have enough evidence to differentiate.

In this chapter, we are motivated by circumstances in which several actions may have similar performance. We focus on improving the usability and acceptance of MDP models in practice by providing flexibility in the implementation of decision strategies. Rather than offering a single decision rule, we present DMs with a set of actions from which they may be able to choose from. We introduce a new method to obtain sets of near-optimal actions and provide conditions for which the actions in these sets can be ordered as a range.

4.1.1 Applications of Markov Decision Processes to Medical Decision-Making

The focus of this chapter will be medical decision-making, as incorporating all the factors that influence the decision-making process is a challenging task and it may be hard to implement decision rules in practice. Overviews of decision models for the management of different diseases can be found at Denton et al. (2011), Capan et al. (2017), Saville et al. (2018), and Chanchaichujit et al. (2019). Treatment decision models in the literature include Long et al. (2008); Lee et al. (2008); Chan et al. (2013); Liu et al. (2017); Negoescu et al. (2017, 2018); Ayer et al. (2019), and Chehrazi et al. (2019). Other studies have focused on finding the optimal time to gather additional information or for a screening procedure including Maillart et al. (2008); Chhatwal et al. (2010); Zhang et al. (2012); Erenay et al. (2014); Skandari et al. (2015); Helm et al. (2015); Deo et al. (2015); Ayer et al. (2016); Sabouri et al. (2017); Hicklin et al. (2018); Suen et al. (2018); Agnihothri et al. (2018); Onen et al.

(2018); Lee et al. (2018); Lin et al. (2018), and Aprahamian et al. (2019, 2020). Despite the ability of MDP models to model sequential decision-making in complex medical settings, they typically cannot be implemented in their current form as they lack flexibility in their decision rules.

In medical decision-making situations, the transition dynamics and rewards are often estimated using longitudinal observational patient data and results from the medical literature. As longitudinal data and medical results are derived with a finite number of observations, these estimates are subject to statistical uncertainty (Steimle et al., 2019). But even if we had complete knowledge of the transition dynamics and rewards, the selection of clinical interventions will typically depend on physicians' judgment and their patients' preferences.

Translating medical decision-making models into practice is difficult. General medical practitioners may interpret decision rules as cumbersome, confusing, and lacking in credibility (Cabana et al., 1999; Cohen and Townsend, 2018; Grundy et al., 2019). Therefore, it is important to consider practical implications in the design of decision rules (Classen and Mermel, 2015). One way such implications can be considered is by providing clinicians with flexibility in the implementation of protocols, while continuing to improve patient outcomes.

An example, that we will study in detail, where clinicians may benefit from flexibility is in the management of ASCVD (constituting CHD and stroke). According to the National Vital Statistics, ASCVD is considered among the leading cause of death in the US (Kochanek et al., 2019). The Heart Disease and Stroke Statistics 2020 Update reports that CHD and stroke account for 42.6% and 17.0% of deaths attributable to cardiovascular diseases in the US, respectively (Virani et al., 2020). One of the main controllable risk factors of ASCVD is hypertension or high BP. The most recent hypertension guidelines from the American College of Cardiology and American Heart Association (Whelton et al., 2018) have generated considerable controversy among practitioners (Ioannidis, 2018; Cohen and Townsend, 2018; Solberg and Miller, 2018; Wilt et al., 2018). Further, these guidelines provide conflicting recommendations regarding when to initiate pharmacological interventions with other guidelines such as Williams et al. (2018).

Controversy and conflicting recommendations aggravate the already difficult problem of deciding how to manage patients' BP. Moreover, clinicians' opinion and patients' inclinations may influence the selection of treatment plans, independently of the guidelines' suggestions (Cabana et al., 1999). To benefit from physicians' expertise and account for any potentially conflicting recommendations, we design personalized data-driven ranges of treatment options that are within a margin of certainty of the best treatment alternative,

based on the estimated transition dynamics and rewards.

4.1.2 Modeling Approach

The problem of providing clinicians with options to treat their patients motivates the general research topic of adapting stochastic dynamic programming methods to offer multiple actions per state and time period. The need for flexibility in the implementation of decision strategies is not specific to medical domains. For example, this requirement may be needed for a decision support system that provides suggestions about how to manage the inventory levels of a product. To provide easily interpretable decision support, this problem also triggers the need to obtain insights into how the sets of near-optimal actions behave with respect to the modeling data. In this chapter, we present a new approach to address these topics based on SBDP, nonoverlapping batch means, and statistical MCC.

We choose SBDP to estimate action-value functions because it does not require knowledge of the true underlying probability distribution of the evolution of the system of interest. Instead, SBDP relies on sample realizations of the transition dynamics, which may be obtained through simulation. The SBDP framework also allows us to estimate optimal actions with high degree of accuracy, which serve as controls in the multiple comparison procedures. Rather than using re-sampling techniques, we divide the output of simulations into batches because the sample size of a simulation model can always be managed by the DM. We use MCC because, once an optimal action is identified as the control, we are interested in comparing the remaining alternatives with such control. The MCC method requires the least number of comparisons and the strongest inference for our purposes.

4.2 Organization of the Chapter

This remainder of this chapter is organized as follows. We begin by providing a review of the relevant literature in Section 4.3. In Section 4.4, we provide additional background on MDP models and MCC. We formally define the sets and ranges of near-optimal actions in Section 4.5. In Section 4.6, we introduce our algorithms to obtain the sets of near-optimal actions as well as our analysis of the algorithms. We present our case study of flexible hypertension management in Section 4.7. Finally, conclusions and future research directions are discussed in Sections 4.8 and 4.9.

4.3 Literature Review

The relevant literature to this research lies in the following fields: (1) simulation-based algorithms for MDP models; (2) batching and independent replication methods for the output analysis of simulation models; (3) statistical multiple comparisons approaches; (4) decision support models that provide sets of actions; and (5) treatment decision models for the management of cardiovascular diseases. In this section, we highlight some prominent papers in each category and briefly describe how our proposed methodology differs from them.

The description of an MDP as a simulation model can be found in Hernández-Lerma and Lasserre (1996) and Chang et al. (2013). A large portion of simulation-based algorithms have focused on solving MDP models with large state spaces or in which there is no model of the system dynamics. These methods principally fall under the umbrella of approximate dynamic programming (ADP) or reinforcement learning (RL). Summaries of ADP/RL methods include Powell (2011), Bertsekas (2012), and Sutton and Barto (2018). Other simulation-based models have concentrated on solving MDP models with large action spaces Chang et al. (2013). A key difference between the model-based methods in the ADP/RL literature and our SBBI algorithm is that, while they simulate the system dynamics as episodes, our algorithm simulates each time period independently. Closest to our work, there have been methods that use simulation to estimate the expectation in dynamic programming with small to moderately sized state and action spaces (Powell, 2011). In discounted infinite-horizon settings, Haskell et al. (2016) introduced simulation-based value iteration and policy iteration algorithms. Our method differs from the approaches presented by Haskell et al. (2016) in that we focus on finite horizon and allow for random immediate rewards. Another closely related area to our work is the sample-average approximation in discrete stochastic programming (Kleywegt et al., 2002). Our SBBI algorithm is different from standard multistage discrete stochastic programming models in that actions affect the system dynamics in the future.

The idea of grouping the output of simulation models to derive confidence intervals for the mean of a simulation outcome was introduced by Fishman (1978). There are two notable grouping approaches in steady-state simulation analysis: nonoverlapping batch means and independent replications. Alexopoulos and Goldsman (2004) compared the benefits and drawbacks of batching a single long simulation run and executing independent replications. In our context, both approaches are equivalent as there is no initialization bias and the observations are independent due to the Markov property. The nonoverlapping batch means and independent replication methods usually

rely on asymptotic normality arguments to obtain confidence intervals (Goldsman, 1992; Alexopoulos and Seila, 1996). Steiger and Wilson (2002) and Steiger et al. (2002, 2005) introduced algorithms to attain confidence intervals of a specific precision hinging on normality and independence tests. In our work, we present a nonparametric approach to obtain confidence intervals in steady-state simulation analysis without any distributional assumptions.

Our work is also related to the theory of statistical multiple comparisons. Overviews of multiple comparisons procedures can be found in Hsu (1996) and Toothaker (2012). Among the types of multiple comparisons, MCC is the most relevant to this research (Dunnett, 1955). Similar to many classic statistics methods, MCC assumes normality and equal variances across alternatives. Westfall and Young (1993) proposed a generalization of multiple comparisons procedures that allowed for general distributions using the bootstrap. This approach was later applied in the context of MCC by Westfall (2011). Another line of work focused on developing MCC methods without the assumption of equal variances (Li and Ning, 2012). Even though these alternative formulations allow for general distributions or unequal variances, none of them allow for both of them. Our SBMCC algorithm is a nonparametric approach that does not require equal variances.

There has been limited research in the area of decision support models that provide more than one decision rule. Laber et al. (2014) developed sets of decision rules in the context of dynamic treatment regimes based on clinically significant differences. A crucial distinction between this work and ours is that we focus on statistical significance instead of practical significance. Another difference is that we focus on Markov policies whereas they center on history dependent policies. This allows us to consider more than two decision epochs and actions. Ertefaie et al. (2016) also considered the problem of providing a set of suggestions in the context of dynamic treatment regimes. Although our approach has many similarities with this work, a vital distinction is that we identify a control before the statistical inference. This results in fewer comparisons and improved statistical power. Ertefaie et al. (2016) also concentrated in 2-stage history dependent policies while we focus on Markov policies over a finite planning horizon. The closest work to this article is by Fard and Pineau (2011), where the authors consider the problem of developing sets of near-optimal actions for discounted infinite-horizon MDP models. Compared to this work, we define a new sense of near-optimality in terms of statistical significance whereas they specify their near-optimality in the same units of the value function. A key contribution of our work to this field is that we characterize the behavior of the sets with respect to the modeling data, including the case where the actions contained in the set can be ordered as a range.

Treatment decision models for patients at risk of cardiovascular diseases include Hauskrecht and Fraser (2000); Stanford R.E. (2004); Cooper et al. (2006); Lee et al. (2018), and Zargoush et al. (2018). Researchers have also developed MDP models for the management of cardiovascular diseases (Denton et al., 2009; Kurt et al., 2011; Mason et al., 2014; Schell et al., 2016; Steimle et al., 2019). In contrast to these models, which recommend a single decision rule, our work provides physicians and their patients with flexibility in the implementation of treatment plans.

4.4 Preliminaries

In this chapter, we focus on finding a sequence of sets of near-optimal actions in the context of discrete-time finite-horizon MDP models with finite state and action spaces. This section introduces the main notions behind finite MDP models and MCC as well as our mathematical notation.

4.4.1 Markov Decision Processes

MDP models are used to model the interactions of a DM with a fully observable system of interest. At a specific time $t \in \mathcal{T}$, the DM observes the state of the system $s \in \mathcal{S}$ and chooses an action $\pi_t(s) \in \mathcal{A}$, according to a decision rule $\pi_t : \mathcal{S} \mapsto \mathcal{A}$. It is assumed that the system is always in a state and that the evolution of system only depends on the current state and action. A sequence of decision rules over all decision epochs $t \in \mathcal{T} \setminus \{T\}$ is called a policy $\pi := (\pi_t : t \in \mathcal{T} \setminus \{T\})$. After an action $\pi_t(s) = a$ is chosen, the DM receives a reward $r_t(s, a, \omega) \in \mathbb{R}$, where $\omega \in \Omega$ is an outcome of an exogenous process representing the uncertainty in the system. The system then evolves to state $s' \in \mathcal{S}$ according to a transition function $f_{t+1}(s, a, \omega)$. This process continues over a set of decision epochs $\mathcal{T}$ until the DM chooses the last action at time $T - 1$. The system finally evolves to state $\bar{s} \in \mathcal{S}$ at time T, when the DM receives a terminal reward $r_T(\bar{s}, \omega) \in \mathbb{R}$. If the DM prefers rewards now over rewards in the future, a discount factor $\gamma \in (0, 1]$ may be included in the formulation.

The goal of the DM is to find the policy π that maximizes the total expected discounted reward over the planning horizon, i.e.:

$$v_0(s) := \max_{\pi} \mathbb{E}\left[\sum_{t=0}^{T-1} \gamma^t r_t(s_t, a_t, \omega_t) + \gamma^T r_T(s_T, \omega_T) \middle| s_0 = s\right],$$

where s_t, $a_t = \pi_t(s_t)$, and ω_t are the state, action, and random disturbance of the system

at decision epoch t, respectively, and the expectation is taken with respect to the joint distribution of $\omega_0, \ldots, \omega_T$. Dividing the problem into each decision epoch, we get the set of dynamic programming equations:

$$Q_t(s,a) \coloneqq \mathbb{E}\Big[r_t(s,a,\omega) + \gamma v_{t+1}\left(f_{t+1}(s,a,\omega)\right)|s,a\Big],$$

where $v_t(s) \coloneqq \max_{a\in\mathcal{A}} Q_t(s,a)$ and $Q_T(s,a) = v_T(s) = \mathbb{E}[r_T(s,\omega)|s]$. Starting from the terminal time period T and proceeding backwards until decision epoch 0 we can find the optimal set of actions $\mathcal{A}_t^*(s) \coloneqq \operatorname{argmax}_{a\in\mathcal{A}} Q_t(s,a)$ at each decision epoch t. An optimal decision rule at state s is given by $\pi_t^*(s) \in \mathcal{A}_t^*(s)$ and an optimal policy is defined as $\pi^* = (\pi_t^* : t \in \mathcal{T}\setminus\{T\})$.

In summary, a simulation MDP is formally defined by the tuple $(\mathcal{T}, \mathcal{S}, \mathcal{A}, f, r, \gamma)$. We use the following notation throughout the chapter:

t	index of discrete time periods; $t \in \mathcal{T}$, where $\mathcal{T} \coloneqq \{0,1,\ldots,T\}$ is a finite set of time periods. Decisions are made until time $T-1$; the time periods $\mathcal{T}\setminus\{T\}$ will be referred to as decision epochs.
s	state of the system; $s \in \mathcal{S}$, where $\mathcal{S} \coloneqq \{1,\ldots,S\}$ is a finite set of states.
a	DM's action; $a \in \mathcal{A}$, where $\mathcal{A} \coloneqq \{1,\ldots,A\}$ is a finite set of actions.
$r_t(s,a,\omega)$	reward associated with a state s, an action a, and an outcome of the exogenous process $\omega \in \Omega$, where the reward function is defined as $r : \mathcal{S}\times\mathcal{A}\times\Omega \mapsto \mathbb{R}_+ \coloneqq \{x\in\mathbb{R}\|x\geq 0\}$.
$f_{t+1}(s,a,\omega)$	transition function which produces the next state s' given a state s, an action a, and an outcome of the exogenous process $\omega \in \Omega$; $s' = f_{t+1}(s,a,\omega)$, where $f : \mathcal{S}\times\mathcal{A}\times\Omega \mapsto \mathcal{S}$.
$Q_t(s,a)$	action-value function associated with state s and action a at decision epoch t; $Q_t(s,a) \coloneqq \mathbb{E}\Big[r_t(s,a,\omega) + \gamma \max_{a'\in\mathcal{A}} Q_{t+1}\left(f_{t+1}(s,a,\omega),a'\right)|s,a\Big]$.
$v_t(s)$	value function for state s at decision epoch t; $v_t(s) \coloneqq \max_{a\in\mathcal{A}} Q_t(s,a)$.
$\mathcal{A}_t^*(s)$	set of optimal actions in state s at decision epoch t, where $\mathcal{A}_t^*(s) \subset \mathcal{A}$; $\mathcal{A}_t^*(s) \coloneqq \operatorname{argmax}_{a\in\mathcal{A}} Q_t(s,a)$.
$\mathcal{A}_t^-(s)$	set of sub-optimal actions in state s at decision epoch t, where $\mathcal{A}_t^-(s) \subset \mathcal{A}$; $\mathcal{A}_t^-(s) \coloneqq \mathcal{A}\setminus\mathcal{A}_t^*(s)$.
$\pi_t^*(s)$	optimal decision rules at state s; $\pi_t^*(s) \in \mathcal{A}_t^*(s)$. The sequence of optimal decision rules $\pi^* = (\pi_t^* : t\in\mathcal{T}\setminus\{T\})$ is referred to as a policy.
γ	discount factor of the model; $\gamma \in (0,1]$.

Note that any simulation MDP $(\mathcal{T}, \mathcal{S}, \mathcal{A}, f, r, \gamma)$ with transition function f and rewards r

can be transformed into a standard MDP $(\mathcal{T}, \mathcal{S}, \mathcal{A}, \boldsymbol{P}, \rho, \gamma)$ with transition probabilities $\boldsymbol{P} \in [0,1]^{|\mathcal{T}\times\mathcal{S}\times\mathcal{S}\times\mathcal{A}|}$ and rewards $\rho : \mathcal{S}\times\mathcal{A} \mapsto \mathbb{R}_+$. Simply let $p_t(s'|s,a) \coloneqq \mathbb{E}[\mathbb{1}\{f_{t+1}(s,a,\omega) = s'|s,a\}]$ and $\rho_t(s,a) \coloneqq \mathbb{E}[r_t(s,a,\omega)|s,a]$, where the expectation is taken over ω and $\mathbb{1}\{\cdot\}$ represents an indicator function.

4.4.2 Multiple Comparisons With a Control

When a control or standard action is available, a DM may be interested in comparing the performance of every other action with the performance of such control. In this section, we adapt the ideas in MCC to the context of simulated MDP models. The parameters of interest are $Q_t(s,a^*) - Q_t(s,a)$ for $a^* \in \mathcal{A}_t^*(s)$ and $a \in \mathcal{A}_t^-(s)$, at each decision epoch t and state s. In here, $\mathcal{A}_t^*(s) \subset \mathcal{A}$ denotes the set of optimal actions (potential controls) and $\mathcal{A}_t^-(s) \subset \mathcal{A}$ denotes the set of sub-optimal actions. If a^* is an action such that $Q_t(s,a^*) \geq Q_t(s,a)$ for any $a \in \mathcal{A}$, our objective is to identify as many actions as possible to be inferior than a^*. Assuming equal sample sizes across actions, the $1-\alpha$ simultaneous confidence lower bounds for the difference between a control $Q_t(s,a^*)$ and the remaining action-value functions $\{Q_t(s,a) : a \in \mathcal{A}\}$ are given by:

$$Q_t(s,a^*) - Q_t(s,a) > \hat{Q}_t(s,a^*) - \hat{Q}_t(s,a) - d_t(s,\alpha)\sqrt{N^{-1}\left[\hat{\sigma}_t^2(s,a^*) + \hat{\sigma}_t^2(s,a)\right]}, \tag{4.1}$$

where $\alpha \in (0,1)$, $\hat{Q}_t(s,a)$ is an estimate of $Q_t(s,a) < \infty$, N is the number of observations used to estimate $\hat{Q}_t(s,a)$, and $\hat{\sigma}_t^2(s,a)$ is an estimate of $\sigma_t^2(s,a) < \infty$, the variance of the action-value function associated with state s and action a. If there are reasons to believe that the action-value functions are normally distributed, $d_t(s,\alpha)$ can be obtained by solving a double integral whose numerical evaluations are readily available in standard statistical software Dunnett (1955). However, in most practical situations this assumption may not be true. Westfall and Young (1993) proposed an alternative formulation that allows for general probability distributions. Extending their formulation to unequal variances, it aims to find a constant $d_t(s,\alpha)$ such that:

$$\mathbb{P}\Big(\max_{a\in\mathcal{A}} \psi_t(s,a) \leq d_t(s,\alpha)\Big) = 1-\alpha, \tag{4.2}$$

where

$$\psi_t(s,a) \coloneqq \frac{\hat{Q}_t(s,a^*) - \hat{Q}_t(s,a) - (Q_t(s,a^*) - Q_t(s,a))}{\sqrt{N^{-1}\left[\hat{\sigma}_t^2(s,a^*) + \hat{\sigma}_t^2(s,a)\right]}},$$

is a root statistic corresponding to state s and action a.

4.5 Ranges of Near-Optimal Actions

From equation (4.1) we can conclude that an optimal action $a^* \in \mathcal{A}_t^*(s)$ is significantly better than some other action $a \in \mathcal{A}_t(s)$ at a significance level α if:

$$\hat{Q}_t(s,a^*) - \hat{Q}_t(s,a) - d_t(s,\alpha)\sqrt{N^{-1}\left[\hat{\sigma}_t^2(s,a^*) + \hat{\sigma}_t^2(s,a)\right]} > 0.$$

Thus, we cannot conclude that action $a^* \in \mathcal{A}_t^*(s)$ is significantly different from $a' \in \mathcal{A}_t(s)$ if:

$$\hat{Q}_t(s,a^*) - \hat{Q}_t(s,a') - d_t(s,\alpha)\sqrt{N^{-1}\left[\hat{\sigma}_t^2(s,a^*) + \hat{\sigma}_t^2(s,a')\right]} \leq 0.$$

This leads to our definition of a set of near-optimal actions.

Definition 4.1 *Given N observations, an optimal action $a^* \in \mathcal{A}_t^*(s)$ such that $Q_t(s,a^*) \geq Q_t(s,a)$ for $s \in \mathcal{S}$ and all $a \in \mathcal{A}$, and a quantile $d_t(s,\alpha)$, a set of actions $\Pi_t(s,\alpha)$ is said to be α-nonsignificant with $\alpha \in (0,1)$ if it satisfies:*

$$\Pi_t(s,\alpha) \coloneqq \left\{a \in \mathcal{A} : \hat{Q}_t(s,a^*) - \hat{Q}_t(s,a) \leq d_t(s,\alpha)\sqrt{N^{-1}\left[\hat{\sigma}_t^2(s,a^*) + \hat{\sigma}_t^2(s,a)\right]}\right\}.$$

A desirable property of $\Pi_t(s,\alpha)$ in practice is to be ordered according to the effect of the actions in the action space $\mathcal{A}_t(s) \subseteq \mathcal{A}$ at state s and decision epoch t. For example, clinicians may find a set of treatment choices more interpretable if the set follows a natural order such as an increasing number of medications. An ordered set of actions in the sense of Definition 4.2 will be referred as a range of actions.

Definition 4.2 *$\Pi_t(s,\alpha)$ is said to be a range of α-nonsignificant actions if it satisfies Definition 4.1 and if $a, a'' \in \Pi_t(s,\alpha)$ implies that $a' \in \Pi_t(s,\alpha)$ for any actions ordered as $a \leq a' \leq a'' \in \mathcal{A}_t(s)$.*

4.6 Solution Approach

In this section, we describe our approach to identify sets of α-nonsignificant actions. We first introduce our SBBI algorithm. Subsequently, we study the finite sample, convergence, and the asymptotic structural properties of the algorithm. We then present our SBMCC method. Lastly, we examine the asymptotic behavior of our method and characterize the sets of near-optimal actions. Our analytical results are summarized in Figure 4.1.

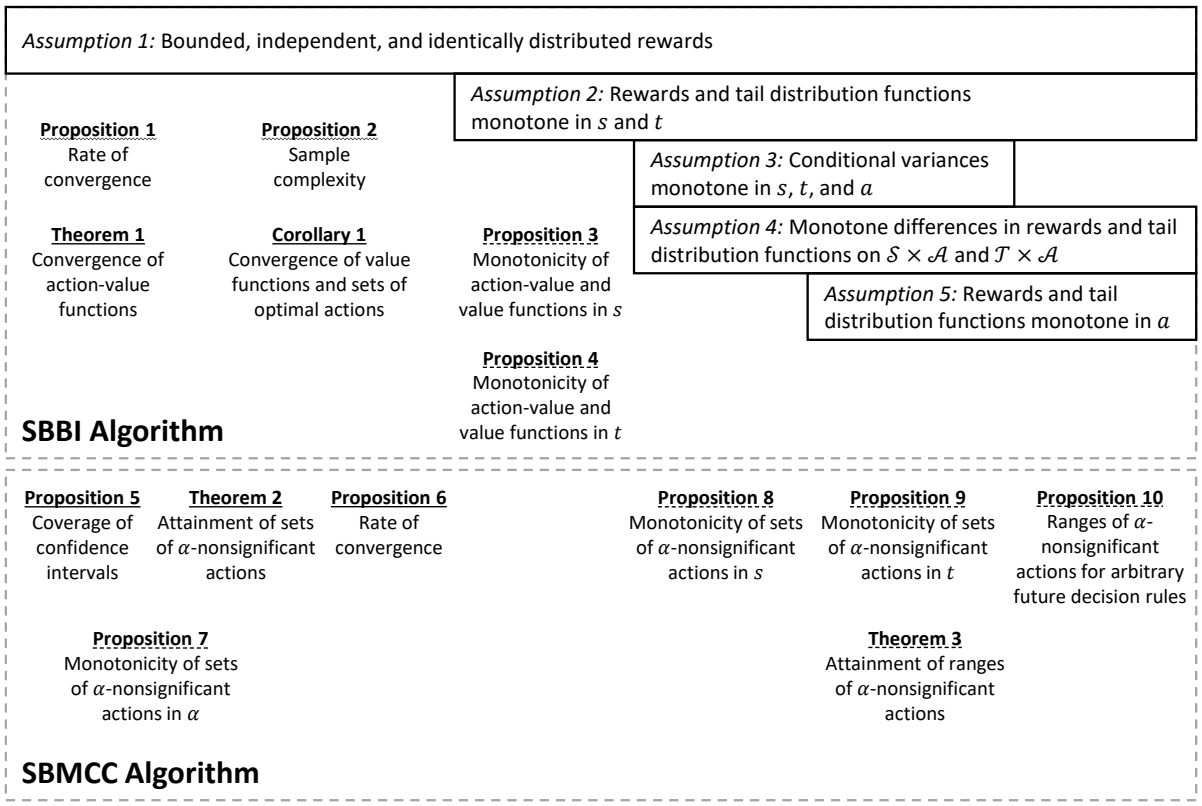


Figure 4.1: Summary of solution approach analysis. Font in italics gives the assumption number each result relies on followed by a short description of the assumption. Boldface font gives the result number in the chapter followed by a brief description of the result. A result uses an assumption if it is placed below the box of the assumption. We divide our theoretical results into three categories: finite sample properties (underlined with a wavy line), general asymptotic properties (underlined with a solid line), and asymptotic structural properties (underlined with a dashed line).

4.6.1 Simulation-Based Backwards Induction

We now introduce our algorithm to estimate the action-value functions and identify optimal actions. The algorithm is presented for finding actions $a^* \in \mathcal{A}_t^*(s)$ such that $Q_t(s, a^*) \geq Q_t(s, a)$ for all $a \in \mathcal{A}$ and each $s \in \mathcal{S}$. If the goal is to find actions a^* such that $Q_t(s, a^*) \leq Q_t(s, a)$ for all a, replace the max and argmax operators by min and argmin, respectively. We aim to estimate $Q_t(s, a)$ for all actions $a \in \mathcal{A}$ and states $s \in \mathcal{S}$ as well as $v_t(s)$ and $\mathcal{A}_t^*(s)$ for every state s.

Our SBBI algorithm is included as Algorithm 1 in Section C.1 of Appendix C. Given a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$, we define a discrete-time stochastic process $(\omega^n : n \in \{1, \ldots, N\})$ as the exogenous information process in the sequential decision problem. Without loss of generality we let $(\omega^n : n \in \{1, \ldots, N\})$ be a sequence of independent and identically distributed (iid) random variables uniformly distributed on $[0, 1]$, denoted by $\mathcal{U}(0, 1)$. This is consistent with past work on simulating MDP models (Chang et al., 2013; Haskell et al., 2016).

At each decision epoch t, state s, and action a, we simulate a sequence $(Q_t^n(s, a) : n \in \{1, \ldots, N\})$ of $N \in \mathbb{N}_+ := \mathbb{N} \setminus \{0\}$ observations of the value associated with state s, action a. Once N observations of each action-value function have been simulated, we approximate the action-value function $Q_t(s, a)$ by its sample mean as:

$$\hat{Q}_t(s, a) := \frac{1}{N} \sum_{n=1}^{N} Q_t^n(s, a) = \frac{1}{N} \sum_{n=1}^{N} r_t(s, a, \omega^n) + \gamma \hat{v}_{t+1}(f_{t+1}(s, a, \omega^n)). \tag{4.3}$$

From $\hat{Q}_t(s, a)$, we estimate the value function $v_t(s)$ as $\hat{v}_t(s) := \max_{a \in \mathcal{A}} \hat{Q}_t(s, a)$, the set of optimal actions $\mathcal{A}_t^*(s)$ as $\hat{\mathcal{A}}_t^*(s) := \operatorname{argmax}_{a \in \mathcal{A}} \hat{Q}_t(s, a)$, the set of sub-optimal actions $\mathcal{A}_t^-(s)$ as $\hat{\mathcal{A}}_t^-(s) := \mathcal{A} \setminus \hat{\mathcal{A}}_t^*(s)$, and the optimal decision rules as $\hat{\pi}_t^*(s) \in \hat{\mathcal{A}}_t^*(s)$. If the terminal conditions $Q_T(s, a) = v_T(s) = \mathbb{E}[r_T(s, \omega)|s]$ are not known in advanced but can be simulated, we also estimate them through their sample mean.

Given ω^n, $f_{t+1}(s, a, \omega^n)$ and $r_t(s, a, \omega^n)$ are deterministic functions of s and a. That is, the distribution of $Q_t^n(s, a)$ is completely determined by ω^n. Because of the Markov property, $Q_t^n(s, a)$ is independent from $Q_t^{n'}(s, a)$ for any $n \neq n'$. Therefore, $(Q_t^n(s, a) : n \in \{1, \ldots, N\})$ is a sequence of iid random variables. This allows us to simulate $(Q_t^n(s, a) : n \in \{1, \ldots, N\})$ in parallel, which leads to great computational speed gains. We use $\mathbb{F}_t(\cdot, s, a)$ to denote the cumulative distribution function (cdf) of $Q_t^n(s, a)$ and $\mathbb{F}_t(\cdot, s)$ to denote the joint cdf of $\{Q_t^n(s, a) : a \in \mathcal{A}\}$. The empirical estimates of the distribution functions are denoted by $\hat{\mathbb{F}}_t(\cdot, s, a)$ and $\hat{\mathbb{F}}_t(\cdot, s)$, respectively.

4.6.1.1 Finite Sample Properties of the SBBI Algorithm

In this subsection, we provide results on the behavior of the SBBI algorithm with finite number of observations. The proofs of the claims in this subsection can be found in Section C.2.1 of Appendix C. We begin with the following assumption:

Assumption 4.1 *The immediate rewards $r_t(s, a, \omega)$ are known constants or non-negative iid random variables from a possibly unknown probability distribution bounded by $R_t(s, a) < \infty$. Further, the terminal rewards $r_T(s, \omega)$ are non-negative iid random variables from a possibly unknown probability distribution bounded by $R_T(s) < \infty$.*

This assumption implies that the action-value functions $Q_t(s, a)$ are bounded and that their estimates $\hat{Q}_t(s, a)$ are bounded random variables. We now state a result on the convergence rate of the SBBI algorithm.

Proposition 4.1 *Under Assumption 4.1 it follows that:*

$$1 - \mathbb{P}(\hat{\mathcal{A}}_t^*(s) \subseteq \mathcal{A}_t^*(s)) \leq A \exp\left\{\frac{-N}{2\kappa_t^2}\right\},$$

with $\kappa_t \coloneqq \sum_{\tau=t}^{T} \gamma^{\tau-t} R_\tau$, where $R_t \coloneqq \max_{(s,a)\in\mathcal{S}\times\mathcal{A}} R_t(s, a)$.

The result in Proposition 4.1 implies that the SBBI algorithm converges exponentially fast on the number of observations N. This finding implies that the SBBI algorithm can efficiently estimate $Q_t(s, a)$, provided that κ_t is not too small compared to N. A consequence of Proposition 4.1 is the required sample size to guarantee that any action in the set of estimates of optimal actions $\hat{\mathcal{A}}_t^*(s)$ is an optimal action with probability of at least $1 - \beta$.

Proposition 4.2 *Suppose Assumption 4.1 holds. Then, for any $\beta \in (0, 1)$ and a fixed sample size N satisfying $N \geq 2\kappa_t^2 \log(A/\beta)$ it holds that $\mathbb{P}(\hat{\mathcal{A}}_t^*(s) \subseteq \mathcal{A}_t^*(s)) \geq 1 - \beta$.*

In the context of medical decision-making problems, such as our case study, $\mathcal{A}_t^*(s)$ is usually a singleton. In this case, Proposition 4.2 gives the number of observations required to ensure that $\mathbb{P}(\hat{\mathcal{A}}_t^*(s) = a^*) \geq 1 - \beta$ for $\mathcal{A}_t^*(s) = \{a^*\}$. A key feature of this proposition is that the number of observations N depends logarithmically on the size of the action space A. This suggests that the sample complexity for finding an action $a^* \in \hat{\mathcal{A}}_t^*(s)$ such that $a^* \in \mathcal{A}_t^*(s)$ with probability of at least $1 - \beta$ increases polynomially even if A increases exponentially with N.

4.6.1.2 Analysis of the SBBI Algorithm

We now present our results on the convergence of the SBBI algorithm. The proofs of the claims in this subsection can be found in Section C.2.2 of Appendix C. We begin by showing the uniform almost sure convergence of $\hat{Q}_t(s,a)$ to $Q_t(s,a)$:

Theorem 4.1 *Suppose Assumption 4.1 holds. Then, $\hat{Q}_t(s,a)$ converges to $Q_t(s,a)$ with probability 1 uniformly on $\mathcal{A}$.*

Uniform convergence implies that the required number of observations for convergence is independent of the action a. The following corollary is an immediate consequence of Theorem 4.1:

Corollary 4.1 *Suppose Assumption 4.1 holds. Then, $\hat{v}_t(s)$ converges to $v_t(s)$ and $\hat{\mathcal{A}}_t^*(s) \subseteq \mathcal{A}_t^*(s)$ with probability 1 for N large enough.*

We use similar arguments as in Proposition 2.1 of Kleywegt et al. (2002) to prove Corollary 4.1 and use their definition for the statement "an event happens with probability 1 for N large enough". An event happens with probability 1 for N large enough if for $\mathbb{P}$-almost every realization of a random sequence $\boldsymbol{\omega} := \{\omega^1, \omega^2, \ldots\}$ for $\omega^1, \omega^2, \ldots \in \Omega$ there exists an integer $N(\boldsymbol{\omega})$ such that the event happens for all samples $\{\omega^1, \ldots, \omega^N\} \in \boldsymbol{\omega}$ with $N \geq N(\boldsymbol{\omega})$.

Remark 4.1 *To prove the claims in this subsection it suffices that $(Q_t^n(s,a) : n \in \{1, \ldots, N\})$ is a sequence of independent random variables. The assumption of identically distributed random variables is not needed.*

4.6.1.3 Asymptotic Structural Properties of the SBBI Algorithm

We proceed to present the asymptotic structural properties of the SBBI algorithm. The proofs of the claims in this subsection can be found in Section C.2.3 of Appendix C. Structured policies tend to be more intuitive for DMs and are typically easier to implement in practice. Throughout this subsection, we discuss the implications of the results using a medical decision-making example. We begin with the following definition:

Definition 4.3 *Let X be a partially ordered set and $g : X \mapsto \mathbb{R}$. We say that g is an ϵ-nonincreasing (ϵ-nondecreasing) function if for $x^+ \geq x^-$ in X it holds that $g(x^+) \leq g(x^-) + \epsilon$ $\Big(g(x^+) \geq g(x^-) - \epsilon\Big)$ for $\epsilon > 0$. In a similar way, we say that g is ϵ-constant if $|g(x^+) - g(x^-)| \leq \epsilon$ for $\epsilon > 0$.*

Definition 4.3 will be useful to deal with ties in the rest of our analysis. For example, if $(X_n : n \in \mathbb{N})$ and $(Y_n : n \in \mathbb{N})$ are sequences of random variables that converge to the same quantity, then for any $\epsilon > 0$ we can find an $N^* \in \mathbb{N}$ such that for every $n \geq N^*$ we have that $|X_n - Y_n| \leq \epsilon$. That is, X_n and Y_n are ϵ-constant. Let $\bar{p}_t(s'|s,a) \coloneqq \sum_{s'' \geq s'} p_t(s''|s,a)$ denote the tail distribution of the transition probabilities, or the probability that the state at decision epoch $t+1$ exceeds s' after choosing action a at state s and decision epoch t. We make the following assumption:

Assumption 4.2 *The state space $\mathcal{S}$ can be ordered such that the tail distribution functions $\bar{p}_t(s'|s,a)$ are nondecreasing in s and t, the expected immediate rewards $\mathbb{E}[r_t(s,a,\omega)|s,a]$ are nonincreasing in s and t, and the terminal rewards are nonincreasing in s and $\mathbb{E}[r_T(s,\omega)|s] \geq \mathbb{E}[r_{T-1}(s,a,\omega)|s,a]$.*

The conditions in Assumption 4.2 with respect to s are the same as in Proposition 4.7.3 in Puterman (2014). This assumption, along with Assumption 4.1, provide sufficient conditions to ensure that the estimates of the action-value functions and the value functions are monotone in s and t for N large enough. We present the monotonicity in s in the following proposition:

Proposition 4.3 *Under assumptions 4.1 and 4.2, $\hat{Q}_t(s,a)$ and $\hat{v}_t(s)$ are ϵ-nonincreasing in s with probability 1 for N large enough.*

In medical decision-making settings, states commonly represent the health condition of patients and actions represent clinical interventions. The action-value and value functions typically represent a measure of how long and how well patients are expected to live given a clinical intervention, such as life-years or quality-adjusted life-years. If the health conditions are ordered from the healthiest to the sickest, Proposition 4.3 implies that sicker patients will have shorter total expected lifetime than healthier patients. The next proposition present the monotonicity of the estimates of the action-value and value functions in t for N large enough.

Proposition 4.4 *Suppose assumptions 4.1 and 4.2 hold. Then, $\hat{Q}_t(s,a)$ and $\hat{v}_t(s)$ are ϵ-nonincreasing in t with probability 1 for N large enough.*

The implication of Proposition 4.4 in medical decision-making problems is that patients' total expected lifetime will never increase with their age, a common assumption in the medical literature.

4.6.2 Simulation-Based Parallel Multiple Comparisons with a Control

In this section, we present our method to identify sub-optimal actions that are not statistically different from an optimal action at a significance level α. To derive a set of near-optimal actions, we compare the performance of an optimal action with the rest of the actions. Similar to Westfall and Young (1993), our formulation aims to find a constant $d_t(s, \alpha)$ for each state s such that it satisfies equation (4.2). Since $Q_t(s, a)$ is unknown, so are $\max_{a \in \mathcal{A}} \psi_t(s, a)$ and its cdf. We denote this cdf as $\mathbb{H}_t$ and use $\mathbb{H}_t(\cdot, \mathbb{F}_t(s))$ when its dependence on the unknown cdf $\mathbb{F}_t(\cdot, s)$ must be emphasized. To address this challenge, we adapt the concept of nonoverlapping batch means to simulated MDP models (Fishman, 1978).

Consider N observations of a simulated MDP. The method of nonoverlapping batch means divides the sequence of N outputs of a simulation into M adjacent nonoverlapping batches, each of size K. Because $(Q_t^n(s, a) : n \in \{1, \ldots, N\})$ is a sequence of iid random variables, dividing N outputs of a simulated MDP into M batches is equivalent to executing M independent simulations of the MDP, each with K observations. The m^{th} batch (or simulation replicate) consists of the random variables: $Q_t^{m,1}(s, a), Q_t^{m,2}(s, a), \ldots, Q_t^{m,K}(s, a)$, for $m = 1, \ldots, M$. For each batch m, we then estimate the action-value functions by the sample mean and the variance associated with the action-value functions over K observations. After batching, the grand sample mean can be obtained as:

$$\hat{Q}_t(s, a) = \frac{1}{M} \sum_{m=1}^{M} \bar{Q}_t^m(s, a) = \frac{1}{MK} \sum_{m=1}^{M} \sum_{k=1}^{K} Q_t^{m,k}(s, a),$$

and the variance of the batch sample means as:

$$\hat{\zeta}_t^2(s, a) = \frac{1}{M - 1} \sum_{m=1}^{M} \left(\bar{Q}_t^m(s, a) - \hat{Q}_t(s, a) \right)^2,$$

where $\bar{Q}_t^m(s, a)$ is the sample mean for the m^{th} batch. We obtain an estimate of $\sigma_t^2(s, a)$ by multiplying the variance of the batch sample means by the number of observations per batch. That is, $K\hat{\zeta}_t^2(s, a)$ is an estimator of $\sigma_t^2(s, a)$.

The nonoverlapping batch means method then allows us to generate an estimator for the root statistic $\psi_t(s, a)$ as:

$$\hat{\psi}_t(s, a) := \frac{\hat{Q}_t(s, a^*) - \hat{Q}_t(s, a) - (Q_t(s, a^*) - Q_t(s, a))}{\sqrt{M^{-1} \left[\hat{\zeta}_t^2(s, a^*) + \hat{\zeta}_t^2(s, a) \right]}}.$$

Note that in the nonoverlapping batch means method formulation we use the number of batches M instead of the total number of observations N to calculate standard errors because we are estimating $\sigma_t^2(s,a)$ with $K\hat{\zeta}_t^2(s,a)$. Using the variability across the M batches, we generate an empirical estimate of $\mathbb{H}_t$, denoted by $\hat{\mathbb{H}}_t(\cdot, \hat{\mathbb{F}}_t(s))$ or simply $\hat{\mathbb{H}}_t$ when is not necessary to highlight its dependence to the empirical cdf $\hat{\mathbb{F}}_t(\cdot, s)$.

We now introduce our algorithm to generate $\hat{\mathbb{H}}$ and estimate the quantile $d_t(s,\alpha)$ for each state s at decision epoch t. Our SBMCC method is included as Algorithm 2 in Section C.1 of Appendix C. For each batch (or simulation replicate), we generate an estimate of the root statistic as:

$$\bar{\psi}_t^m(s,a) \coloneqq \frac{\bar{Q}_t^m(s,a^*) - \bar{Q}_t^m(s,a) - \left(\hat{Q}_t(s,a^*) - \hat{Q}_t(s,a)\right)}{\sqrt{K^{-1}\left[\bar{\sigma}_t^2(s,a^*,m) + \bar{\sigma}_t^2(s,a,m)\right]}},$$

for $m = 1,\ldots,M$, where $a^* \in C_t(s)$, $\bar{Q}_t^m(s,a)$ is the sample mean for the m^{th} batch and $\bar{\sigma}_t^2(s,a,m)$ is the sample variance of the the m^{th} batch.

A key assumption in MCC is that the control is known before observing the data that will be used to evaluate the actions. Thus, we must generate $C_t(s) \in \mathcal{C}$ without knowing $\{\hat{Q}_t(s,a) : t \in \mathcal{T} \setminus \{T\}, s \in \mathcal{S}, a \in \mathcal{A}\}$. Several approaches could be used to identify $\mathcal{C}$, such as solving the standard version of the MDP $(\mathcal{T}, \mathcal{S}, \mathcal{A}, \boldsymbol{P}, \rho, \gamma)$ through backwards induction before simulating the MDP and letting $C_t(s) = \mathcal{A}_t^*(s)$, simulating an initial independent replication of the SBBI algorithm and letting $C_t(s) = \hat{\mathcal{A}}_t^*(s)$, or using the estimates of a single batch and letting $C_t(s) = \operatorname{argmax}_{a\in\mathcal{A}} \bar{Q}_t^m(s,a)$. In the later case, the batch used to obtain $C_t(s)$ must be excluded from Algorithm 2. Proposition 4.2 provides a lower bound on the sample size required such that $C_t(s) \subseteq \mathcal{A}_t^*(s)$ with high probability when $C_t(s)$ is obtained from a simulation model.

Once we generate an estimate of the root statistic for each batch, we estimate $d_t(s,\alpha)$ as:

$$\hat{d}_t(s,\alpha) \coloneqq \inf\left\{x \in \mathbb{R} : \hat{\mathbb{H}}_t(x, \hat{\mathbb{F}}_t(s)) \geq 1-\alpha\right\}.$$

Since the estimates of the root statistics $\{\bar{\psi}_t^m(s,a) : m \in \{1,\ldots,M\}, t \in \mathcal{T} \setminus \{T\}, s \in \mathcal{S}, a \in \mathcal{A}\}$ are mutually independent, the SBMCC algorithm can be executed in parallel across M, $\mathcal{S}$, $\mathcal{A}$, and $\mathcal{T} \setminus \{T\}$. However, we present the parallel execution of the algorithm across M, $\mathcal{S}$, and $\mathcal{A}$ as it allows for the integration of the SBMCC algorithm to the SBBI algorithm. The combination of the algorithms is included as Algorithm 3 in Section C.1 of Appendix C. This integration allows us to investigate the effect of future α-nonsignificant actions in the sets of near-optimal actions in the current decision epoch.

4.6.2.1 Analysis of the SBMCC Algorithm

We now proceed to present our asymptotic results of the SBMCC algorithm. The proofs of the claims in this subsection can be found in Section C.2.4 of Appendix C. Let $\Theta \subseteq \mathbb{R}$ denote the set of all possible values of $Q_t(s,a^*) - Q_t(s,a)$. In the following proposition, we show that the SBMCC algorithm produces the correct overall asymptotic coverage $1-\alpha$.

Proposition 4.5 *Suppose Assumption 4.1 holds. Then, for any $\alpha \in (0,1)$ we have that:*

$$\mathbb{P}\Big(Q_t(s,a^*) - Q_t(s,a) \in \Theta : \hat{\mathbb{H}}_t\left(\max_{a\in\mathcal{A}}\{\hat{\psi}_t(s,a)\}, \hat{\mathbb{F}}_t(s)\right) \leq 1-\alpha\Big) = 1-\alpha$$

for N large enough, $a^* \in C_t(s)$, and all $a \in \mathcal{A}$.

The result in Proposition 4.5 means that the true difference between the performance of a control action and the remaining actions at state s will asymptotically be in a subset of Θ such that the empirical cdf of $\max_{a\in\mathcal{A}} \psi_t(s,a)$ evaluated at $\max_{a\in\mathcal{A}} \hat{\psi}_t(s,a)$ is at most the confidence level $1-\alpha$ with probability of exactly such confidence level. While all the conditions in Θ involve random variables, all the relevant quantities converge with probability 1 to their true values as $M \to \infty$ and $K \to \infty$ (see lemmas C.5 and C.6 in Appendix C). Note that our method has similar asymptotic coverage properties to the nonparametric bootstrap method (Beran, 1988; Tu and Zhou, 2000; Westfall, 2011). Proposition 4.5 allows us to show that a set of actions $\Pi_t(s,\alpha) \subseteq \mathcal{A}_t(s)$ with a quantile $\hat{d}_t(s,\alpha)$ derived from the SBMCC algorithm will asymptotically be a set of α-nonsignificant actions with probability 1. We present this result in the following theorem:

Theorem 4.2 *Under Assumption 4.1 we have that:*

$$\Pi_t(s,\alpha) = \left\{a \in \mathcal{A} : \hat{Q}_t(s,a^*) - \hat{Q}_t(s,a) \leq \hat{d}_t(s,\alpha)\sqrt{M^{-1}\left[\hat{\zeta}_t^2(s,a^*) + \hat{\zeta}_t^2(s,a)\right]}\right\}$$

with $\hat{d}_t(s,\alpha) = \hat{\mathbb{H}}_t^{-1}(1-\alpha, \hat{\mathbb{F}}_t(s))$ is a set of α-nonsignificant actions with probability 1 for N large enough and $a^* \in C_t(s)$.

Theorem 4.2 generalizes the theoretical basis of MCC as described in Section 3 of Dunnett (1955) to the nonparametric case. This theorem also extends the simultaneous confidence interval methods for MCC without the equal variances assumption described in Section 2 of Li and Ning (2012). It is worth observing that Theorem 4.2 as well as the results in Dunnett (1955) and Li and Ning (2012) are based on the implicit null hypothesis that all actions are equally good. We now state a result on the rate of convergence of the SBMCC algorithm.

Proposition 4.6 *Suppose that Assumption 4.1 is satisfied. Then,*

$$\lim_{N\to\infty} N^{1/2} \sup_{x\in\mathbb{R}} \left|\hat{\mathbb{H}}_t(x, \hat{\mathbb{F}}_t(s)) - \mathbb{H}_t(x, \mathbb{F}_t(s))\right| \leq CA^{5/4}\sqrt{2\kappa_t^3},$$

where C is the constant appearing in the multivariate Berry-Esseen bound.

Although the value of the constant C is an area of open research, its current best estimate is $C = 42A^{1/4} + 16$ by Raic (2019). The result in Proposition 4.6 provides a bound on the convergence rate of Algorithm 2 of order $O(N^{-1/2})$. We note that this rate of convergence is equivalent to the convergence rate of the central limit theorem (Serfling, 1980, Theorem 1.9.5). In the proof of this proposition, we once again observe similarities between our method and the nonparametric bootstrap (Singh, 1981, Theorem 1.C).

4.6.2.2 Asymptotic Structural Properties of the SBMCC Algorithm

We provide asymptotic structural results of the SBMCC algorithm next. The proofs of the claims in this subsection can be found in Section C.2.5 of Appendix C. Similar to subsection 4.6.1.3, we discuss the connotations of the assumptions and results in the context of medical decision-making scenarios. We start by stating the relationship between the cardinality of the sets of α-nonsignificant actions and the significance level α.

Proposition 4.7 *Suppose that Assumption 4.1 holds. Then, $|\Pi_t(s,\alpha)|$ is nonincreasing in $\alpha \in (0,1)$. Moreover, there exist an α such that $\Pi_t(s,\alpha) \subseteq \mathcal{A}_t^*(s)$ with probability 1 for N large enough.*

Under the classical null hypothesis that all actions are equally good, the significance level α indicates the strength of the evidence that a DM, such as a clinician and/or a patient, requires before concluding that there is sufficient evidence to reject the null hypothesis. The result in Proposition 4.7 means that if we increase the significance level (e.g., from 0.01 to 0.05), the sets of α-nonsignificant actions will not include more choices. This result allows clinicians to control the cardinality of the sets of α-nonsignificant actions based on their confidence in the rewards and transition functions. If clinicians are not exceptionally certain in the parameterization of their model, smaller values of α, such as 0.01 and 0.05, may be reasonable. On the other hand, if clinicians are highly confident in their model, larger values of α, such as 0.1 and 0.2 may suffice.

We now present sufficient conditions to ensure that the cardinality of the sets of α-nonsignificant actions are monotone for N large enough. Consistent with the work of Mannor and Tsitsiklis (2013), we find that there are no recursive algorithms when dealing with the variance of value functions. As a result, we make assumptions directly on the

conditional variance of the value functions. The following assumption gives conditions on the conditional variance of the expected rewards and value functions:

Assumption 4.3 *The conditional variance of the immediate rewards* $\mathbb{E}[r_t^2(s,a,\omega)|s,a] - \mathbb{E}[r_t(s,a,\omega)|s,a]^2$ *and the value functions* $\mathbb{E}[v_t^2(s')|s,a] - \mathbb{E}[v_t(s')|s,a]^2$ *are nonincreasing in* s, t, *and* $a \in \mathcal{A}_t(s)$.

Recall our medical decision-making setting from Subsection 4.6.1.3 where the time periods and states represent patients' age and health conditions, respectively, and the actions are clinical interventions at different intensities. Health conditions are ordered from the healthiest to the sickest and clinical interventions are order from the lowest to the highest intensity. The action-value and value functions represent how long and how well patients are expected to live given a clinical intervention. In this setting, Assumption 4.3 indicates that the effect of clinical interventions will not become more uncertain when patients are sicker or older. It also suggests that more intense clinical interventions will be more certain than less aggressive interventions. Although $\mathbb{E}[v_t^2(s')|s,a] - \mathbb{E}[v_t(s')|s,a]^2$ is not part of the basic model, the conditions can be directly verified after obtaining $v_t(s)$ either via standards backwards induction or after approximating it through $\hat{v}_t(s)$ with the SBBI algorithm.

Besides requirements on the conditional variances, we make an assumption on the nature of the differences of the tail distribution functions and the expected immediate rewards:

Assumption 4.4 *The tail distribution functions* $\bar{p}_t(s'|s,a)$ *are subadditive functions on* $\mathcal{S} \times \mathcal{A}$ *and* $\mathcal{T} \times \mathcal{A}$. *Further, the expected immediate rewards* $\mathbb{E}[r_t(s,a,\omega)|s,a]$ *are superadditive functions on* $\mathcal{S} \times \mathcal{A}$ *and* $\mathcal{T} \times \mathcal{A}$.

Assumption 4.4 means that the impact of more intense clinical interventions on the probability that patients get sicker is larger if the patient is sicker or older. This assumption also implies that more aggressive interventions have a larger effect on patients' health when the patients are sicker or older.

Incorporating Assumptions 4.1 through 4.4 provide sufficient conditions to make sure that the sets of α-nonsignificant actions are monotone in s for N large enough. We present this result in the following proposition:

Proposition 4.8 *Suppose Assumptions 4.1 through 4.4 hold. Then,* $|\Pi_t(s,\alpha)|$ *is* ϵ*-nonincreasing in* s *with probability 1 for* N *large enough.*

The result in Proposition 4.8 means that sicker patients will receive less choices than healthier patients. This may be useful from a clinical perspective because sicker patients are more likely to experience adverse events. To ensure that the sets of α-nonsignificant actions are monotone in t we also need conditions on the action space $\mathcal{A}_t(s) \subseteq \mathcal{A}$ associated with state s and decision epoch t. We make the following assumption:

Assumption 4.5 *The action space $\mathcal{A}_t(s)$ can be ordered such that the tail distribution functions $\bar{p}_t(s'|s,a)$ are nonincreasing in $a \in \mathcal{A}_t(s)$ and the expected immediate rewards $\mathbb{E}[r_t(s,a,\omega)|s,a]$ are nondecreasing in $a \in \mathcal{A}_t(s)$.*

If clinical interventions are ordered from the lowest to the highest intensity, Assumption 4.5 implies that the more aggressive interventions will result in better immediate and future expected health outcomes. Integrating Assumption 4.5 to assumptions 4.1 through 4.4 provides sufficient conditions to assure that the sets of α-nonsignificant actions will be monotone in t for N large enough.

Proposition 4.9 *Suppose Assumptions 4.1 through 4.5 hold. Furthermore, assume that $v_t(s) - v_{t+1}(s)$ is nondecreasing in s. Then, $|\Pi_t(s,\alpha)|$ is ϵ-nonincreasing in t with probability 1 for N large enough.*

The additional condition in Proposition 4.9 indicates that healthier patients will experience smaller differences in terms of life expectancy and quality of life than sicker patients over the planning horizon. Proposition 4.9 means that patients will receive less choices in their sets of α-nonsignificant actions as they get older. This result may be useful in clinical practice because older patients are typically more likely to experience adverse events. We highlight that the conditions in Propositions 4.8 and 4.9 are also sufficient to show monotonicity in the approximately optimal decision rules $\hat{\pi}_t^*(s)$ for N large enough in the following remark:

Remark 4.2 *The conditions in Proposition 4.8 and Proposition 4.9 are sufficient to prove that there exist approximately optimal decision rules $\hat{\pi}_t^*(s)$ that are ϵ-monotone on s and t with probability 1 for N large enough, respectively. We provide the proof of this remark in Section C.2.5 of Appendix C.*

Combining Assumptions 4.1, 4.2, 4.3, and 4.5 we get sufficient conditions to guarantee that the actions contained in the sets of α-nonsignificant actions are arranged as a range (see Definition 4.2). We present this result in the following theorem:

Theorem 4.3 *Suppose that assumptions 4.1, 4.2, 4.3, and 4.5 are satisfied. Then,* $\Pi_t(s, \alpha)$ *is an* α*-nonsignificant range of actions at state* s *and decision epoch* t *with probability 1 for* N *large enough.*

If there are two clinical interventions of varying intensities contained in a set of α-nonsignificant actions, the result in Theorem 4.3 implies that any clinical intervention with an intensity between them will also be included in the set. Moreover, once a clinical intervention is proven to not be part of $\Pi_t(s, \alpha)$ it is certain that any intervention that is more extreme will also not be part of the range. This results in computational gains, especially for the case of large action spaces. A range of near-optimal actions may be more intuitive and interpretable for clinicians than a set without any particular order. As a result, the ranges of α-nonsignificant actions may be easier to implement in medical practice. We make the following remark to account for the case that all the action-value functions have the same variance:

Remark 4.3 *If* $\{Q_t(s, a) : a \in \mathcal{A}\}$ *for state* s *and decision epoch* t *have equal variances (i.e.* $\sigma_t^2(s, a) = \sigma_t^2(s, a')$ *for all* $a \neq a'$*), Assumption 4.3 is not required to show that* $\Pi_t(s, \alpha)$ *is a range of actions at state* s *with probability 1 for* N *large enough.*

A consequence of the conditions in this subsection, in particular Assumption 4.5, is that clinicians do not need to assume that patients will receive approximately optimal interventions in the next decision epoch. If these conditions are satisfied, clinicians could use any decision rule in the next decision epoch and asymptotically reach at least a subset of the recommendations in the current period. Let $\mathscr{A}_t^*(s, \tilde{a})$ and $\mathit{\Pi}_t(s, \alpha, \tilde{a})$ denote the set of optimal actions and the set of α-nonsignificant actions, respectively, assuming that action $\tilde{a} \in \mathcal{A}_{t+1}(f_{t+1}(s, a, \omega))$ is taken at the next decision period. We present our result in the following proposition:

Proposition 4.10 *Suppose assumptions 4.1, 4.2, 4.4 and 4.5 hold. Then, we have that* $\mathscr{A}_t^*(s, \tilde{a}) \subseteq \mathcal{A}_t^*(s)$ *and* $\mathit{\Pi}_t(s, \alpha, \tilde{a}) \subseteq \Pi_t(s, \alpha)$ *for* N *large enough.*

The result in Proposition 4.10 indicates that if assumptions 4.1, 4.2, 4.4, and 4.5 are satisfied then any decision rule can be followed in the subsequent decision epochs. This result may be beneficial in clinical practice as future clinical interventions may be unclear due to the uncertainty in patients' health progression. This result provides clinicians and their patients with confidence in the ranges of near-optimal actions in the current decision epoch without the burden of potential ambiguity in patients' future health.

4.7 Case study: Personalized Hypertension Treatment Plans

In this section, we apply of our methodology to obtain flexible hypertension treatment plans for primary prevention of ASCVD. We begin by providing some background on hypertension treatment as well as motivating the need for flexible treatment protocols. Subsequently, we describe our MDP, data source, model parameters, and simulation framework. Lastly, we present the treatment plans and health outcomes of patients following treatment choices contained in our ranges of near-optimal actions.

4.7.1 Background on Hypertension Treatment

Using the definition from the 2017 Hypertension Clinical Practice Guidelines, 45.6% of adults in the US have hypertension (Whelton et al., 2018). Stage 1 hypertension is defined as SBP of 130-139 mm Hg or diastolic blood pressure (DBP) of 80-89 mm Hg, and stage 2 hypertension is defined as an SBP of at least 140 mm Hg or a DBP of at least 90 mm Hg. The 2017 Hypertension Clinical Practice Guidelines provide non-pharmacological and pharmacological recommendations for patients with hypertension as well as with elevated BP, defined as an SBP of 120-129 mm Hg and a DBP smaller than 80 mm Hg. In this case study, we focus on the pharmacological recommendations.

A key distinction between clinical practice guidelines, such as Chobanian et al. (2003); James et al. (2014); Williams et al. (2018) and Whelton et al. (2018), and optimal decision models in the literature is that they provide clinicians with flexibility in the implementation of hypertension treatment plans. To benefit from clinicians' judgment and account for their patients' preferences, we develop ranges of near-optimal treatment choices for the personalized management of hypertension.

4.7.2 Markov Decision Process Formulation

The process of sequentially determining antihypertensive medications over a planning horizon is modeled as a finite MDP. We adapt the standard MDP formulation in Schell et al. (2016) to a primary prevention simulation MDP. The objective of the MDP model is to determine the treatment strategy that maximizes the expected discounted life-years before an adverse event. The elements of our simulation MDP $(\mathcal{T}, \mathcal{S}, \mathcal{A}, f, r, \gamma)$ are as follows:

$\mathcal{T}$	planning horizon of $T = 10$ years where decisions are made at the beginning of each year; $\mathcal{T} = \{0, 1, \ldots, 10\}$.
$\mathcal{S}$	state space consisting patients' demographic information, clinical observations, and health condition. We separate the state space $\mathcal{S}$ into healthy states $\mathcal{H}$ and absorbing states $\mathcal{E}$, based on patients' health conditions (i.e. $\mathcal{S} = \mathcal{H} \cup \mathcal{E}$).
$\mathcal{A}$	action space composed of 0 to 5 antihypertensive medications at half and standard dosage, for a total of $A = 21$ treatment choices.
$s' = f_{t+1}(s, a, \omega)$	transition function based on the estimated transition probability $p_t(s'\|s, a)$. The transition probabilities are derived from patients' risk for ASCVD events, the benefit from treatment, fatality likelihoods, and non-ASCVD mortality.
$r_t(s', a, \omega)$	reward associated with a transition to state s' after action a and outcome ω. We define $r_t(s', a, \omega) = 1 - \Delta(a)$ if $s' \in \mathcal{H}$ and 0 otherwise, where $\Delta(a)$ denotes the treatment-realted disutility from medication a.
γ	discount factor of the model; $\gamma = 0.97$.

We describe each of the elements of the MDP in detail in the following subsections.

4.7.2.1 Planning Horizon

Based on communications with clinical collaborators, we evaluate the health outcomes and treatment plans of each patient in our populations over 10 years. That is, $\mathcal{T} = \{0, 1, \ldots, 10\}$. Treatment decisions are made at the beginning of each year $t \in \mathcal{T} \setminus \{10\}$. We use $T = 10$ to represent the lifetime effects of the treatment decisions on each patient.

4.7.2.2 State Space

A state $s \in \mathcal{S}$ consists of a patient's demographic information, clinical observations, and health condition. The patient's demographic information includes age, race, sex, smoking status, and diabetes status. The clinical observations are measurements of the patient's untreated SBP, DBP, HDL, and TC. We separate the state space $\mathcal{S}$ into healthy states $\mathcal{H}$ and absorbing states $\mathcal{E}$, based on the health condition of the patient (i.e. $\mathcal{S} = \mathcal{H} \cup \mathcal{E}$). A healthy state $s \in \mathcal{H}$ is defined by the patient's demographic and clinical observations prior to an ASCVD event or death. The absorbing states $\mathcal{E}$ represent ASCVD events or death. We consider the following absorbing states: survival of a CHD event, survival of a stroke, death from CHD event, death from stroke, and death from a non-cardiovascular

disease related cause.

4.7.2.3 Action Space

Our action space $\mathcal{A}$ is the set of possible number of medications that can be prescribed. We considered between 0 and 5 antihypertensive medications at half and standard dosage, for a total of $A = 21$ treatment choices. Once a treatment decision is made, we assume that treatment has a near-immediate effect on patients' health. The estimates of the effects of antihypertensive drugs on patients' BP and risk for ASCVD events are derived from the work performed by the Blood Pressure Lowering Treatment Trialists' Collaboration (BPLTTC) as described in Sundström et al. (2014, 2015); Salam et al. (2019). To be consistent with previous studies, we assume that a medication at half dosage has two-thirds of the strength of a medication at standard dosage and half of its side effects (Sussman et al., 2013). Based on communications with our clinical collaborators, we constraint the set of possible actions by a minimum allowable BP. To prevent treatment harm from excessive medication, we used a minimum allowable BP of 120/55 mm Hg.

4.7.2.4 Transition Functions

Before treatment decisions are made, we calculate a pre-treatment, one-year risk of ASCVD events for each patient using the risk calculator described in Yadlowsky et al. (2018). This risk calculator was developed using a logistic regression model with age, sex, race (Black or White), smoking status, diabetes status, SBP, HDL, and TC as explanatory variables. Since CHD and stroke have the same basic risk factors (and because there are currently no separate risk scores for CHD and stroke) we assume that 70% of the ASCVD risk is due to CHD events and 30% is due to stroke events (Virani et al., 2020).

The risk for CHD and stroke are modified when a patient receives treatment. For patients who have an ASCVD event in a given year, we calculate how likely are the patients to die by applying fatality likelihoods to the post-treatment risk of CHD and stroke events (Virani et al., 2020; Kochanek et al., 2019; Sussman et al., 2013). We calculate these fatality likelihoods as the ratio of known fatal event rates from National Center for Health Statistics (NCHS) to the overall event rates in our population predicted by the risk calculator, adjusted for age and sex (Yadlowsky et al., 2018; NCHS, 2017). Lastly, we calculate the probability of non-ASCVD mortality using life-tables (Arias and Xu, 2019).

The post-treatment risk, fatality likelihoods, and non-ASCVD mortality provide all the necessary information to estimate transition probabilities $p_t(s'|s, a)$ for every state s, treatment choice a, and year t. Note that since $\mathcal{E}$ is a set of absorbing states, if $s \in \mathcal{E}$ then

$p_t(s'|s,a) = 1$ for $s' = s$. From the transition probabilities we obtain the following transition functions:

$$s' = f_{t+1}(s,a,\omega) = \begin{cases} 1 & \text{if } \omega \leq p_t(s'=1|s,a), s \in \mathcal{H} \\ j & \text{if } \sum_{k=1}^{j-1} p_t(s'=k|s,a) < \omega \leq \sum_{k=1}^{j} p_t(s'=k|s,a), s \in \mathcal{H} \text{ for } j \in \mathcal{S} \setminus \{1\}, \\ s & \text{if } s \in \mathcal{E}. \end{cases}$$

for all s, a, and t.

4.7.2.5 Rewards

A patient in state s transition to state s' after receiving an antihypertensive medication a. Following the transition to s', the patient receives a reward $r_t(s',a,\omega)$ defined as:

$$r_t(s',a,\omega) = \begin{cases} 1 - \Delta(a) & \text{if } s' \in \mathcal{H}, \\ 0 & \text{otherwise,} \end{cases}$$

where $\Delta(a)$ represents the disutility, or treatment harm, from medication a. The disutility associated with each treatment choice are obtained from Schell et al. (2016) and Sussman et al. (2013). Once the last treatment decision is made at year 9, the patient transitions to state $\bar{s}$ and receives a terminal reward defined as:

$$r_T(\bar{s},\omega) = \begin{cases} \mathcal{L}(\bar{s}) & \text{if } \bar{s} \in \mathcal{H}, \\ 0 & \text{otherwise,} \end{cases}$$

where $\mathcal{L}(s)$ denotes the expected lifetime for a patient in state s. The expected lifetime of each patient is obtained from Arias and Xu (2019). As suggested in Neumann et al. (2016), we use a discount factor of $\gamma = 0.97$ per year.

The clinical parameters used throughout our numerical study are listed in Table 4.3.

4.7.3 Data Source

To parameterize our models, we use the NHANES dataset from 2009 to 2016 (Centers for Disease Control and Prevention, 2020). Our primary sample is composed of adult Black or White patients from 50 to 54 years old with no history of heart attack, stroke, or congestive heart failure, for a total population of 16.72 million people. We impute any missing data in the NHANES dataset using the MissForest package in R (Stekhoven and Buhlmann,

Table 4.3: Base case parameters

Parameter	Value	Source
BP reduction: standard dosage (half dosage)		Sundström et al. (2014, 2015); Salam et al. (2019); Sussman et al. (2013)
SBP	5.5 (3.7) mm Hg	
DBP	3.3 (2.2) mm Hg	
ASCVD risk reduction: standard dosage (half dosage)		Sundström et al. (2014, 2015); Salam et al. (2019); Sussman et al. (2013)
CHD	13% (7%)	
Stroke	21% (14%)	
Treatment-related disutility		Schell et al. (2016); Sussman et al. (2013)
Half dosage	0.001	
Full dosage	0.002	
Life expectancy	Varies by patient	Arias and Xu (2019)
Proportion of ASCVD risk due to CHD	70%	Virani et al. (2020)
Mortality from ASCVD events		NCHS (2017)
CHD	Varies by patient	
Stroke	Varies by patient	
Non-ASCVD Mortality	Varies by patient	Arias and Xu (2019)

2012). To model how the risk factors of each patient may evolve over time, we estimate their progression using linear regression. See Section C.3.1 in Appendix C for details.

4.7.4 Ordering of States and Actions

To obtain a range of near-optimal treatment choices with probability 1 if $\mathcal{S}$ and $\mathcal{A}_t(s)$ must be ordered such that $\mathbb{E}[r_t(s', a, \omega)|s, a]$ and $\bar{p}_t(s'|s, a)$ are monotone on $s \in \mathcal{S}$ and $a \in \mathcal{A}_t(s)$, and $\sigma_t(s, a)$ is monotone on $\mathcal{A}_t(s)$ (see Theorem 4.3).

Given the progression of a patient's risk factors, the state transitions only depend on their health condition. To ensure the monotonicity of $\mathbb{E}[r_t(s', a, \omega)|s, a]$ and $\bar{p}_t(s'|s, a)$ on s, we ordered patients' states at each year in terms of their health condition. As $p_t(s|s, a) = 1$ for $s \in \mathcal{E}$ and $p_t(s'|s, a) \in (0, 1)$ if $s \in \mathcal{H}$, we order our states so that $\hat{s} < \tilde{s}$ if $\hat{s} \in \mathcal{H}$ and $\tilde{s} \in \mathcal{E}$. This ordering guarantees that $\bar{p}_t(s'|s, a)$ is monotone in s. Since patients only receive a nonzero reward if the patient transitions from $s \in \mathcal{H}$ to $s' \in \mathcal{H}$, $\mathbb{E}[r_t(s', a, \omega)|s, a]$ is monotone in s by construction.

To make sure $\mathbb{E}[r_t(s, a, \omega)|s, a]$, $\bar{p}_t(s'|s, a)$, and $\sigma_t(s, a)$ are monotone in a, we ordered $\mathcal{A}_t(s)$ in terms of number of medications. We note that this ordering achieves the desired result because the reduction in ASCVD risk from treatment is linear in the number of medications.

4.7.5 Simulation Framework

We develop a simulation model to evaluate the hypertension treatment plans contained in our ranges of treatment choices. For comparison purposes, we also evaluate optimal treatment plans as described in Schell et al. (2016) and a treatment strategy based on the 2017 Hypertension Clinical Practice Guidelines (Whelton et al., 2018).

The trajectory of a single patient in our modeling framework is summarized in Figure 4.2. Before developing treatment plans, we calculate the risk for ASCVD events at each year. We then estimate transition probabilities and develop transition functions. Subsequently, we determine treatment policies based on: (1) optimal treatment plans, (2) the 2017 Hypertension Clinical Practice Guidelines, and (3) treatment choices contained in our ranges of near-optimal actions. To derive our ranges of near-optimal treatment choices we use the combined version of the SBBI and SBMCC algorithms included as Algorithm 3 in Section C.1 of Appendix C. The control treatment choices are identified by using the estimates of the first batch of outputs from the SBBI algorithm.

We consider three types of treatments contained in the ranges of near-optimal actions: the best performing treatment in the range, the treatment choice with the median number

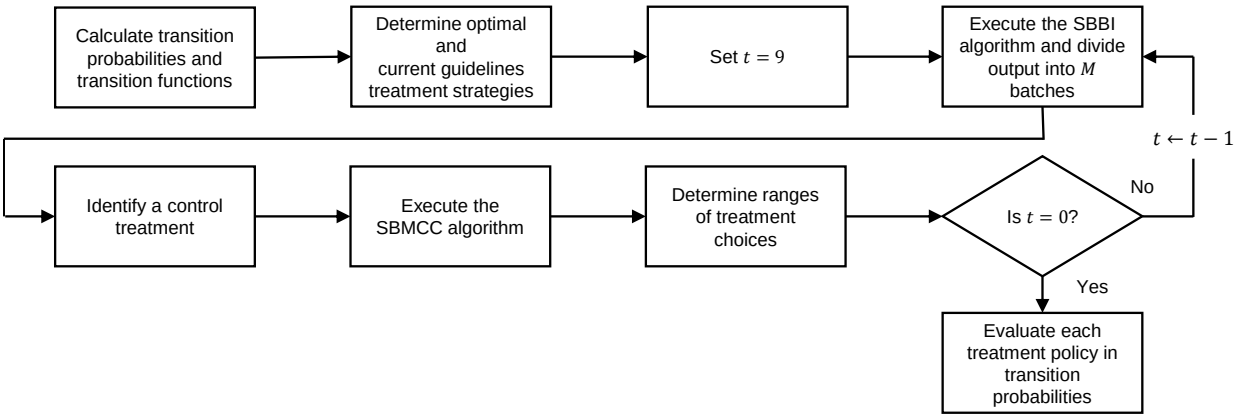


Figure 4.2: Summary of simulation framework for a single patient. The index t represents represents the year in the planning horizon (10 years).

of medications, and the treatment choice with the least amount of medications. These strategies will be referred to as: best in range, median in range, and fewest in range, respectively. We choose the first type of treatment choice to mimic the behavior of the kind of patient that want the best possible treatment and completely adheres to prescriptions. Note that the best in range strategy corresponds to the approximately optimal treatment choice obtained with the SBBI algorithm. The other two types of treatment plans (i.e. median in range and fewest in range) aim to represent potential physicians' reactions to patients nonadherence. Research has shown that prescribing less medications may increase patients adherence to prescriptions (Claxton et al., 2001; Saini et al., 2009). Once we obtain the different treatment strategies, we evaluate each treatment policy based on the Markov chain embedded in the MDP (Puterman, 2014, Section 4.2). The optimal policies and the current clinical guidelines treatment strategy are described in Section C.3.2 of Appendix C.

4.7.6 Selection of Number of Batches and Observations per Batch

Before evaluating the impact of flexible treatment plans, we derive the number of batches to divide the output of the simulated MDP for each patient in our population. We are interested in obtaining simultaneous confidence interval widths that are narrow enough to identify as many treatment choices as possible to be inferior to the control treatment with finite number of batches, finite observations per batch, and specified significance level α. Consequently, we use the maximum confidence interval width across all patients at the first year of our study to evaluate the convergence of the confidence interval widths. A significance level of $\alpha = 0.05$ is used for all analyses. The selection of the number of

batches in described in detail in Section C.3.3 of Appendix C.

4.7.7 Analysis

To understand the implications of flexible treatment plans at a patient level, we examine the effect of the characteristics of a patient on the width of the ranges of treatment choices. We then study the policy implications of the ranges of near-optimal treatment alternatives by comparing our methodology to optimal treatment plans and current clinical guidelines. First, we inspect the distribution of the number of medications recommended by each treatment strategy. Second, we explore the proportion of patients the ranges of near-optimal actions included the optimal treatment and the current clinical guidelines treatment strategy. Third, we examine the life-years saved, ASCVD events averted, and expected time to adverse event (including ASCVD events and non-ASCVD related death) by each treatment strategy, compared to no treatment.

To study the policy implications of each treatment strategy, we divide our population by sex, race, and BP group. We create the BP groups on the basis of the 2017 Hypertension Clinical Practice Guidelines: normal BP, elevated BP, stage 1 hypertension, and stage 2 hypertension. We also perform sensitivity analysis on the treatment strategies by varying the model parameters and assumptions.

4.7.8 Sensitivity Analyses

The sensitivity analyses on model parameters are described in Table 4.4. These parameters and their sensitivity analysis values are selected based on the existing literature and communications with our clinical collaborators (Sussman et al., 2013).

Table 4.4: Sensitivity analysis parameter values.

Parameter	Base case (sensitivity analysis values)
ASCVD risk	Yadlowsky et al. (2018) (half, double)
ASCVD risk reductions	BPLTTC (half, double)
Half dosage disutility	0.001 (0.0005, 0.0020, 0.0092)
Standard dosage disutility	0.002 (0.0010, 0.0040, 0.0184)
Action-value function distribution	Empirical (Gaussian)
Future action	Best in range (fewest in range, median in range)
Population	Ages 50-54 (ages 70-74)
Parameter misestimation	None (± 50% estimated risk, 50% nonadherence)

We first consider the case that the ASCVD risk estimated by the score in Yadlowsky et al.

(2018), the ASCVD risk reductions from obtained from Sundström et al. (2014, 2015); Salam et al. (2019), and the treatment-related disutility are halved or doubled. We also examine an scenario where the treatment-related disutility results in equal number of medications recommended by the optimal treatment strategy and the current clinical guidelines. In this scenario, we used a disutility of 0.0092 for medications at half dosage and 0.0184 for medications at standard dosage. Second, we perform a sensitivity analysis on the distribution of the action-value functions. Rather than using an empirical estimation of their true distribution, we assume that the action-value functions, including terminal rewards, are normally distributed. In another sensitivity analysis, we use the treatment choices with the least amount of medications and median number of medications in the next year's range of near-optimal actions, instead of the best treatment treatment choice in the range. We also compare the performance of treatment choices in the range of near-optimal actions to the optimal treatment and the current clinical guidelines in a secondary population. Each policy is applied in a sample representative of all Black or White adults in the US with ages between 70 and 74 years old (7.55 million people). Finally, we study the case the parameters are misestimated. We contemplate three misestimation scenarios: patients' true risk is half the estimated risk, patients' true risk is double the estimated risk, and patients' true benefit from treatment is half the estimated benefit.

4.7.9 Numerical Results

In this subsection, we evaluate the effect of flexible hypertension treatment plans. We provide insights into the patient and population-level implications of flexible treatment. The results of our sensitivity analyses are included in subsection C.3.4.1 of Appendix C.

4.7.9.1 Patient-Level Insights from Flexible Treatment

We now evaluate the ranges of near-optimal actions in a series of patient profiles, based on patients from the NHANES dataset. For comparison purposes, we also determine the optimal treatment plans and the current clinical guidelines for each patient profile. We first obtain ranges of antihypertensive medications for the following patient profile: 54-year-old, non-diabetic, non-smoker, White male with stage 1 hypertension, and low TC, HDL, and LDL. This patient profile will be referred to as the base patient profile. Note that this profile has two major risk factors for ASCVD events, the BP and the HDL levels. We then modify the following characteristics of the patient: sex, race, diabetes status, smoking status, and age.

Figure 4.3 shows the ranges of near-optimal treatment choices for our selection of

patient profiles. In the base patient profile, we observe ranges from 4 to 7 treatment choices that correspond to recommending from 0 to 2 medications at standard dosage and 1 at half dosage over the planning horizon. We also notice that the best treatment in the ranges match the optimal treatment at every year. The current clinical guidelines do not recommend any pharmacological treatment until year 10, when the patient reaches a 10-year risk for ASCVD events slightly above 10%.

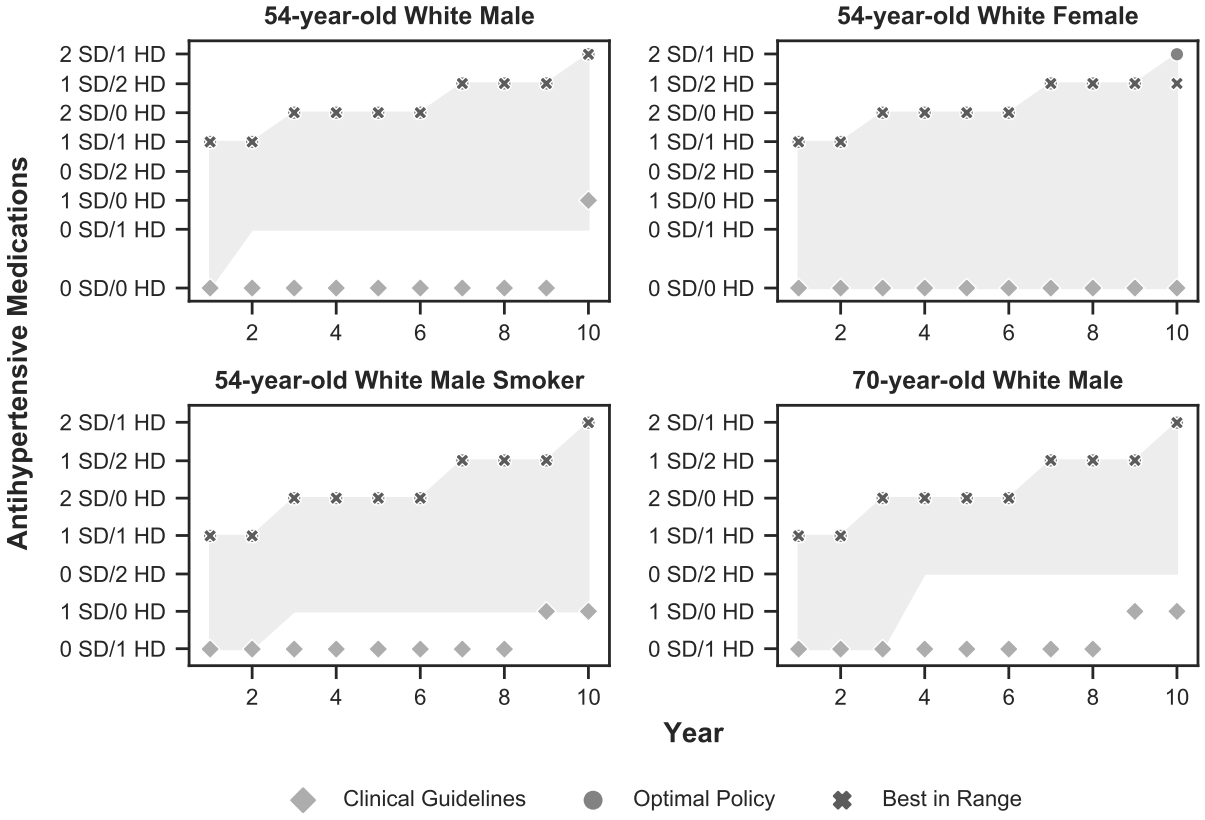


Figure 4.3: Ranges of near-optimal treatment choices per patient profile. The ranges are highlighted with the gray shaded area in each profile. The labels "SD" and "HD" denote antihypertensive medications at standard dosage and half dosage, respectively.

Changing the sex of the patient to female increases the width of the ranges but decreases the level of aggressiveness of the treatments prescribed. Using this patient profile, we find that the best treatment in range is slightly lower than the optimal treatment at year 10 of our study. However, the optimal treatment is contained in the range, and the optimality gap is relatively small (0.0026 life-years). We find a similar behavior changing the race of the patient to Black. A potential reason for this behavior is that, with otherwise-identical risk factors, a White female and a Black male patient would have lower risks than a White

male patient for the considered set of characteristics (Yadlowsky et al., 2018; Goff et al., 2014).

Modifying the diabetes and smoking status of the profile seems to have comparable effects. In both cases the ranges become narrower and more aggressive treatment is prescribed. We also note that the current clinical guidelines recommend more aggressive treatment as the risk for ASCVD events increases. Increasing the age of the base patient profile to 70 years has the biggest impact in the behavior of the ranges. We discover ranges from 4 to 5 treatment choices that correspond to recommending from 1 medication at half dosage to 2 medications at standard dosage and 1 at half dosage over the planning horizon.

4.7.9.2 Population-Level Insights from Flexible Treatment

Out of a population of 16.72 million people, 1.33 million (7.96%) are Black females, 7.58 million (45.34%) are White females, 1.08 million (6.44%) are Black males, and 6.73 million (40.26%) are White males. The number of people by sex, race, and BP group (normal BP, elevated BP, stage 1 hypertension, and stage 2 hypertension) at the first year of our study are shown in Figure C.2 in Section C.3.4 of Appendix C. We observe that male patients generally have higher BP than female patients. We also notice that 6.44 million (38.49%) people have stage 1 hypertension. This is the largest proportion of adults in the US with ages 50 to 54. Nevertheless, this finding varies by race. While there are more Black adults with stage 1 hypertension than any other BP group (45.83%), the largest proportion of White adults have normal BP (38.59%). These findings are consistent with the most recent age-adjusted hypertension prevalence trends across adults in the US (Virani et al., 2020).

Comparison of Treatment Recommendations: By comparing the treatment strategies contained in our ranges of near-optimal actions to the optimal treatment policies and the current clinical guidelines, we are able to obtain insights into how treatment changes by demographic. The distribution of treatment recommended by each policy per BP category at year 1 and 10 of our study is shown in Figure 4.4. Other than more intense treatment over time, we note that the distribution of treatment did not change considerably in years 2 through 9. The results segregated by BP category, sex, and race are shown in figures C.3 and C.4 in Section C.3.4 of Appendix C.

From the distribution of treatment recommendations, we observe that almost no patients receive treatment in the normal BP category at any given year (less than 0.3% of the population). Comparing the treatment strategies contained in our ranges to the optimal treatment plans and the current clinical guidelines, we observe that recommending the

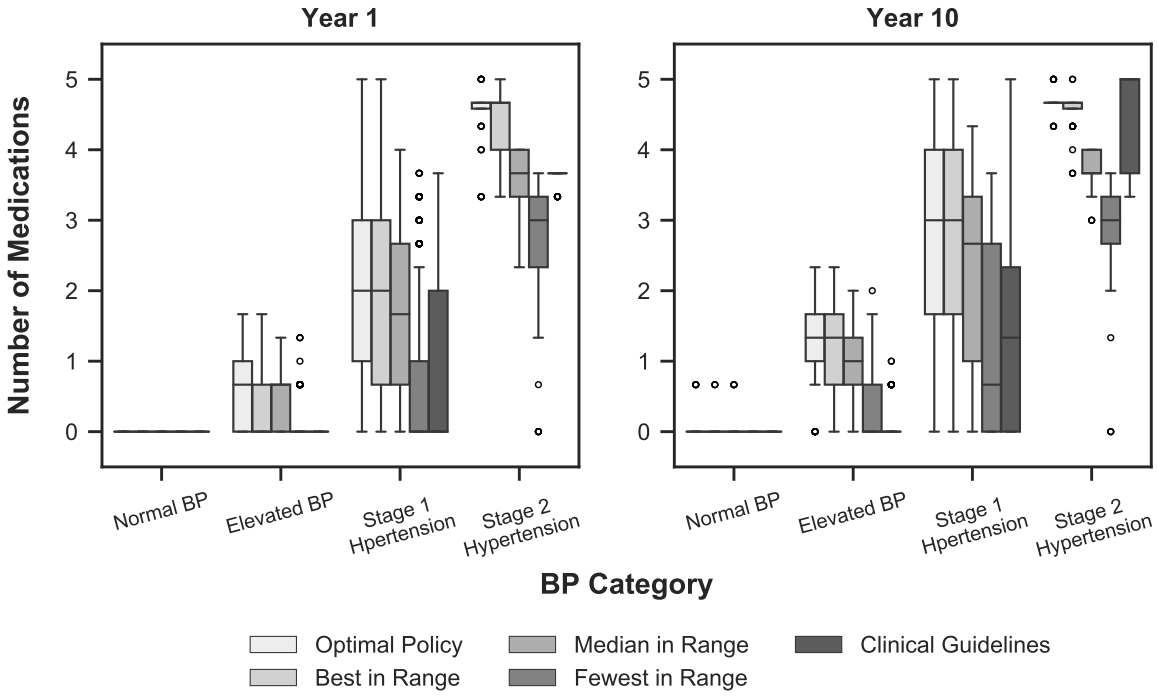


Figure 4.4: Distribution of treatment at year 1 and year 10 of the study. BP categories are made based on patients' characteristics at year 1.

best treatment in the range is typically close to optimal. We find that recommending the best treatment in the ranges is often equivalent to recommending the largest number of medications contained in the ranges. The best treatment in range is between the optimal treatment and the current guidelines treatment strategy, but considerably closer to optimal treatment in all BP categories. We also note that current clinical guidelines are between recommending the fewest and the median number of medications in the ranges in normal BP, elevated BP, and stage 1 hypertension. The current clinical guidelines are generally between recommending the largest number of medications and the median number of medications in the ranges for patients with stage 2 hypertension.

Examination of Treatment Choices Contained in the Ranges: Since very few people receive treatment under any of the policies in the normal BP category, we focus on patients with elevated BP, stage 1 hypertension, and stage 2 hypertension. We now study the proportion of patients whose ranges of near-optimal actions contain the optimal treatment choice and the current clinical guidelines by sex, race, and BP group.

The ranges of near-optimal actions always contain the optimal treatment plans in all years, demographics, and BP groups. However, we observe an overall decreasing trend in the proportion of patients treated according to the current clinical guidelines that are included in the ranges. This may be because the ranges of near-optimal treatment choices are informed by risk and the current clinical guidelines are mainly driven by BP levels. Another reason could be that the current guidelines do not consider the impact of present decisions on the future health of patients, while the ranges of α-nonsignificant do. The proportion of patients for whom the ranges cover the actions recommended by current clinical guidelines over the 10-year planning horizon stratified by sex, race, and BP category is shown in Figure C.5 in Section C.3.4 of Appendix C.

Effect of Treatment Recommendations: We now proceed to evaluate the outcomes of patients under each treatment strategy in terms of the life-years saved, the ASCVD events prevented, and the expected time to an averse event, compared to no treatment. In total, the best treatment, the median number of medications, and the fewest number of medications contained in the ranges save 2.92, 2.55, and 1.75 million life-years. The optimal treatment plans and the clinical guidelines treatment strategy save 3.02 and 1.83 million life-years, respectively.

Evaluating our results by BP category, we find that patients with stage 1 hypertension receive the greatest benefit from treatment (Figure 4.5). We note that patients' health outcomes under the best treatment choice in the ranges are not substantially different

from the health outcomes based on optimal treatment. In patients with elevated BP, our treatment strategies outperform the clinical guidelines. The health outcomes of patients with stage 1 hypertension under the clinical guidelines are similar to the life-years saved under the strategy that recommends the fewest number of medications contained in the ranges. All the treatment policies result in similar life-years saved in patients with stage 2 hypertension. We include our results separated by sex, race, and BP category in figures C.6 and C.7 in Section C.3.4 of Appendix C.

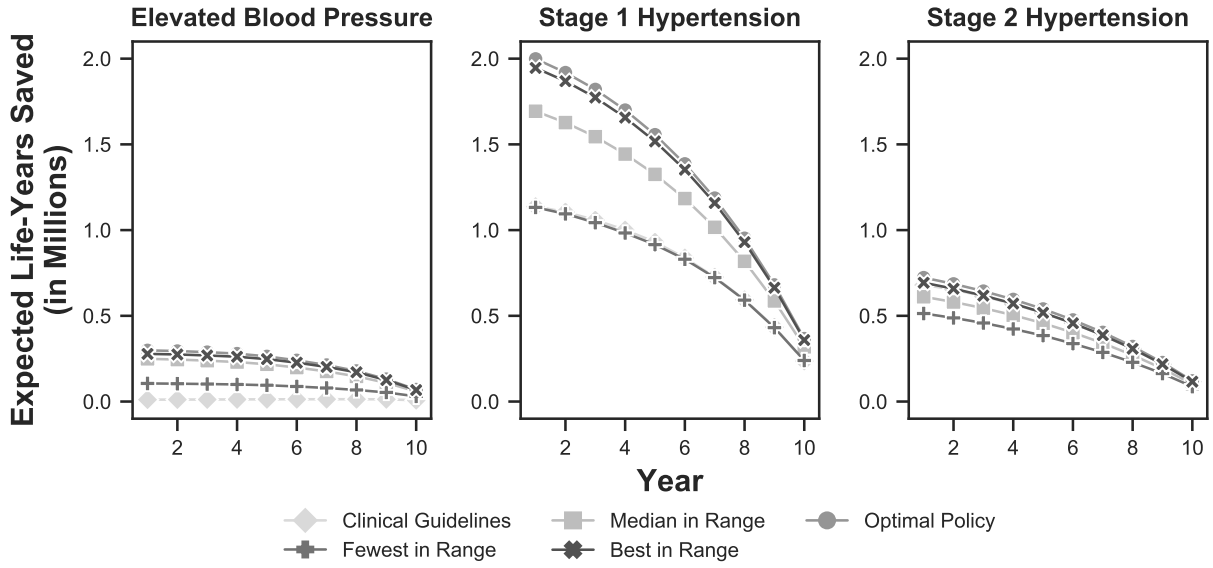


Figure 4.5: Life-years saved by each treatment policy compared to no treatment per BP group over the planning horizon.

We notice a similar pattern when comparing the policies in terms of ASCVD events averted. Over the 10-year planning horizon, the best, the median number of medications, and the fewest number of medications contained in the ranges prevent 176, 154, and 103 thousand ASCVD events, compared to no treatment.

With regards to the expected time to an averse event, we also observe that the best treatment in range is close to the optimal policies, and the treatment choice with the fewest number of medications is similar to clinical guidelines. We notice that events are expected to be delayed in 5.48, 4.61, and 2.93 life-years by the best, the median number of medications, and the fewest number of medications contained in the ranges, respectively. The optimal treatment and current guidelines delay adverse events in 5.60 and 2.96 life-years, respectively.

4.7.9.3 Population-Level Insights from Sensitivity Analyses

We proceed to study how the treatment strategies are affected by changing the model parameters and assumptions. The results of our sensitivity analysis are summarized in Table 4.5. All values correspond to the results at the first year of our study.

Table 4.5: Summary of sensitivity analyses at the first year of our study.

Sensitivity analysis scenario	Life-years Saved[a]						
	Optimal Treatment	Best in Range	Median in Range	Fewest in Range	Clinical Guidelines	Number of Medications[b]	Range Width[b]
Base case	3.02	2.92	2.55	1.75	1.83	1.93 (0, 4.33)	3.07 (1, 9)
ASCVD event rates							
Halved	3.02	2.95	2.67	2.07	1.65	2 (0, 4.33)	2.86 (1, 9)
Doubled	3.02	2.96	2.62	1.94	2.10	2.02 (0, 4.33)	3.06 (1, 10)
Treatment benefit							
Halved	2.11	2.02	1.67	1.05	0.99	2.44 (0, 4.67)	5.09 (1, 15)
Doubled	3.44	3.38	3.17	2.47	2.94	1.31 (0, 4.33)	1.81 (1, 4)
Treatment-related disutility							
Halved	3.18	3.07	2.68	1.83	1.89	1.93 (0, 4.33)	3.03 (1, 9)
Doubled	2.73	2.62	2.30	1.58	1.70	1.93 (0, 4.33)	3.21 (1, 10)
Equal treatment	1.29	1.21	0.78	0.71	0.79	1.6 (0, 4)	3.37 (1, 10)

[a] The life-years saved by each policy are presented in millions.
[b] The value outside the parenthesis is the average, the values within the parenthesis are the 5th and 95th quantile across the population of adults in the US with ages between 50 and 54.

If the event rate for ASCVD events is half of the estimated rate by Yadlowsky et al. (2018), the life-years saved by the treatment strategies are generally closer to the optimal treatment policy. We also note that all the treatment strategies contained in the ranges outperform current clinical guidelines in this scenario. If the event rate is doubled, the current clinical guidelines save more life-years than recommending the fewest number of medications in the ranges. This may be because the current clinical guidelines start treatment for patients with stage 1 hypertension if they have a 10-year ASCVD risk of at least 10%. As a result, if the ASCVD risk is changed, a different number of people with stage 1 hypertension benefit from treatment. We observe that the number of medications covered by the ranges and the number of treatment choices contained in the ranges (or range width) do not change considerably if the risk for ASCVD events is halved or doubled.

We find that changes in the benefit from treatment have a large effect in the treatment policies than changes in the event rates. If the benefit from treatment is halved, all

policies save less life-years than in the base case. We also note that the ranges contain more treatment choices and that treatment is more aggressive than in the base case. The opposite effect is observed if the benefit from treatment is doubled. In this scenario, less medications are necessary to ensure the well being of patients and less treatment choices are within 0.02 life-years of the best action, which results in narrower ranges.

In general, we notice that the treatment-related disutility considerably affects the life-years saved by each policy but the treatment strategies themselves to a lesser extent. If the treatment-related disutility is increased until the optimal treatment policy recommends the same number of medications as the clinical guidelines, we observe a dramatic reduction in the life-years saved by each strategy, the number of medications covered in ranges, and the width of the ranges. In this scenario, the optimal, best in range, median in range, and fewest in range strategies tend to recommend less aggressive treatment, which results in a lower number of life-years saved. Although the current clinical guidelines do not use disutility as a driver for recommending treatment, this strategy also result in less life-years saved when evaluated in the Markov chain embedded in the MDP.

Normally Distributed Action-Value Functions Scenario: We observe that assuming that the immediate rewards and terminal rewards are normally distributed does not have a substantial effect of the width of the ranges. Overall, we find that the quantile values $d_t(\alpha)$ obtained using the parametric method developed by Dunnett (1955) is reasonably robust to the type of nonnormality exhibited in the action-value functions associated with each patient's state and treatment recommendation. This finding is consistent with previous studies on the robustness of Dunnett's method (Bretz and Hothorn, 2003; Westfall, 2011). In line with Proposition 4.10, we also notice that the width of the ranges and the approximately optimal actions do not change with the treatment choice at the next decision epoch if $\Pi_t(s, \alpha)$ can be ordered as a range. The average range width and number of medications for our base case, assuming normality in the action-value functions, using the action that corresponds to the median number of medications in next year's range, and using the action to corresponds to the fewest number of medications in next year's range is included in Figure C.8.

Ages 70 to 74 Scenario: Applying each treatment strategy to the adult population in the US with ages between 70 and 74 year old, we can draw similar conclusions than with out base case population (adults with ages from 50 to 54). In this population, the treatment strategies contained in the ranges of near-optimal actions save more life-years than the current clinical guidelines in every BP category and demographic. This may be because

older patients tend to have higher risk for ASCVD events than younger patients, which translates to more intense treatment by the policies contained in the ranges. We also note that the best treatment in the ranges result in similar health outcomes to the optimal treatment plans. The life-years saved by each policy segregated by sex and BP category as well as by race and BP category are included as Figures C.9 and C.10.

Benefit Misestimation Scenarios: Figure C.11 shows the proportion of patients whose treatment is covered by the ranges of near-optimal actions despite parameter misestimation. We notice that the ranges of treatment choices are generally robust against event rate misestimation. The largest difference between the proportion of patients whose treatment is covered in the ranges in the base case and the event rate misestimation scenarios is 4.58%. Furthermore, we find that the optimal policies are always contained in the ranges of near-optimal actions. While the proportion of patients whose treatment is covered in the ranges remained unchanged in the clinical guidelines treatment strategy in the treatment benefit misestimation scenario, this proportion drops by up to 53.73% in the optimal treatment plans. A potential explanation for this decrease in coverage is that the optimal treatment strategy treats almost twice as aggressively if the true benefit from treatment is half of the misestimated benefit.

4.8 Discussion

In this chapter, we introduced a new method to obtain sets of near-optimal actions in finite MDP models. Additionally, we presented an alternative notion of optimality, which we called α-nonsignificance. We propose two algorithms to achieve the sets of α-nonsignificant actions: the SBBI and the SBMCC algorithm. The SBBI algorithm works in a similar way to the standard backwards induction algorithm, except that we replace the expectation by a sample-average approximation. We showed that the estimates attained with the SBBI algorithm converge to their true values with probability 1 exponentially fast. The SBMCC algorithm leverages ideas from the nonoverlapping batch means method to produce simultaneous confidence intervals without any distributional or equal variance assumptions. We proved that the method reaches the correct coverage with probability 1 at an asymptotic rate of $\sqrt{N}$. In addition, we provided sufficient conditions to ensure the monotonicity of the sets of α-nonsignificant actions in time and states. Lastly, we gave conditions to guarantee that the sets of near-optimal actions will be ordered as a range. By providing DMs with a set of actions from which they are able to choose from, we improve the usability and acceptance of MDP models in practice.

In our case study, we examined the implications of flexible hypertension treatment plans at a patient and a population-level. Several conclusions can be made from this study. First, how much flexibility a clinician may receive to treat a patient depends on the patient's sex, race, and BP. As a consequence, the benefit from treatment choices contained in the ranges vary by sex, race, and BP group. Second, current clinical guidelines may be under-treating patients with elevated BP and over-treating patients with stage 2 hypertension. This results in life-year losses compared to the strategies contained in the ranges of α-nonsignificant treatment choices. Finally, the estimates of the benefit of treatment have a major effect on the type of treatment patients receive and the amount of flexibility clinicians may have to treat patients. The estimates of the risk for ASCVD events and treatment-related disutility have a smaller impact. As new evidence of the effectiveness of BP treatment becomes available, the ranges of near-optimal antihypertensive medications may become more accurate in medical practice.

There are opportunities for future work that build upon our ranges of α-nonsignificant actions. Our SBBI algorithm takes a maximum over a set of random variables, which can lead to positive statistical bias. Nevertheless, this bias is expected to decrease as the number of observations increase, in a similar way to the statistical bias of the sample-average approximation in stochastic programming (Mak et al., 1999). Ideas such as the double Q-learning method introduced by Van Hasselt (2010) may be used to correct for this source of bias in finite sample settings. Another extension of this work could be to adapt our methodology to partially observable and infinite-horizon MDP models. For the later type of models, the ideas from Haskell et al. (2016) could be used to obtain empirical estimates of the value and action-value functions. It may also be worthwhile to allow for continuous state and action spaces. Structural results may be necessary to guarantee the convergence in these cases. We acknowledge that the SBBI and SBMCC algorithms inherit some of the curses of dimensionality associated with standard backwards induction. Overcoming these difficulties may be an additional area for future work. Additionally, our algorithms are limited by their storage requirements. This issue could be addressed with the development of an online method to obtain ranges of α-nonsignificant actions.

From a clinical perspective, our work could be extended by incorporating other conditions, such as high cholesterol or diabetes. Based on communications with clinical collaborators, we decided to develop ranges of antihypertensive medications as a starting point. Integrating the treatment of multiple conditions will likely result in greater flexibility. Our results provide a lower bound on the amount of flexibility clinicians and their patients could receive in the implementation of decision strategies. An alternative to the modeling approach presented in this chapter could be to model the uncertainty in

the choices made by primary care providers and their patients directly from observational data (such as interviews). In this case, the inference would be made around patients' characteristics. However, the conclusions achieved with this alternative technique would require assumptions on the level of rationality of the patients.

4.9 Conclusions

The ranges of α-nonsignificant actions presents a new line of work by handling stochastic optimization problems as hypothesis testing problems. Providing several suggestions at each state and decision epoch in a sequential decision problem presents domain experts with an effective way to integrate their knowledge into mathematical models. A range of near-optimal choices could have many benefits in practice, such as better user experience and flexibility.

Chapter 5

Conclusions and Future Work

This dissertation presented new methods that use population and patient-level data to make better healthcare decisions. The analyses included in this dissertation not only make contributions to the medical and health policy fields, but also make technical advances to the body of literature in decision-making under uncertainty.

We divided the research presented in this dissertation into two parts. The first part of the dissertation (Chapter 2) presented methods to predict the supply and demand of organs for transplantation. The second part of this dissertation (Chapters 3 and 4) introduced approaches to improve the management of ASCVD.

Chapter 2 incorporated the use of heuristics, predictive modeling, and stochastic simulation methods to predict the supply, demand, and allocation of organ for transplantation. First, we built a population-based method to forecast liver supply. Second, we modeled the temporal relationship between changes in population characteristics and additions to the LT waiting list. Third, we created a model to understand how liver allocation policies and demographic changes may impact future liver availability. Finally, we developed a machine learning to predict deceased donor organ yield. The implementation of this work can potentially change the current liver allocation policies. It may serve as a guide so that steps taken for future planning alleviate some of the mismatch between liver donors and recipients.

In Chapter 3, we integrated dynamic programming, VoI analysis, and stochastic simulation to assess the impact of genetic information in cholesterol treatment policies. We presented a simulation-based framework to model genetic risk. Also, we aimed to understand how treatment protocols can be driven by genetic information. Lastly, we introduced a mathematical model to simultaneously determine optimal cholesterol treatment plans and genetic testing strategies. This work provides medical doctors and health insurance policy makers with a better understanding of the potential benefits of performing genetic testing in different populations. In addition, our models could serve as an aid for health

insurance policies by identifying patients whose cholesterol treatment will likely improve after acquiring genetic information.

Chapter 4 proposed a new approach to obtain sets of near-optimal actions for finite MDP models. This new approach combined elements from SBDP, non-overlapping batching, and statistical multiple comparisons. We provided finite sample, convergence, and asymptotic structural properties of our method. By analyzing the structure of the sets, we characterized their behavior with respect to the modeling data and identified when they can be ordered as a range. To show the scalability of our approach, we obtained ranges of near-optimal actions by applying our method to the management of hypertension. A range of treatment choices could have many advantages in medical practice, such as flexibility and better user experience.

In the remainder of this chapter, we briefly discuss possible extensions of our work. First, we present a potential method to bridge the gap between organ supply and demand, motivated from our research in Chapter 2. Then, we discuss the prospect of allowing for improvements in the quality of genetic testing as an extension to our work in Chapter 3. Afterwards, we propose the consideration of input data uncertainty as future work originating from our research on Chapter 4. We end the dissertation with some final remarks.

5.1 Future Work: Evaluating the Impact of Interventions to Narrow the Gap Between Organ Supply and Demand

Several approaches have been suggested in the medical literature to increase deceased organ donation. These approaches include: donor management and better consent practices, improving the number of donors per eligible death, increasing the use of organs from donation after cardiac death, improving public opinion and willingness of next-of-kin to provide consent, and increasing the number of organs transplanted per donor or organ yield (Jenkins et al., 1999; Manninen and Evans, 1985; Marks et al., 2006; Rodrigue et al., 2006; Selck et al., 2008). Population-level data could be used to estimate the potential implications of each of these approaches in the US. For example, DeRoos et al. (2019) simulated the impact of a presumed consent policy in organ availability. In this analysis, the authors found that presumed consent alone is not likely to solve the organ shortage in the US.

One way to continue to address the organ shortage in the US is to combine multiple interventions. The potential implications of each of these interventions must be evaluated

individually and combined. There are three key challenges in this analysis: (1) the future population growth and demographic shifts are uncertain, (2) the true benefits and harms of each intervention are unknown, and (3) it is unclear if combining any of these interventions may have negative effects. The future population growth and shifts could be estimated from historical data. The benefits and harms of each intervention could be estimated using data from other countries (e.g., Rithalia et al. (2009)). Sensitivity analyses are warranted to ensure that any possible negative consequences are studied in detail. Future research in this area can focus on addressing the following questions:

1. What is the best way to forecast the population growth in the US?
2. What would be the effect of demographic shifts in organ availability?
3. What has been the effect of similar interventions to reduce the organ shortage in developed countries? How does the donation system in these countries compare to the US?
4. When an intervention to reduce the gap between organ supply and demand would produce adverse results? When the combination of interventions result in negative effects?
5. How to model the effect of a change in the donation system in donors and waiting list candidates' behavior?

Future research in this area could also consider decision-making tasks. For instance, the prediction of the effect of interventions to alleviate the future disparity among the expected organ demand and availability could also inform allocation policies. Specifically, policies to allocate the organs for transplantation within and across DSAs could be developed based on these prediction models.

5.2 Future Work: Understanding the Role of Improvements in the Quality of Information in Genetic Testing

Genetic testing is a rapidly evolving field. Since the mapping of the human genome in 2003, over 75,000 genetic tests have become available in the market (Phillips et al., 2018). In Chapter 3, we assumed that the quality of genetic testing will remain constant over time (Khera et al., 2018). However, tests with greater accuracy may be available in the future. An extension of the work presented in this chapter could be to allow for potential

improvements in the quality of genetic testing over time. The difficulty of this setting lies in the additional source of variability. Even if a genetic test is beneficial for a patient, it might be better to wait as a future test could provide greater gains. This uncertainty may also impact treatment decisions. The development of new stochastic optimization methods may be required to address such a setting. Alternatively, the patient-level genetic test adoption decisions could be modeled as an optimal stopping problem (Liu et al., 2017). Simulation or Markov models could then be used to evaluate the impact of the test adoption decisions over the lifetime of patients. The following questions may be addressed by future research in this area:

1. When is best to perform a genetic test? Is there a threshold-type policy for the testing decisions?
2. How the testing decisions impact the downstream treatment decisions and vice versa?
3. How do the treatment policies and testing strategies compare with and without improvements in the quality of genetic information?
4. How patients' characteristics affect the treatment and testing decisions?
5. How the simultaneous treatment and testing policies compare to individual treatment and testing strategies? Could the integrated problem be decomposed into two smaller ones?
6. How do the optimal treatment and testing policies compare to the clinical guidelines?

Future work in this area could go beyond cardiovascular diseases. For example, hereditary cancer tests have second highest spending percentage in the US (Phillips et al., 2018). Genetic tests in this domain are evolving at a fast pace, and new tests frequently become available. Treatment decisions could be influenced by the results of these tests. This body of research could be used to determine simultaneous testing and treatment decisions.

5.3 Future Work: Flexible Decision Support Under Input Data Uncertainty

Using the flexible protocols introduced in Chapter 4, physicians can obtain decision support given an accurate model of their system of interest. However, there may be circumstances when there is disagreement among medical experts regarding how to model

patients' disease progression and the consequences of clinical interventions. Debate on how to characterize the evolution of patients' health a system introduces extrinsic uncertainty, which may influence stochastic models in ways that are difficult to predict. One way to address this setting is to develop adaptive guidelines that allow for multiple estimates of the health progression and the impact of clinical interventions. These sources of variability could then be used to obtain sets of near-optimal treatment choices. This new framework could require advances in stochastic optimization, statistical multiple comparisons, and ADP/RL. Alternatively, the uncertainty could be expressed as intervals and linear programming with interval coefficients could be used to obtain the best and the worst optimal solutions (Chinneck and Ramadan, 2000). New methods in interval linear programming may be needed to obtain the solutions between the best and the worst optimal solutions. Future work in this area could focus on the following research questions:

1. How the confidence in the stochastic model affect the adaptive guidelines?
2. How the sets of near-optimal actions compare with and without input data uncertainty?
3. What values of the input data would lead a particular action to be optimal or a set of actions to be near-optimal?
4. How patients' characteristics affect the sets of near-optimal actions?

This work has the potential to be applied in many situations where humans play a role in the decision-making process. Moreover, the work in this area could also be applied in non-healthcare domains, such as inventory management and transportation science.

5.4 Future Work: Synergies Across Chapters

Another extension of the work presented in this dissertation is to exploit the potential for synergies in the methods presented throughout Chapters 2, 3, and 4. Population-level data could be used to identify trends across different grouping variables, such as demographic characteristics. These trends could serve to inform patient-level decisions. Conversely, combining the outcomes of individuals across a population could may be used to aid policy-level decisions. Furthermore, it may be worthwhile to explore the combination of multiple population-level methods or patient-level approaches. Future research in this area can address the following questions:

1. How do patient-level outcomes affect population-level policies?

2. How do population-level trends influence individual-level interventions?

3. How to perform testing decisions given flexible treatment plans (i.e., multiple suggestions instead of a single decision rule)?

The first and second research questions provide examples of synergies between the methods contained in the first part (Chapter 2) and the second part (Chapters 3 and 4) of this dissertation. On the other hand, the third research question gives a potential integration among the techniques proposed in Chapters 3 and 4.

5.5 Final Remarks

In this dissertation, we developed new techniques that use population and patient-level data to inform health policy and medical decisions. Generally, we addressed the following questions:

1. How can we use population-level data to model the future supply and demand of valuable resources?

2. How can we use patient-level data to personalize disease progression over time?

3. How can we use the personalized estimates of disease progression to obtain tailored treatment and testing strategies?

Our findings highlight the importance of population shifts and patients' characteristics in public health and clinical decisions. This research also advances the theory of data-driven decision making by providing new frameworks to make multiple decisions simultaneously and to obtain sets of near-optimal actions in the context of MDP models. As more healthcare data becomes available, new methods that leverage these data to make better health policy and medical decisions will be needed. Our work could serve as a foundation for future research at the intersection of operations research, statistics, and healthcare.

Appendix A

Appendix for Modeling Supply, Demand and Allocation in Liver Transplantation

A.1 Machine Learning Models

The machine learning models included in our analysis are briefly described in this section.

A.1.1 Mean-Only Model

This model represents using the average overall yield across all the deceased donors included in 80% of the derivation cohort to make predictions in the random holdouts (remaining 20%).

A.1.2 Poisson Regression

A generalized linear models (GLM) is an extension of the linear model that allows for response variables with error distribution models other than normal, such as binomial or Poisson (Faraway, 2006). Since the overall deceased donor organ yield is a count response variable, it can be modeled using the Poisson distribution. The Poisson GLM assumes that the mean of the response variable is equal to its variance. We test this assumption using a regression based test for overdispersion and the AER package in R (Cameron and Trivedi, 1990; Kleiber and Zeileis, 2012). A stepwise variable selection method is used to obtain a useful set of covariates to predict the overall deceased donor organ yield. Likelihood ratio tests are also used to verify that simpler models could not be obtained after removing any of the selected variables (Hastie et al., 2009).

A.1.3 Negative Binomial Regression

The NB GLM is a generalization of the Poisson GLM where the rate parameter follows a gamma distribution (Faraway, 2006). It is of particular importance when overdispersion, or the variance of the response variable is greater than its mean, is present in the Poisson GLM. For this model a stepwise variable selection method and likelihood ratio test are also used to attain useful set of covariates to predict the overall deceased donor organ yield (Hastie et al., 2009).

A.1.4 Generalized Additive Model

A GAM is an extension of the linear model that allows for non-linear relationships between the response variable and the predictors. This model express the relationship between the response variable and the predictors using a sum of smooth functions. By adding a smooth function to each predictor, GAMs can capture non-linear relationships between the predictors and the response variable that traditional linear models are not able to capture (Hastie and Tibshirani, 1986). The degree of the smooth functions is chosen to be large enough to represent the true relationship between the predictors and the response variable reasonably well, but small enough to maintain computational efficiency. Model selection is executed with penalized regression splines and the mgcv package in R (Wood, 2006; Wood and Augustin, 2010; Wood, 2013).

A.1.5 Tree-Based Methods

Four tree-based models are considered during this analysis. First, a CART model is considered. The CART models are first grown and then pruned using cost-complexity pruning to avoid over-fitting (Breiman et al., 1984). The second tree-based method is boostrap aggregation. Bagging trees has the potential benefit of reducing the model variance and improving model accuracy if each individual tree is unbiased (or has low bias) with uncorrelated (or weakly correlated) predictions (Breiman, 1996). A total of 500 trees are bagged for this model.

The third tree-based model is a random forest. This model is created by averaging the predictions from individual CART models trained on separate bootstrapped resamples of the data, just as for the tree-based boostrap aggregation. Nevertheless, this model is modified such that each tree is fit using only a subset of the predictors with the purpose of reducing the correlation among predictions. During this analysis, we use 5-fold cross-validation to determine the proportion of predictors used to fit each of the 500 bagged

trees (Hastie et al., 2009; Kuhn, 2008).

Finally, tree-based gradient boosting is considered. The boosting algorithm sequentially grows trees using information from previous trees. This algorithm creates an ensemble of weak learners to produce final predictions (Hastie et al., 2009). We use 5-fold cross-validation to select the depth of variable interactions and learning rate of the tree ensemble based on the literature (Hastie et al., 2009; Kuhn, 2008; Ridgeway, 2007).

A.1.6 Bayesian Additive Regression Trees

BART are a Bayesian sum-of-trees model that could be used either for regression or classification problems. Each tree in this model is constrained by a prior probability to be a weak learner, biasing the tree towards a simpler structure. This constraint ensures that each tree will only contribute slightly to the overall fit. The fitting and inference are then achieved using an iterative back-fitting Markov chain Monte Carlo algorithm that generates samples from a posterior distribution (Chipman et al., 2012). We use 5-fold cross-validation and the existing literature to determine the number of trees used in the sum-of trees, prior probabilities, and variance prior parameters (Kapelner and Bleich, 2013).

A.1.7 Multivariate Adaptive Regression Splines

MARS are another extension of the linear model. The MARS model relates the predictors to the response variable using a weighted sum of basis functions. This non-parametric regression method allows for interactions and non-linear relationship between the predictors and response variable. Model selection is performed by pruning and 5-fold cross-validation (Friedman, 1991).

A.1.8 Artificial Neural Networks

ANN are usually composed of three layers of nodes: input, hidden and output layers. Each node in the input layer of the network represents a predictor in the model and the output layer of the network is usually comprised of a single node that represents the response variable (Hastie et al., 2009). The hidden layers of the network permit the ANN to model the relationship between the predictors and the response variable. Throughout this analysis, we use a single hidden layer and we choose the number of hidden nodes and rate of decay based on existing literature and 5-fold cross-validation (Klimberg and McCullough, 2013; Kuhn, 2008).

A.1.9 Selck, Deb, and Grossman's Ordinary Least Squares Regression

Among the models developed by Selck et al. (2008), we include the ordinary least square model adjusted with management variables since it presented the best predictive performance in their study. While this model includes management variables (three or more inotropes at recovery, T4 given at admission, steroids given at admission, diuretics given at admission, number of transfusions, terminal serum sodium, terminal pO2 >350 mmHg on 100%, non-missing indicator of pO2, any inotropes used, and desmopressin acetate given at admission) as predictors, management variables are not considered in the rest of models. To have a consistent definition of overall organ yield, we apply the methodology described in Selck et al. (2008) to our dataset with the objective of predicting 0 to 7 organs per donor.

A.1.10 Messersmith et al.'s Ordinal Logistic Regression

While Messersmith and coauthors developed overall and organ-specific yield models, we only consider their overall yield ordinal logistic regression model since it is directly comparable to our machine learning models. To implement this model, we incorporate donor age and BMI using natural cubic splines with knots at the values specified by Messersmith et al. (2011). Splines are piecewise polynomial functions used to approximate arbitrary functions. Natural cubic splines are commonly used splines built of piecewise third-order polynomials that pass through a set of knots. Each knot represents a place where the piecewise polynomial functions meet (Hastie et al., 2009). To have a consistent definition of overall organ yield, we apply the methodology described in Messersmith et al. (2011) to our dataset with the objective of predicting 0 to 7 organs per donor.

A.1.11 SRTR's Aggregate of Binomial and Multinomial Logistic Regressions

We re-build the SRTR models using splines with knots at the values specified by the SRTR (Scientific Registry of Transplant Recipients, 2019; Hastie et al., 2009). To avoid complete or quasi-complete separation while considering all the variables included the SRTR models, we implement penalized binomial and multinomial logistic regression models using the ridge regularization with the glmnet package in R (Friedman et al., 2010). We select the shrinkage parameters in the models using 5-fold cross-validation. Separation may occur when all observations of a predictor (or set of predictors) have the same outcome in the response variable. One way to address this issue is through regularization methods.

A popular method is the ridge regularization (also known as L2 regularization), which decreases model complexity while keeping all variables in the model. The shrinkage parameter of the ridge regularization method controls the amount of penalty applied to the regression coefficients in the model (Hastie et al., 2009).

A.2 Model Adequacy Checks

The assumptions and convergence of each model are verified through every replication the Monte Carlo cross-validation analysis. We discard any potential model for overall deceased donor organ yield if any of its assumptions are violated or if convergence is not achieved. However, any machine learning model discarded as a potential final model is conserved in the Monte Carlo cross-validation analysis for comparison purposes.

A.2.1 Poisson Regression

The hypothesis of mean-variance equality in the Poisson GLM with all the available covariates and after variable selection is tested using the regression based test for overdispersion (Cameron and Trivedi, 1990; Kleiber and Zeileis, 2012). We find that there is not enough evidence to conclude that the expected values of the models are equal to the variances (all $P < 0.001$ over the 30 replications). Therefore, this model may not be appropriate to model the response variable and it is discarded as a potential final model.

A.2.2 NB Regression

The fitting algorithm of the NB regression with all the available covariates and the covariates selected after stepwise variable selection generally does not converge throughout the Monte Carlo cross-validation analysis. Moreover, we observe that in the cases that the algorithm converged, the dispersion parameter approached zero (essentially giving the results of a Poisson regression model). This finding suggests that our dataset may be underdispersed, violating the assumption that the conditional variance of the response variable is greater than or equal than its conditional mean. Thus, this model is also discarded as a potential final model to predict overall deceased donor organ yield.

Appendix B

Appendix for Understanding the Role of Risk in Optimal Cardiovascular Treatment Planning

B.1 Progression of Risk Factors Over Time

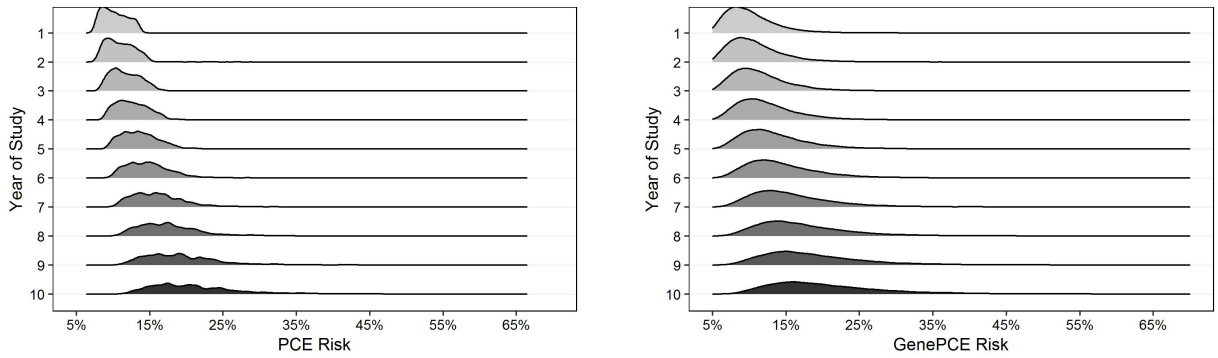
Table B.1: Linear regression models coefficients.

Variable	SBP	LDL	HDL	TC
Intercept	93.85	36.87	33.96	96.95
Age	0.55	3.17	0.45	3.69
Age squared	0.00	-0.03	0.00	-0.03
Sex (Female)	-2.71	1.92	11.12	8.87
Race (Black)	6.84	-3.03	1.69	-5.64
Smoking	1.36	1.48	-1.72	1.83
Diabetes	2.59	-13.96	-8.48	-17.35
Moderate intensity treatment	-1.04	-15.37	-2.66	-17.28
High intensity treatment	-0.27	-22.93	-2.37	-26.70

B.2 GenePCE Treatment Threshold Identification

To ensure fairness in the amount of treatment given by both policies, we empirically derive treatment cut-points for the GenePCE score. A new treatment thresholds is needed mainly because the distribution of the odds ratio for CHD is right-skewed. The asymmetry in the GRS causes most GenePCE scores to be less than the PCE scores (Figure B.1).

We model the states of the simulation and their transitions over time using as time-variant Markov chains. This allows us to calculate the expected amount of treatment given by each treatment strategy using backwards induction and policy evaluation methods (Puterman, 2014). We change the treatment threshold in the GenePCE strategy iteratively



(a) Distribution of PCE risk scores.

(b) Distribution of GenePCE risk scores.

Figure B.1: Distribution of risk scores over the 10-year planning horizon (only includes population that received genetic testing).

until both policies recommend the same amount of treatment for each patient over the 10-year planning horizon in expectation. The GenePCE treatment threshold that gives the same amount of treatment as the PCE strategy for the expected health trajectory of the patients in our primary population is 8.47%. We call this treatment threshold the equal 10-year treatment threshold. Note that this method is equivalent to running the state-based simulation "infinite" number of times for each patient and calculate the average amount of treatment given by both policies.

The same method is used to empirically derive treatment thresholds for different scenarios in our sensitivity analyses. For the case there is a 30% decline in the population ASCVD event rate due to unrelated factors, the GenePCE treatment threshold results in 9.50%. If the PCE risk increased by an odds ratio of 1.38 per standard deviation of the genetic score, the GenePCE treatment threshold is 9.30%. On the other hand, if the PCE risk increases by an odds ratio of 2 per standard deviation of the genetic score, the GenePCE treatment cut-point is 7.67%. For the cases that genetic testing is performed for anyone for whom a GRS outside 0.5 and 2 SDs away from the mean would alter their care, the GenePCE treatment thresholds result in 8.01% and 9.16%, respectively.

As part of our sensitivity analyses, we identify the treatment threshold in the GenePCE strategy that results in the same amount of treatment as the PCE strategy during the first year of our study. This is an iterative process where we change the treatment cut-point for the GenePCE strategy until the amount of treatment in both treatment strategies is the same. The GenePCE treatment threshold that gives the same amount of treatment as the PCE strategy during the first year of the study is 9.42%. We call this treatment threshold the equal initial treatment threshold. See Table B.2 for a summary of the GenePCE treatment

thresholds used in our analyses.

Table B.2: Summary of GenePCE treatment thresholds.

GenePCE treatment threshold	Description	Recommended moderate statin by PCE strategy, million patient-years	Recommended moderate statin by GenePCE strategy, million patient-years	Number of people receiving genetic testing, million
10%	Same treatment threshold as PCE strategy	127.51	110.41	16.17
8.47%	Equal 10-year treatment	127.51	127.51	16.17
9.30%	Equal 10-year treatment if the PCE risk increases by an odds ratio of 1.38 per standard deviation of the polygenic score	127.52	127.54	16.17
7.67%	Equal 10-year treatment if the PCE risk increases by an odds ratio of 2 per standard deviation of the polygenic score	127.55	127.51	16.17
9.08%	Equal 10-year treatment if the population ASCVD event rate decreases by 30%	67.46	67.44	16.17
8.01%	Equal 10-year treatment if genetic testing is performed for anyone for whom a GRS outside 0.5 SDs from the mean genetic score would alter their care	66.19	66.21	7.94
9.16%	Equal 10-year treatment if genetic testing is performed for anyone for whom a GRS outside 2 SDs from the mean genetic score would alter their care	221.62	221.77	30.70
9.42%	Equal initial treatment	127.51	117.22	16.17

B.3 Developing the AdjustedGenePCE Strategy

We hypothesize that the impact of genetic risk alters the relative benefit of treatment, not just the absolute risk of developing ASCVD. This alteration in relative benefit could substantially change when genetic risk is most useful and have been seen by Mega et al. (2015). The effect iz less clear in Natarajan et al. (2017). Because it can only be found in randomized trials that develop a genetic risk score, other studies have not been able to test for it.

To incorporate this information into clinical decisions, we first estimated the benefit of statins for patients with different degrees of genetic risk. If patients are in the lowest quintile of the genetic score, not only their likelihood of having a CHD event is decreased, but the RRR of statinsis also altered by 0.994. If their genetic score is in the second, third, or fourth quintile, we alter it by 0.889. If the simulated genetic risk is classified in the highest genetic score quintile, we adjust the RRR by a factor of 1.389 (Mega et al., 2015). These adjustments are normalized so that the effect of genetics does not change the statin benefit on average (Collins et al., 2016). The normalization of the adjustments is shown in Table B.3.

Table B.3: Impact of genetic risk on relative benefit of treatment.

Quintile	Genetic RRR	Moderate Intensity Statins		High Intensity Statins		Normalizing factor
		Statin RRR	Normalized Genetic RRR	Statin RRR	Normalized Genetic RRR	
1	0.34	0.194	0.183	0.278	0.263	0.944
2	0.32	0.194	0.173	0.278	0.247	0.889
3	0.32	0.194	0.173	0.278	0.247	0.889
4	0.32	0.194	0.173	0.278	0.247	0.889
5	0.5	0.194	0.27	0.278	0.386	1.389

If this adjustment is correct, choosing statins based on PCE risk alone would fail to recognize the variation in statin benefit. To develop an effective treatment guideline that would account for this, we create thresholds based on the expected absolute risk reduction (ARR) of treatment. This would treat the same number of people with statins (allowing a fair comparison with the other treatment strategies) but incorporate the added information of the genetic effects on treatment. To do this, we calculate the average effect of moderate intensity statins on the 10% 10-year PCE risk threshold (which is 1.94% 10-year ARR). We then calculate an ARR on patients for whom the effect of genetics on both the risk and the RRR of treatment is applied. If this ARR is greater than 1.94, this policy would recommend treatment with a moderate intensity statin. This policy would also recommend high intensity statins for anyone with history of ASCVD events. This will

treat the population similarly on average but take genetic risk into greater account in the decision-making process.

B.4 Convergence Analysis

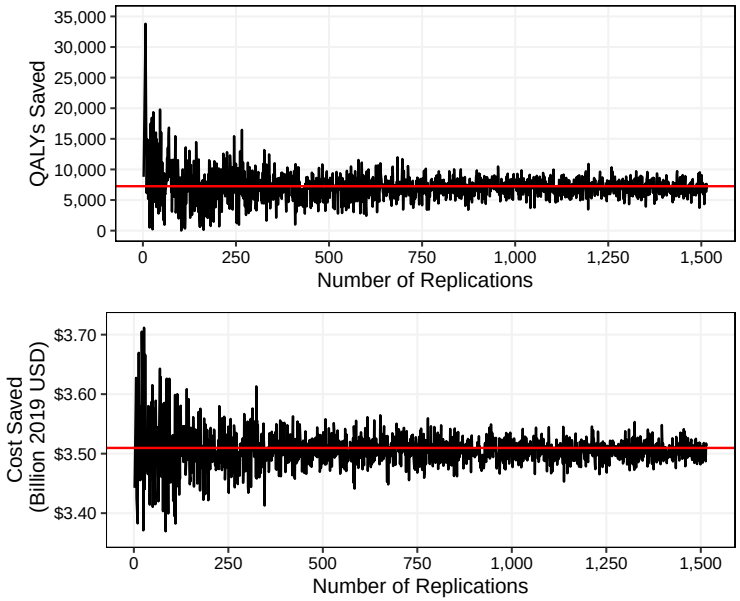


Figure B.2: Convergence of QALYs and cost saved in our simulation over the number of health trajectory replications under a single GRS realization per patient. Red line represents the QALYs and cost saved at 2,000 health trajectory replications.

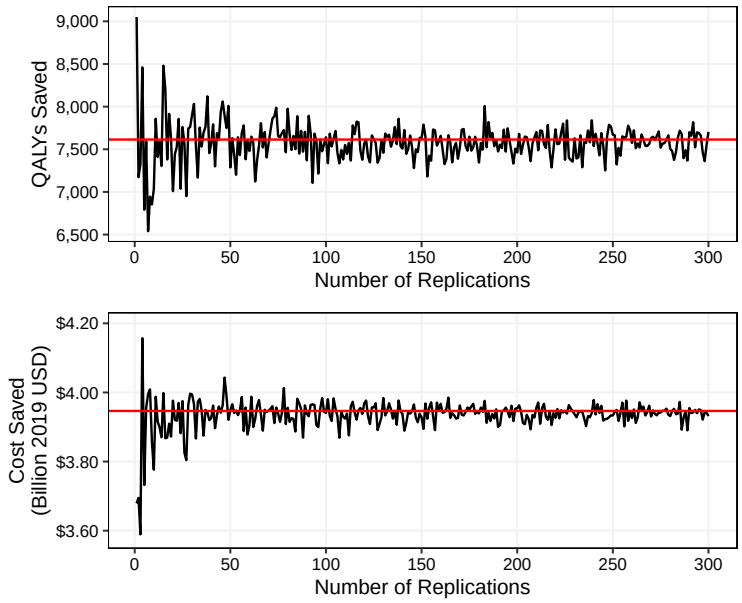


Figure B.3: Convergence of QALYs and cost saved in our simulation over the number of GRS realizations with a fixed number of heath trajectory replications. Red line represents the QALYs and cost saved at 500 GRS realizations and 500 health trajectory replications.

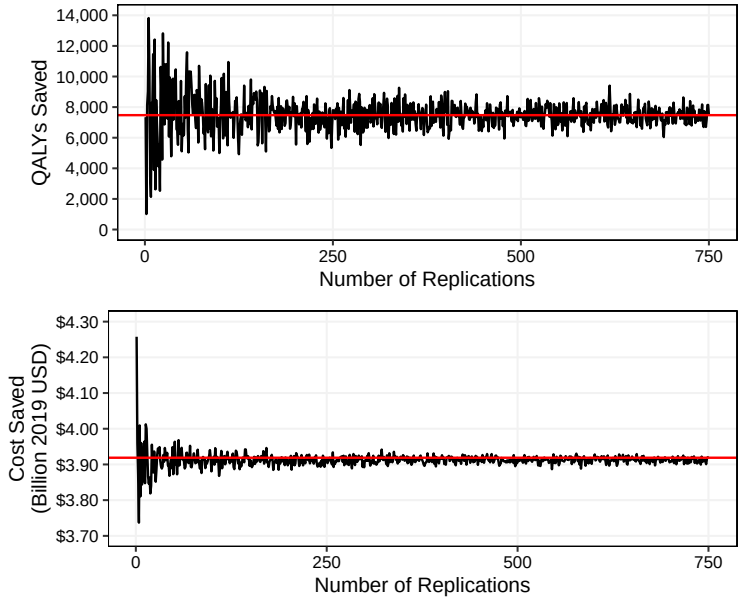


Figure B.4: Convergence of QALYs and cost saved in our simulation over the number of health trajectory replications under 100 GRS realizations per patient. Red line represents the QALYs and cost saved at 750 health trajectory replications and 100 GRS realizations.

B.5 Results of Sensitivity Analysis by Age Groups

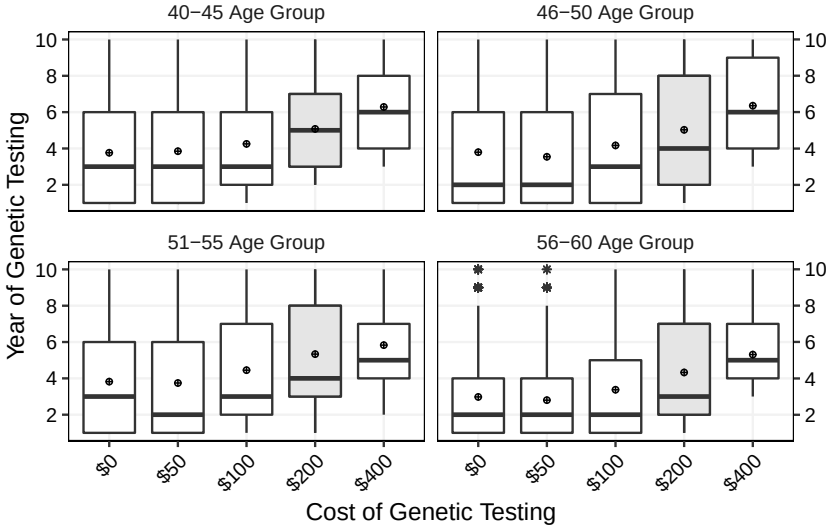


Figure B.5: Testing cost sensitivity analysis results by age group. Shaded boxplots represent the base case, asterisks represent outliers, and points in the center of the boxes represent the average testing year.

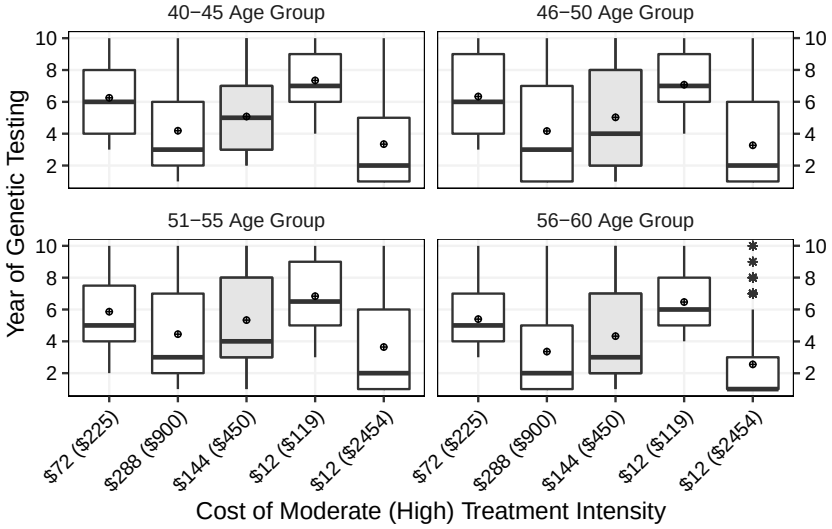


Figure B.6: Treatment cost sensitivity analysis results by age group. Shaded boxplots represent the base case, asterisks represent outliers, and points in the center of the boxes represent the average testing year.

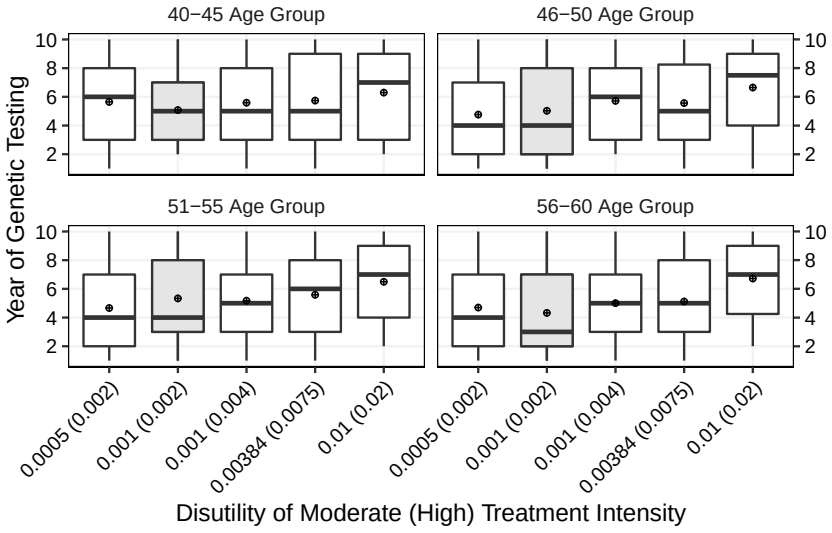


Figure B.7: Treatment disutility sensitivity analysis results by age group. Shaded boxplots represent the base case, asterisks represent outliers, and points in the center of the boxes represent the average testing year.

Appendix C

Appendix for Data-Driven Ranges of Near-Optimal Actions for Finite Markov Decision Processes

C.1 Algorithms

In this section, we provide additional details of our algorithms. We begin with the SBBI algorithm included as Algorithm 1. The algorithm requires the following inputs:

- $(\mathcal{T}, \mathcal{S}, \mathcal{A}, f, r, \gamma)$: a simulation MDP with transition functions f and rewards function r.
- N: desired number of observations; $N \in \mathbb{N}_+ \coloneqq \mathbb{N} \setminus \{0\}$.
- $\boldsymbol{v}_T$: set of terminal conditions; $\boldsymbol{v}_T \coloneqq \{v_T(s) : s \in \mathcal{S}\}$.

The outputs of the algorithm are as follows:

- $\boldsymbol{Q}$: set of sequences of observations of action-value functions across all decision epochs, states, and actions; $\boldsymbol{Q} \coloneqq \{\boldsymbol{Q}_t(s,a) : t \in \mathcal{T} \setminus \{T\}, s \in \mathcal{S}, a \in \mathcal{A}\}$, where $\boldsymbol{Q}_t(s,a) \coloneqq (Q_t^n(s,a) : n \in \{1,\ldots,N\})$ is a sequence of observations of the action-value function associated with state s and action a at decision epoch t.
- $\hat{\boldsymbol{Q}}$: set of action-value function estimates across all decision epochs, states, and actions; $\hat{\boldsymbol{Q}} \coloneqq \{\hat{Q}_t(s,a) : t \in \mathcal{T} \setminus \{T\}, s \in \mathcal{S}, a \in \mathcal{A}\}$.
- $\hat{\mathcal{A}}^*$: set of estimates of optimal actions; $\hat{\mathcal{A}}^* \coloneqq \{\hat{\mathcal{A}}_t^*(s) : t \in \mathcal{T} \setminus \{T\}, s \in \mathcal{S}\}$.
- $\hat{\mathcal{A}}^-$: set of estimates of sub-optimal actions; $\hat{\mathcal{A}}^- \coloneqq \{\mathcal{A} \setminus \hat{\mathcal{A}}_t^*(s) : t \in \mathcal{T} \setminus \{T\}, s \in \mathcal{S}\}$.

Algorithm 1: Simulation-based backwards induction (SBBI) algorithm.

Input : $\mathcal{T}, \mathcal{S}, \mathcal{A}, N, f, r, v_T$.

for $t = T-1, T-2, \ldots, 0$ **do**
  for all $s \in \mathcal{S}$ **do**
    for all $a \in \mathcal{A}$ **do**
      Set $\boldsymbol{Q}_t(s,a) \leftarrow \emptyset$.
      for $n \leftarrow 1$ **to** N **do in parallel**
        Simulate $\omega^n \sim \mathcal{U}(0,1)$ and determine $s' = f_{t+1}(s,a,\omega^n)$.
        Compute $Q_t^n(s,a) = r_t(s,a,\omega^n) + \gamma \hat{v}_{t+1}(s')$.
        Update $\boldsymbol{Q}_t(s,a) \leftarrow \boldsymbol{Q}_t(s,a) \cup \{Q_t^n(s,a)\}$.
      end forpar
      Calculate $\hat{Q}_t(s,a)$ and $\hat{\sigma}_t^2(s,a)$.
      Set $\hat{\mathcal{A}}_t^*(s) \leftarrow \operatorname{argmax}_{a \in \mathcal{A}} \hat{Q}_t(s,a)$ and $\hat{\mathcal{A}}_t^-(s) \leftarrow \mathcal{A} \setminus \hat{\mathcal{A}}_t^*(s)$ for all $s \in \mathcal{S}$.
    end for
  end for
end for

Output: Q, $\hat{Q}$, $\hat{\mathcal{A}}^*$, $\hat{\mathcal{A}}^-$.

We now give additional details on the SBMCC algorithm included as Algorithm 2. The inputs for the algorithm are the following:

- $(\mathcal{T}, \mathcal{S}, \mathcal{A})$: finite sets of decision epochs, states, and actions.
- C: set of controls across all decision epochs and states; $C := \{C_t(s) : t \in \mathcal{T} \setminus \{T\}, s \in \mathcal{S}\}$.
- M: number of batches or independent MDP simulations; $M \in \mathbb{N}_+$.
- $\bar{\boldsymbol{Q}}$: set of action-value function estimates per batch by decision epoch, state, and action; $\bar{\boldsymbol{Q}} := \{\bar{Q}_t^m(s,a) : m \in \{1, \ldots, M\}, t \in \mathcal{T} \setminus \{T\}, s \in \mathcal{S}, a \in \mathcal{A}\}$.
- $\hat{\boldsymbol{Q}}$: set of action-value-function estimates across all decision epochs, states, and actions.

The output of the SBMCC algorithm is the estimate of $d_t(s,\alpha)$ for each state s at decision epoch t, denoted by $\hat{d}_t(s,\alpha)$. Combining algorithms 1 and 2 we produce Algorithm 3. This algorithm is used in our case study.

Algorithm 2: Simulation-based multiple comparisons with a control (SBMCC) algorithm.

Input : $\mathcal{T}$, $\mathcal{S}$, $\mathcal{A}$, C, M, $\bar{Q}$, $\hat{Q}$.

for all $t \in \mathcal{T} \setminus T$ **do**
  Set $\hat{\Psi}_t(s) \leftarrow \emptyset$ for all $s \in \mathcal{S}$.
  for $m \leftarrow 1$ **to** M **do in parallel**
    for all $s \in \mathcal{S}$ **do**
      for all $a \in \mathcal{A}$ and $a^* \in C_t(s)$ **do**
        Calculate $\bar{\psi}_t^m(s, a)$.
      end for
      Update $\hat{\Psi}_t(s) \leftarrow \hat{\Psi}_t(s) \cup \{\max_{a \in \mathcal{A}} \bar{\psi}_t^m(s, a)\}$.
    end for
  end forpar
  Let $\hat{d}_t(s, \alpha)$ be the empirical $1 - \alpha$ quantile of $\hat{\Psi}_t(s)$ for all $s \in \mathcal{S}$.
end for

Output: $\hat{d}_t(s, \alpha)$ for all $s \in \mathcal{S}$ and $t \in \mathcal{T} \setminus \{T\}$.

Algorithm 3: Combined SBBI and SBMCC algorithm.

Input : $\mathcal{T}$, $\mathcal{S}$, $\mathcal{A}$, C, N, M, f, r, $\boldsymbol{v}_T$.

for $t = T-1, T-2, \ldots, 0$ **do**
  for all $s \in \mathcal{S}$ **do**
    for all $a \in \mathcal{A}$ **do**
      Set $\boldsymbol{Q}_t(s,a) \leftarrow \emptyset$.
      for $n \leftarrow 1$ **to** N **do in parallel**
        Simulate $\omega^n \sim \mathcal{U}(0,1)$ and determine $s' = f_{t+1}(s,a,\omega^n)$.
        Compute $Q_t^n(s,a) = r_t(s,a,\omega^n) + \gamma \hat{Q}_{t+1}(s', \tilde{a})$, for $\tilde{a} \in \Pi_t(s')$.
        Update $\boldsymbol{Q}_t(s,a) \leftarrow \boldsymbol{Q}_t(s,a) \cup \{Q_t^n(s,a)\}$.
      end forpar
      Calculate $\hat{Q}_t(s,a)$ and $\hat{\sigma}_t^2(s,a)$.
    end for
  end for
  Divide $\boldsymbol{Q}_t(s,a)$ into M batches of K observations each.
  Set $\hat{\Psi}_t(s) \leftarrow \emptyset$ for all $s \in \mathcal{S}$.
  for $m \leftarrow 1$ **to** M **do in parallel**
    for all $s \in \mathcal{S}$ **do**
      for all $a \in \mathcal{A}$ and $a^* \in C_t(s)$ **do**
        Calculate $\bar{Q}_t^m(s,a)$, $\bar{\sigma}_t^2(s,a,m)$, and $\bar{\psi}_t^m(s,a)$.
      end for
      Update $\hat{\Psi}_t(s) \leftarrow \hat{\Psi}_t(s) \cup \{\max_{a \in \mathcal{A}} \bar{\psi}_t^m(s,a)\}$.
    end for
  end forpar
  Compute $\hat{d}_t(s,\alpha)$ and estimate $\Pi_t(s)$ for all $s \in \mathcal{S}$.
end for

Output: $\hat{d}_t(s,\alpha)$ and $\Pi_t(s)$ for all $s \in \mathcal{S}$ and $t \in \mathcal{T} \setminus \{T\}$.

C.2 Proofs of Analytical Results

This section of Appendix C contains the proof of all our claims. For ease of reading, we have repeated the claims. In addition, we have separated claims with multiple parts with lower case Roman numerals. Since the output of our algorithms can be obtained through independent simulations or by dividing a single simulation into batches, we split the simulation output into M batches (or independent simulations) of K observations, for a total of $N = MK$ samples (see Section 4.6.2). We present our results in terms of M and K, unless otherwise noted.

C.2.1 Proofs of Section 4.6.1.1

Proposition 1 (Proposition 4.1.) *Under Assumption 4.1 it follows that:*

$$1 - \mathbb{P}(\hat{\mathcal{A}}_t^*(s) \subseteq \mathcal{A}_t^*(s)) \leq A \exp\left\{\frac{-N}{2\kappa_t^2}\right\},$$

with $\kappa_t \coloneqq \sum_{\tau=t}^{T} \gamma^{\tau-t} R_\tau$, *where* $R_t \coloneqq \max_{(s,a)\in\mathcal{S}\times\mathcal{A}} R_t(s,a)$.

The proof of this proposition depends on the following lemma.

Lemma C.1 *Let* $\hat{\theta}_t(s,a,a') \coloneqq \hat{Q}_t(s,a) - \hat{Q}_t(s,a')$ *for* $a, a' \in \mathcal{A}$. *Under Assumption 4.1, (i)* $\hat{Q}_t(s,a) \leq \kappa_t$ *and (ii)* $|\hat{\theta}_t(s,a,a')| \leq \kappa_t$.

Proof.

(i) We first show that $0 \leq \hat{Q}_t(s,a) \leq \kappa_t$. Because the rewards are non-negative by Assumption 4.1 it follows that $\hat{Q}_t(s,a) \geq 0$. The proof proceeds by backwards induction on t with $t = T$ as the base case. For $\hat{Q}_T(s,a) \coloneqq \hat{r}_T(s)$ we have:

$$\hat{Q}_T(s,a) = \frac{1}{MK} \sum_{m=1}^{M} \sum_{k=1}^{K} r_T(s, \omega^{m,k}).$$

Since there is always a positive probability of observing an outcome $\tilde{\omega}^{m,k} \in \Omega$ such that $r_T(s,\omega^{m,k}) \leq r_T(s,\tilde{\omega}^{m,k}) \leq R_T(s)$ for every m and k, it holds that:

$$\hat{Q}_T(s,a) \leq \frac{1}{MK} \sum_{m=1}^{M} \sum_{k=1}^{K} r_T(s, \tilde{\omega}^{m,k}).$$

But then again, there is always a positive probability of observing an outcome $\bar{\omega}^{m,k} \in \Omega$ such that $r_T(s, \tilde{\omega}^{m,k}) \leq r_T(s, \bar{\omega}^{m,k}) \leq R_T(s)$ for every m and k. Extending this logic until $r_T(s, \bar{\omega}^{m,k}) = R_T(s)$ for all m and k we get:

$$\hat{Q}_T(s,a) \leq \frac{1}{MK} \sum_{m=1}^{M} \sum_{k=1}^{K} R_T(s) = R_T(s) = \gamma^0 R_T(s) \leq \gamma^0 R_T,$$

showing that the claim is true for $t = T$. Suppose $\hat{Q}_{t+1}(s,a) \leq \sum_{\tau=t+1}^{T} \gamma^{\tau-(t+1)} R_\tau$ as the induction hypothesis. For $\hat{Q}_t(s,a)$ it follows that:

$$\hat{Q}_t(s,a) = \frac{1}{MK} \sum_{m=1}^{M} \sum_{k=1}^{K} r_t(s,a,\omega^{m,k}) + \gamma \hat{v}_{t+1}(s^{m,k}),$$

where $s^{m,k} = f_{t+1}(s,a,\omega^{m,k})$. By the same argument used to show the claim in the induction base we get that:

$$\hat{Q}_t(s,a) \leq R_t(s,a) + \frac{\gamma}{MK} \sum_{m=1}^{M} \sum_{k=1}^{K} \hat{v}_{t+1}(s^{m,k}).$$

Moreover, from the induction hypothesis it holds that:

$$\hat{v}_{t+1}(s^{m,k}) \leq \sum_{\tau=t+1}^{T} \gamma^{\tau-(t+1)} R_\tau$$

Therefore,

$$\begin{aligned}
\hat{Q}_t(s,a) &\leq R_t(s,a) + \gamma \sum_{\tau=t+1}^{T} \gamma^{\tau-(t+1)} R_\tau \\
&\leq \gamma^0 R_t + \sum_{\tau=t+1}^{T} \gamma^{\tau-t} R_\tau \\
&= \sum_{\tau=t}^{T} \gamma^{\tau-t} R_\tau,
\end{aligned}$$

completing the inductive step.

(ii) The claim that $|\hat{Q}_t(s,a) - \hat{Q}_t(s,a')| \leq \sum_{\tau=t}^{T} \gamma^{\tau-t} R_\tau$ follows immediately since $0 \leq \hat{Q}_t(s,a)$ by Assumption 4.1 and because $\hat{Q}_t(s,a) \leq \sum_{\tau=t}^{T} \gamma^{\tau-t} R_\tau$ by the above analysis. □

Proof of Proposition 4.1. For a fixed s it holds that:

$$\{\hat{\mathcal{A}}_t^*(s) \nsubseteq \mathcal{A}_t^*(s)\} = \bigcup_{a \in \mathcal{A} \setminus \mathcal{A}_t^*(s)} \bigcap_{a' \in \mathcal{A}} \left\{ \hat{Q}_t(s,a) \geq \hat{Q}_t(s,a') \right\}.$$

which implies:

$$\begin{aligned}
\mathbb{P}(\hat{\mathcal{A}}_t^*(s) \nsubseteq \mathcal{A}_t^*(s)) &\leq \sum_{a \in \mathcal{A} \setminus \mathcal{A}_t^*(s)} \mathbb{P}\left(\bigcap_{a' \in \mathcal{A}} \hat{Q}_t(s,a) \geq \hat{Q}_t(s,a') \right) \\
&\leq \sum_{a \in \mathcal{A} \setminus \mathcal{A}_t^*(s)} \mathbb{P}\left(\hat{Q}_t(s,a) \geq \hat{Q}_t(s,a') \right),
\end{aligned}$$

for any $a' \in \mathcal{A}$. Further, consider a mapping $g : \mathcal{A} \setminus \mathcal{A}_t^*(s) \mapsto \mathcal{A}$ such that $\mathbb{E}[\hat{Q}_t(s, g(a))] \geq \mathbb{E}[\hat{Q}_t(s,a)]$. Note that this mapping always exists because if $a \in \operatorname{argmax}_{\bar{a} \in \mathcal{A}} \mathbb{E}[\hat{Q}_t(s,\bar{a})]$ then we have that $g(a) = a$ so that $\mathbb{E}[\hat{Q}_t(s, g(a))] = \mathbb{E}[\hat{Q}_t(s,a)]$. If $a \notin \operatorname{argmax}_{\bar{a} \in \mathcal{A}} \mathbb{E}[\hat{Q}_t(s,\bar{a})]$ then we can find a $g(a)$ such that $\mathbb{E}[\hat{Q}_t(s, g(a))] > \mathbb{E}[\hat{Q}_t(s,a)]$. For each $a \in \mathcal{A} \setminus \mathcal{A}_t^*(s)$, define:

$$\vartheta_t^{m,k}(s,a) \coloneqq Q_t^{m,k}(s,a) - Q_t^{m,k}(s, g(a)),$$

and its corresponding average:

$$\hat{\vartheta}_t(s,a) \coloneqq \frac{1}{MK} \sum_{m=1}^{M} \sum_{k=1}^{K} \vartheta_t^{m,k}(s,a) = \hat{Q}_t(s,a) - \hat{Q}_t(s, g(a)).$$

Let

$$l'(s) \coloneqq \min_{a \in \mathcal{A} \setminus \mathcal{A}_t^*(s)} \hat{\vartheta}_t(s,a),$$

and

$$u'(s) \coloneqq \max_{a \in \mathcal{A} \setminus \mathcal{A}_t^*(s)} \hat{\vartheta}_t(s,a).$$

Note that $\mathbb{E}[\vartheta_t^{m,k}(s,a)] \leq 0$ for all a. Also, notice that $l'(s)$ and $u'(s)$ can be attained because $\mathcal{S}$ and $\mathcal{A}$ are finite and the rewards are bounded for all s, a, and t. Hence, it follows that:

$$\begin{aligned}
\mathbb{P}(\hat{\mathcal{A}}_t^*(s) \nsubseteq \mathcal{A}_t^*(s)) &\leq \sum_{a \in \mathcal{A} \setminus \mathcal{A}_t^*(s)} \mathbb{P}\left(\hat{\vartheta}_t(s,a) \geq 0 \right) \\
&\leq \sum_{a \in \mathcal{A} \setminus \mathcal{A}_t^*(s)} \mathbb{P}\left(e^{\tau \sum_m \sum_k \vartheta_t^{m,k}(s,a)} \geq 1 \right),
\end{aligned}$$

for $\tau \geq 0$. By Markov's inequality it holds that:

$$\begin{aligned}
1 - \mathbb{P}(\hat{\mathcal{A}}_t^*(s) \subseteq \mathcal{A}_t^*(s)) &\leq \sum_{a \in \mathcal{A} \setminus \mathcal{A}_t^*(s)} \mathbb{E}\Big[e^{\tau \sum_m \sum_k \vartheta_t^{m,k}(s,a)}\Big] \\
&= \sum_{a \in \mathcal{A} \setminus \mathcal{A}_t^*(s)} \mathbb{E}\Big[\prod_m \prod_k e^{\tau \vartheta_t^{m,k}(s,a)}\Big] \\
&= \sum_{a \in \mathcal{A} \setminus \mathcal{A}_t^*(s)} \prod_m \prod_k \mathbb{E}\Big[e^{\tau \vartheta_t^{m,k}(s,a)}\Big],
\end{aligned}$$

where the last equality holds by the independence between each $\vartheta_t^{m,k}(s,a)$ induced by the Markov property. Since $|u'(s) - l'(s)| \leq \kappa_t$, by Hoeffding's Lemma (Hoeffding, 1963, Lemma 1) we get:

$$\begin{aligned}
1 - \mathbb{P}(\hat{\mathcal{A}}_t^*(s) \subseteq \mathcal{A}_t^*(s)) &\leq \sum_{a \in \mathcal{A} \setminus \mathcal{A}_t^*(s)} \prod_m \prod_k e^{\tau \mathbb{E}[\vartheta_t^{m,k}(s,a)] + \frac{1}{2}\tau^2 \kappa_t^2} \\
&= \sum_{a \in \mathcal{A} \setminus \mathcal{A}_t^*(s)} \exp\left\{\tau \sum_m \sum_k \mathbb{E}[\vartheta_t^{m,k}(s,a)] + \frac{1}{2} MK \tau^2 \kappa_t^2\right\}.
\end{aligned}$$

Minimizing over $\tau \geq 0$ it follows that:

$$\begin{aligned}
1 - \mathbb{P}(\hat{\mathcal{A}}_t^*(s) \subseteq \mathcal{A}_t^*(s)) &\leq \min_{\tau \geq 0} \sum_{a \in \mathcal{A} \setminus \mathcal{A}_t^*(s)} \exp\left\{\tau \sum_m \sum_k \mathbb{E}[\vartheta_t^{m,k}(s,a)] + \frac{1}{2} MK \tau^2 \kappa_t^2\right\} \\
&= \sum_{a \in \mathcal{A} \setminus \mathcal{A}_t^*(s)} \exp\left\{\frac{-MK\left(\frac{1}{MK}\sum_m \sum_k \mathbb{E}[\vartheta_t^{m,k}(s,a)]\right)^2}{2\kappa_t^2}\right\}.
\end{aligned}$$

As $\left(\frac{1}{MK}\sum_m \sum_k \mathbb{E}[\vartheta_t^{m,k}(s,a)]\right)^2 > 0$, we get:

$$\begin{aligned}
1 - \mathbb{P}(\hat{\mathcal{A}}_t^*(s) \subseteq \mathcal{A}_t^*(s)) &\leq \sum_{a \in \mathcal{A} \setminus \mathcal{A}_t^*(s)} \exp\left\{\frac{-MK}{2\kappa_t^2}\right\} \\
&\leq |\mathcal{A} \setminus \mathcal{A}_t^*(s)| \exp\left\{\frac{-MK}{2\kappa_t^2}\right\} \\
&\leq A \exp\left\{\frac{-MK}{2\kappa_t^2}\right\}.
\end{aligned}$$

Lastly, since $N = MK$:

$$1 - \mathbb{P}(\hat{\mathcal{A}}_t^*(s) \subseteq \mathcal{A}_t^*(s)) \leq A \exp\left\{\frac{-N}{2\kappa_t^2}\right\}.$$

□

Proposition 2 (Proposition 4.2) *Suppose Assumption 4.1 holds. Then, for any $\beta \in (0,1)$ and a fixed sample size N satisfying:*

$$N \geq 2\kappa_t^2 \log(A/\beta),$$

it holds that:

$$\mathbb{P}(\hat{\mathcal{A}}_t^*(s) \subseteq \mathcal{A}_t^*(s)) \geq 1 - \beta.$$

Proof. From Proposition 4.1 we have that:

$$1 - \mathbb{P}(\hat{\mathcal{A}}_t^*(s) \subseteq \mathcal{A}_t^*(s)) \leq A \exp\left\{\frac{-N}{2\kappa_t^2}\right\}.$$

Setting the right-hand side of this inequality to less or equal than β we get:

$$A \exp\left\{\frac{-N}{2\kappa_t^2}\right\} \leq \beta.$$

Therefore,

$$N \geq 2\kappa_t^2 \log(A/\beta).$$

□

C.2.2 Proofs of Section 4.6.1.2

Theorem 1 (Theorem 4.1.) *Suppose Assumption 4.1 holds. Then, $\hat{Q}_t(s,a)$ converges to $Q_t(s,a)$ with probability 1 uniformly on $\mathcal{A}$.*

The proof for this theorem depends on the following lemmas.

Lemma C.2 *Let $(V_n)_{n\in\mathbb{N}}$ and $(W_n)_{n\in\mathbb{N}}$ be sequences of random variables, and let V and W be two other random variables. Suppose that V_n converges almost surely (a.s.) to V and that W_n converges a.s. to W. Then, (i) $V_n + W_n \overset{a.s.}{\to} V + W$ and (ii) $V_n W_n \overset{a.s.}{\to} VW$.*

Proof.

(i) We first show that $V_n + W_n \overset{a.s.}{\to} V + W$. By the triangle inequality, for any $\delta > 0$ we have that:

$$\begin{aligned}
\mathbb{P}\left(\lim_{n\to\infty} |V_n + W_n - (V+W)| \geq \delta\right) &\leq \mathbb{P}\left(\left\{\lim_{n\to\infty} |V_n - V| \geq \frac{\delta}{2}\right\} \bigcup \left\{\lim_{n\to\infty} |W_n - W| \geq \frac{\delta}{2}\right\}\right) \\
&\leq \mathbb{P}\left(\lim_{n\to\infty} |V_n - V| \geq \frac{\delta}{2}\right) + \mathbb{P}\left(\lim_{n\to\infty} |W_n - W| \geq \frac{\delta}{2}\right) \\
&= 0,
\end{aligned}$$

where the second inequality follows from the addition rule of probability and the equality follows because $V_n \xrightarrow{a.s.} V$ and $W_n \xrightarrow{a.s.} W$. This shows that $V_n + W_n \xrightarrow{a.s.} V + W$.

(ii) We now show that $V_n W_n \xrightarrow{a.s.} VW$. Since $V_n W_n - VW = (V_n - V)(W_n - W) + W(V_n - V) + V(W_n - W)$, for any $\delta > 0$ we get:

$$\begin{aligned}\mathbb{P}\left(\lim_{n\to\infty} |V_n W_n - VW| \geq \delta\right) &= \mathbb{P}\left(\lim_{n\to\infty} |(V_n - V)(W_n - W) + W(V_n - V) + V(W_n - W)| \geq \delta\right) \\ &\leq \mathbb{P}\left(\lim_{n\to\infty} |(V_n - V)(W_n - W)| \geq \frac{\delta}{2}\right) \\ &\quad + \mathbb{P}\left(\lim_{n\to\infty} |W(V_n - V) + V(W_n - W)| \geq \frac{\delta}{2}\right), \qquad \text{(C.1)}\end{aligned}$$

where the inequality follows from the triangle inequality and the the addition rule of probability. We first focus on the first term in the right-hand side of equation (C.1). As $|vw| \leq \frac{1}{2}v^2 + \frac{1}{2}w^2$ for all $v, w \in \mathbb{R}$ it holds that:

$$\begin{aligned}\mathbb{P}\left(\lim_{n\to\infty} |(V_n - V)(W_n - W)| \geq \frac{\delta}{2}\right) &\leq \mathbb{P}\left(\lim_{n\to\infty} \frac{1}{2}(V_n - V)^2 + \frac{1}{2}(W_n - W)^2 \geq \frac{\delta}{2}\right) \\ &\leq \mathbb{P}\left(\lim_{n\to\infty} \frac{1}{2}(V_n - V)^2 \geq \frac{\delta}{4}\right) + \mathbb{P}\left(\lim_{n\to\infty} \frac{1}{2}(W_n - W)^2 \geq \frac{\delta}{4}\right) \\ &= \mathbb{P}\left(\lim_{n\to\infty} |V_n - V| \geq \sqrt{\frac{\delta}{2}}\right) + \mathbb{P}\left(\lim_{n\to\infty} |W_n - W| \geq \sqrt{\frac{\delta}{2}}\right) \\ &= 0,\end{aligned}$$

where the last equality follows because $V_n \xrightarrow{a.s.} V$ and $W_n \xrightarrow{a.s.} W$. We now show that the second term in the right-hand side of equation (C.1) is also 0. By the triangle inequality, we have that:

$$\begin{aligned}\mathbb{P}\left(\lim_{n\to\infty} |W(V_n - V) + V(W_n - W)| \geq \frac{\delta}{2}\right) &\leq \mathbb{P}\left(\lim_{n\to\infty} |W(V_n - V)| \geq \frac{\delta}{4}\right) \\ &\quad + \mathbb{P}\left(\lim_{n\to\infty} |V(W_n - W)| \geq \frac{\delta}{4}\right).\end{aligned}$$

For $\mathbb{P}\left(\lim_{n\to\infty}|W(V_n - V)| \geq \frac{\delta}{4}\right)$ it holds that:

$$\begin{aligned}
\mathbb{P}\left(\lim_{n\to\infty}|W(V_n - V)| \geq \frac{\delta}{4}\right) &= \mathbb{P}\left(\left\{\lim_{n\to\infty}|W(V_n - V)| \geq \frac{\delta}{4}\right\} \bigcap \{|W| \leq \eta\}\right) \\
&\quad + \mathbb{P}\left(\left\{\lim_{n\to\infty}|W(V_n - V)| \geq \frac{\delta}{4}\right\} \bigcap \{|W| > \eta\}\right) \\
&\leq \mathbb{P}\left(\lim_{n\to\infty}\eta\,|V_n - V| \geq \frac{\delta}{4}\right) + \mathbb{P}\left(|W| > \eta\right) \\
&= \mathbb{P}\left(\lim_{n\to\infty}|V_n - V| \geq \frac{\delta}{4\eta}\right) + \mathbb{P}\left(|W| > \eta\right),
\end{aligned}$$

for any $\eta \geq 1$. Allowing $\eta \to \infty$ as $n \to \infty$, we get that $\mathbb{P}\left(\lim_{n\to\infty}|W(V_n - V)| \geq \frac{\delta}{4}\right) = 0$ because $V_n \xrightarrow{a.s.} V$ and $\mathbb{P}\left(|W| > \eta\right) = 0$. Repeating the above arguments for $\mathbb{P}\left(\lim_{n\to\infty}|V(W_n - W)| \geq \frac{\delta}{4}\right)$, it holds that:

$$\mathbb{P}\left(\lim_{n\to\infty}|W(V_n - V) + V(W_n - W)| \geq \frac{\delta}{2}\right) = 0.$$

Combining our results, we can conclude that:

$$\begin{aligned}
\mathbb{P}\left(\lim_{n\to\infty}|V_nW_n - VW| \geq \delta\right) &\leq \mathbb{P}\left(\lim_{n\to\infty}|(V_n - V)(W_n - W)| \geq \frac{\delta}{2}\right) \\
&\quad + \mathbb{P}\left(\lim_{n\to\infty}|W(V_n - V) + V(W_n - W)| \geq \frac{\delta}{2}\right) \\
&= 0,
\end{aligned}$$

and it follows that $V_nW_n \xrightarrow{a.s.} VW$. □

Lemma C.3 *Let $\bar{Q}_t(s,a)$ denote the estimate of $Q_t(s,a)$ generated with a single simulation of Algorithm 1 for state s and action a. Then, $\bar{Q}_t(s,a)$ converges to $Q_t(s,a)$ with probability 1 uniformly on $\mathcal{A}$.*

Proof. Suppose the immediate rewards $r_t(s,a,\omega)$ are bounded iid random variables from an unknown probability distribution at every iteration k at every replication m of the simulation. If the immediate rewards are constant (i.e. $r_t(s,a,\omega) = \mathbb{E}[r_t(s,a,\omega)|s,a]$) their convergence holds trivially. For simplicity of notation, we use s^k instead of $s^{m,k}$ and $r_t(s,a,\omega^k)$ instead of $r_t(s,a,\omega^{m,k})$, as we are interested in a single replication of simulation during this proof ($M = 1$).

We prove this result by backwards induction on t, starting with $t = T$ as the base case. Since $Q_T^k(s,a)$ are iid for all $k = 1,\ldots,K$ with finite mean and variance we have that

$\bar{Q}_T(s,a) \xrightarrow{a.s.} Q_T(s,a)$ by the strong law of large numbers (Casella and Beroer, 2001, Theorem 5.5.9). Furthermore, since $\mathcal{A}$ is a finite set we have that:

$$\begin{aligned}\sup_{a\in\mathcal{A}} |\bar{Q}_T(s,a) - Q_T(s,a)| &= \max_{a\in\mathcal{A}} |\bar{Q}_T(s,a) - Q_T(s,a)| \\ &\leq \sum_{a\in\mathcal{A}} |\bar{Q}_T(s,a) - Q_T(s,a)|.\end{aligned}$$

Hence, $\bar{Q}_T(s,a) \xrightarrow{a.s.} Q_T(s,a)$ implies almost sure convergence uniformly on $\mathcal{A}$.

Suppose that $\bar{Q}_{t+1}(s',a') \xrightarrow{a.s.} Q_{t+1}(s',a')$ uniformly on $\mathcal{A}$ for all s' as the induction hypothesis. Then, for $\bar{Q}_t(s,a)$ it follows that:

$$\begin{aligned}\bar{Q}_t(s,a) &= \frac{1}{K}\sum_k Q_t^k(s,a) \\ &= \frac{1}{K}\sum_k r_t(s,a,\omega^k) + \gamma \max_{a'\in\mathcal{A}} \bar{Q}_{t+1}(s^k,a') \\ &= \bar{r}_t(s,a) + \frac{\gamma}{K}\sum_k \max_{a'\in\mathcal{A}} \bar{Q}_{t+1}(s^k,a'),\end{aligned}$$

where $\bar{r}_t(s,a) \coloneqq \frac{1}{K}\sum_k r_t(s,a,\omega^k)$.

To show that $\bar{Q}_t(s,a) \xrightarrow{a.s.} Q_t(s,a)$ uniformly on $\mathcal{A}$ for all $s \in \mathcal{S}$, notice that:

$$\begin{aligned}\sup_{a\in\mathcal{A}} |\bar{Q}_t(s,a) - Q_t(s,a)| &= \max_{a\in\mathcal{A}} |\bar{Q}_t(s,a) - Q_t(s,a)| \\ &\leq \sum_{a\in\mathcal{A}} |\bar{Q}_t(s,a) - Q_t(s,a)| \\ &= \sum_{a\in\mathcal{A}} \Bigg| \bar{r}_t(s,a) + \frac{\gamma}{K}\sum_{k=1}^{K} \max_{a'\in\mathcal{A}} \bar{Q}_{t+1}(s^k,a') \\ &\qquad - \mathbb{E}\Big[r_t(s,a,\omega) + \gamma \max_{a'\in\mathcal{A}} Q_{t+1}(s',a')\Big|s,a\Big]\Bigg|.\end{aligned}$$

Therefore, the almost sure convergence of $\bar{Q}_t(s,a)$ to $Q_t(s,a)$ for all a implies almost sure convergence uniformly on $\mathcal{A}$.

Since $r_t(s,a,\omega^k)$ are bounded iid random variables, we have that $\mathbb{E}[r_t(s,a,\omega)|s,a]$ is well-defined. Consequently, we can deduce that $\bar{r}_t(s,a) \xrightarrow{a.s.} \mathbb{E}[r_t(s,a,\omega)|s,a]$ by the strong law of large numbers. Because $\mathcal{S}$ is a finite set and $|r_t(s,a,\omega^k)| < \infty$, we have that $\mathbb{E}[\max_{a'\in\mathcal{A}} Q_{t+1}(s',a')|s,a]$ is also well-defined.

Now, note that:

$$\frac{1}{K}\sum_k \max_{a'\in\mathcal{A}} \bar{Q}_{t+1}(s^k, a') = \frac{1}{K}\sum_k \sum_{s'} \mathbb{1}\{s^k = s'|s,a\} \max_{a'\in\mathcal{A}} \bar{Q}_{t+1}(s^k, a').$$

For a fixed state s' we have that:

$$\frac{1}{K}\sum_k \mathbb{1}\{s^k = s'|s,a\} \max_{a'\in\mathcal{A}} \bar{Q}_{t+1}(s', a') = \max_{a'\in\mathcal{A}} \bar{Q}_{t+1}(s', a') K^{-1} \sum_k \mathbb{1}\{s^k = s'|s,a\}.$$

From the induction hypothesis, we can deduce that:

$$\max_{a'\in\mathcal{A}} \bar{Q}_{t+1}(s', a') \xrightarrow{a.s.} \max_{a'\in\mathcal{A}} Q_{t+1}(s', a'),$$

for each s'. Furthermore, by the strong law of large numbers we get that:

$$K^{-1}\sum_k \mathbb{1}\{s^k = s'|s,a\} \xrightarrow{a.s.} p_t(s'|s,a),$$

for every s'. Hence, by Lemma C.2 it follows that:

$$\frac{1}{K}\sum_k \mathbb{1}\{s^k = s'|s,a\} \max_{a'\in\mathcal{A}} \bar{Q}_{t+1}(s', a') \xrightarrow{a.s.} p_t(s'|s,a) \max_{a'\in\mathcal{A}} Q_{t+1}(s', a'). \tag{C.2}$$

Adding over all states, we conclude that:

$$\begin{aligned}\frac{1}{K}\sum_k \sum_{s'} \mathbb{1}\{s^k = s'|s,a\} \max_{a'\in\mathcal{A}} \bar{Q}_{t+1}(s', a') &\xrightarrow{a.s.} \sum_{s'} p_t(s'|s,a) \max_{a'\in\mathcal{A}} Q_{t+1}(s', a') \\ &= \mathbb{E}\Big[\max_{a'\in\mathcal{A}} Q_{t+1}(s', a')\Big| s, a\Big].\end{aligned}$$

Since $\bar{r}_t(s,a) \xrightarrow{a.s.} \mathbb{E}[r_t(s,a,\omega)|s,a]$ and

$$\gamma K^{-1}\sum_k \max_{a'\in\mathcal{A}} \bar{Q}_{t+1}(s^k, a') \xrightarrow{a.s.} \gamma\mathbb{E}[\max_{a'\in\mathcal{A}} Q_{t+1}(s', a')|s,a],$$

by Lemma C.2 it holds that:

$$\bar{r}_t(s,a) + \gamma K^{-1}\sum_k \max_{a'\in\mathcal{A}} \bar{Q}_{t+1}(s^k, a') \xrightarrow{a.s.} \mathbb{E}[r_t(s,a,\omega) + \gamma \max_{a'\in\mathcal{A}} Q_{t+1}(s', a')|s,a].$$

Therefore, we get that $\bar{Q}_t(s,a) \xrightarrow{a.s.} Q_t(s,a)$ for all a and s. Because $\max_{a\in\mathcal{A}} |\bar{Q}_t(s,a) - Q_t(s,a)| \leq \sum_{a\in\mathcal{A}} |\bar{Q}_t(s,a) - Q_t(s,a)|$, the a.s. convergence of $\bar{Q}_t(s,a)$ for all a and s implies that $\bar{Q}_t(s,a) \xrightarrow{a.s.}$

$Q_t(s,a)$ uniformly on $\mathcal{A}$ for all s. This completes the inductive step and the proof. □

Proof of Theorem 4.1. From Lemma C.3, we have that $\bar{Q}_t^m(s,a) \xrightarrow{a.s.} Q_t(s,a)$ uniformly on $\mathcal{A}$ for all s and each m. Since

$$\hat{Q}_t(s,a) = \frac{1}{M}\sum_{m=1}^{M} \bar{Q}_t^m(s,a),$$

for each state-action pair (s,a), it holds that $\hat{Q}_t(s,a) \xrightarrow{a.s.} Q_t(s,a)$ by Lemma C.2. Furthermore, since $\mathcal{A}$ is a finite set we have that:

$$\begin{aligned}\sup_{a\in\mathcal{A}} |\hat{Q}_t(s,a) - Q_t(s,a)| &= \max_{a\in\mathcal{A}} |\hat{Q}_t(s,a) - Q_t(s,a)| \\ &\leq \sum_{a\in\mathcal{A}} |\hat{Q}_t(s,a) - Q_t(s,a)|.\end{aligned}$$

Thus, the almost sure convergence of $\hat{Q}_t(s,a)$ to $Q_t(s,a)$ for all s and a implies almost sure convergence uniformly on $\mathcal{A}$ for all s. □

Corollary 1 (Corollary. 4.1) *Suppose Assumption 4.1 holds. Then, (i) $\hat{v}_t(s)$ converges to $v_t(s)$ and (ii) $\hat{\mathcal{A}}_t^*(s) \subseteq \mathcal{A}_t^*(s)$ with probability 1 for N large enough.*

Proof.

(i) From Theorem 4.1, we have that $\max_{a\in\mathcal{A}} |\hat{Q}_t(s,a) - Q_t(s,a)| \xrightarrow{a.s.} 0$ as $N \to \infty$. Thus, it holds that $\hat{v}_t(s) \xrightarrow{a.s.} v_t(s)$ for all s and t.

(ii) The proof is a contrapositive argument. Let

$$\max_{a\in\mathcal{A}} |\hat{Q}_t(s,a) - Q_t(s,a)| \leq \delta_N, \tag{C.3}$$

for $\delta_N > 0$ and

$$\varrho(s) \coloneqq v_t(s) - \max_{a\in\mathcal{A}_t^-(s)} Q_t(s,a). \tag{C.4}$$

Since for any $a \in \mathcal{A}_t^-(s)$ we have that $v_t(s) > Q_t(s,a)$ and $\mathcal{A}_t^-(s)$ is finite it holds that $\varrho(s) > 0$. Let N be large enough such that $\varrho(s)/2 > \delta_N$. From (C.3) we have that $\hat{v}_t(s) \geq v_t(s) - \delta_N$, which implies that $\hat{v}_t(s) > v_t(s) - \varrho(s)/2$. From (C.4) we get:

$$\begin{aligned} v_t(s) - Q_t(s,a) &\geq \varrho(s) \\ v_t(s) &\geq Q_t(s,a) + \varrho(s) \\ v_t(s) + \delta_N &\geq Q_t(s,a) + \delta_N + \varrho(s), \end{aligned}$$

for any $a \in \mathcal{A}_t^-(s)$. Moreover, since $Q_t(s,a) + \delta_N \geq \hat{Q}_t(s,a)$ by equation (C.3) it holds

that:

$$\begin{aligned} v_t(s) + \delta_N &\geq \hat{Q}_t(s,a) + \varrho(s) \\ v_t(s) + \varrho(s)/2 &> \hat{Q}_t(s,a) + \varrho(s) \\ v_t(s) - \varrho(s)/2 &> \hat{Q}_t(s,a), \end{aligned}$$

where the strict inequality follows because $\varrho(s)/2 > \delta_N$. Hence, $\hat{v}_t(s) > v_t(s) - \varrho(s)/2 > \hat{Q}_t(s,a)$ and $a \notin \mathcal{A}_t^*(s)$ implies $a \notin \hat{\mathcal{A}}_t^*(s)$. The inclusion $\hat{\mathcal{A}}_t^*(s) \subseteq \mathcal{A}_t^*(s)$ follows. □

C.2.3 Proofs of Section 4.6.1.3

Proposition 3 (Proposition 4.3) *Under assumptions 4.1 and 4.2, (i) $\hat{Q}_t(s,a)$ and (ii) $\hat{v}_t(s)$ are ϵ-nonincreasing in s with probability 1 for N large enough.*

The proof of this proposition relies on the following lemma.

Lemma C.4 *Let X and Y be partially ordered finite sets, $g : X \mapsto \mathbb{R}$ be a monotonic function of X, and $h : X \times Y \mapsto [0,1]$ be a monotonic function of Y satisfying:*

$$\sum_{x' \geq x} h(x', y) \leq \sum_{x' \geq x} h(x', \bar{y}), \tag{C.5}$$

for $y \leq \bar{y}$ with $\sum_{x \in X} h(x,y) = \sum_{x \in X} h(x,\bar{y})$. Then, we have:

$$\sum_{x \in X} h(x,y) g(x) \leq \sum_{x \in X} h(x,\bar{y}) g(x),$$

when g is nondecreasing in x and

$$\sum_{x \in X} h(x,y) g(x) \geq \sum_{x \in X} h(x,\bar{y}) g(x),$$

when g is nonincreasing in x.

Proof of Lemma C.4. We first prove that the claim holds in the nondecreasing case. By definition, we have that:

$$\sum_{x \in X} h(x,y) = \sum_{x \in X} h(x,\bar{y}).$$

Thus,

$$
\begin{aligned}
0 &= \sum_{x\in X} h(x,y) - \sum_{x\in X} h(x,\bar{y}) \qquad \text{(C.6)}\\
&= \left[\sum_{x\in X} h(x,y) - \sum_{x\in X} h(x,\bar{y})\right] g(x^1)\\
&\geq h(x^1,y)g(x^1) - h(x^1,\bar{y})g(x^1) + \left[\sum_{x\in X\setminus\{x^1\}} h(x,y) - h(x,\bar{y})\right] g(x^2)\\
&\geq \sum_{x'\in\{x^1,x^2\}} h(x',y)g(x') - h(x',\bar{y})g(x') + \left[\sum_{x\in X\setminus\{x^1,x^2\}} h(x,y) - h(x,\bar{y})\right] g(x^3),
\end{aligned}
$$

where $x^1 \leq x^2 \leq x^3 \in X$. The inequalities above follow by equation (C.5) and because g is nondecreasing in x. Continuing with this pattern we get:

$$
0 \geq \sum_{x'\in X} h(x',y)g(x') - h(x',\bar{y})g(x'),
$$

which implies that:

$$
\sum_{x'\in X} h(x',\bar{y})g(x') \geq \sum_{x'\in X} h(x',y)g(x'),
$$

completing the proof for the nondecreasing case. The proof for the nonincreasing case follows by multiplying (C.6) by -1. See the proof of Lemma 1 in Chhatwal et al. (2010) for the special case that $h(x,y) = p_t(s'|s,a)$ and $p_t(s'|s,a)$ is nonincreasing in t. □

Proof of Proposition 4.3.

(i) We first show that $\hat{v}_t(s)$ is ϵ-nonincreasing in s for N large enough. By Proposition 4.7.4 in Puterman (2014), we have that Assumption 4.2 implies that $v_t(s)$ is nonincreasing in s. Moreover, from Corollary 4.1 we have that $\hat{v}_t(s) \xrightarrow{a.s.} v_t(s)$ for all s. Thus, for any $\epsilon > 0$ there is an $N^* \in \mathbb{N}$ such that for any $N \geq N^*$ it holds that $\hat{v}_t(\tilde{s}) \leq \hat{v}_t(s) + \epsilon$ for $s \leq \tilde{s}$. It follows that $\hat{v}_t(s)$ is ϵ-nonincreasing in s with probability 1 for N large enough.

(ii) We now show that if $Q_t(s,a)$ is nonincreasing in s, then $\hat{Q}_t(s,a)$ is ϵ-nonincreasing with probability 1 for N large enough. If $v_t(s)$ is nonincreasing in s we can deduce

that for any $t \in \mathcal{T} \setminus \{T\}$:

$$\begin{aligned} Q_t(s,a) &= \mathbb{E}[r_t(s,a,\omega)|s,a] + \gamma \sum_{s' \in \mathcal{S}} p_t(s'|s,a) v_{t+1}(s') \\ &\geq \mathbb{E}[r_t(\tilde{s},a,\omega)|\tilde{s},a] + \gamma \sum_{s' \in \mathcal{S}} p_t(s'|s,a) v_{t+1}(s') \\ &\geq \mathbb{E}[r_t(\tilde{s},a,\omega)|\tilde{s},a] + \gamma \sum_{s' \in \mathcal{S}} p_t(s'|\tilde{s},a) v_{t+1}(s') \\ &= Q_{t+1}(\tilde{s},a), \end{aligned}$$

for $s \leq \tilde{s}$. The first inequality follows because $\mathbb{E}[r_t(s,a,\omega)|s,a]$ is nonincreasing in s, and the second inequality follows by Lemma C.4 because $\bar{p}_t(s'|s,a)$ is nondecreasing in s and the nonincreasing behavior of $v_t(s)$ in s. Note that $\mathbb{E}[r_T(s,\omega)|s] \geq \mathbb{E}[r_T(\tilde{s},\omega)|\tilde{s}]$ for $s \leq \tilde{s}$ by Assumption 4.2. Since from Theorem 4.1 we have that $\hat{Q}_t(s,a) \xrightarrow{a.s.} Q_t(s,a)$ for all s and a, it holds that $\hat{Q}_t(s,a)$ is ϵ-nonincreasing in s with probability 1 for N large enough. □

Proposition 4 (Proposition 4.4.) *Suppose assumptions 4.1 and 4.2 hold. Then, (i) $\hat{Q}_t(s,a)$ and (ii) $\hat{v}_t(s)$ are ϵ-nonincreasing in t with probability 1 for N large enough.*

Proof.

(i) We first show this result when $\mathbb{E}[r_t(s,a,\omega)|s,a]$ and $p_t(s'|s,a)$ are known for all s', s, a and t. The proof for $v_t(s)$ proceeds by backwards induction on t, starting with $t = T-1$ as the base case. For $v_{T-1}(s)$ we have:

$$\begin{aligned} v_{T-1}(s) &= \max_{a \in \mathcal{A}} \left\{ \mathbb{E}[r_{T-1}(s,a,\omega)|s,a] + \gamma \sum_{s' \in \mathcal{S}} p_{T-1}(s'|s,a) v_T(s') \right\} \\ &\geq \mathbb{E}[r_{T-1}(s,a,\omega)|s,a] \\ &\geq \mathbb{E}[r_T(s,\omega)|s] \\ &= v_T(s). \end{aligned} \tag{C.7}$$

Suppose that $v_{t+1}(s) \geq v_{t+2}(s)$ as the induction hypothesis. Then, for $v_t(s)$ it follows

that:

$$\begin{aligned}
v_t(s) &= \max_{a\in\mathcal{A}}\left\{\mathbb{E}[r_t(s,a,\omega)|s,a] + \gamma\sum_{s'\in\mathcal{S}} p_t(s'|s,a)v_{t+1}(s')\right\} \\
&\geq \max_{a\in\mathcal{A}}\left\{\mathbb{E}[r_{t+1}(s,a,\omega)|s,a] + \gamma\sum_{s'\in\mathcal{S}} p_t(s'|s,a)v_{t+1}(s')\right\} \\
&\geq \max_{a\in\mathcal{A}}\left\{\mathbb{E}[r_{t+1}(s,a,\omega)|s,a] + \gamma\sum_{s'\in\mathcal{S}} p_{t+1}(s'|s,a)v_{t+1}(s')\right\} \\
&\geq \max_{a\in\mathcal{A}}\left\{\mathbb{E}[r_{t+1}(s,a,\omega)|s,a] + \gamma\sum_{s'\in\mathcal{S}} p_{t+1}(s'|s,a)v_{t+2}(s')\right\} \\
&= v_{t+1}(s),
\end{aligned}$$

where the first inequality because $\mathbb{E}[r_t(s,a,\omega)|s,a]$ is nonincreasing t due to Assumption 4.2. The second inequality holds from the nonincreasing behavior of $v_t(s)$ in s by Assumption 4.2 and Proposition 4.3, and applying Lemma C.4 with $h(x,y) = p_t(s'|s,a)$, $h(x,\bar{y}) = p_{t+1}(s'|s,a)$, and $g(x) = v_{t+1}(s')$. The last inequality follows from the induction hypothesis.

(ii) The proof for $Q_t(s,a)$ follows as a consequence of the nonincreasing behavior of $v_t(s)$ in t. For any $t \in \mathcal{T}\setminus\{T-1,T\}$ we then have:

$$\begin{aligned}
Q_t(s,a) &= \mathbb{E}[r_t(s,a,\omega)|s,a] + \gamma\sum_{s'\in\mathcal{S}} p_t(s'|s,a)v_{t+1}(s') \\
&\geq \mathbb{E}[r_{t+1}(s,a,\omega)|s,a] + \gamma\sum_{s'\in\mathcal{S}} p_t(s'|s,a)v_{t+1}(s') \\
&\geq \mathbb{E}[r_{t+1}(s,a,\omega)|s,a] + \gamma\sum_{s'\in\mathcal{S}} p_{t+1}(s'|s,a)v_{t+1}(s') \\
&\geq \mathbb{E}[r_{t+1}(s,a,\omega)|s,a] + \gamma\sum_{s'\in\mathcal{S}} p_{t+1}(s'|s,a)v_{t+2}(s') \\
&= Q_{t+1}(s,a),
\end{aligned}$$

where the first inequality follows because $\mathbb{E}[r_t(s,a,\omega)|s,a]$ is nonincreasing t, the second inequality holds because $\bar{p}_t(s'|s,a)$ is nondecreasing in t and the nonincreasing behavior of $v_t(s)$ in s, and the last inequality follows from the nonincreasing behavior of $v_t(s)$ in t. Note that $Q_{T-1}(s,a) \geq Q_T(s,a)$ by Assumption 4.2.

From Theorem 4.1 we have that $\hat{Q}_t(s,a) \xrightarrow{a.s.} Q_t(s,a)$ uniformly on $\mathcal{A}$ and from Corollary 4.1 we get that $\hat{v}_t(s) \xrightarrow{a.s.} v_t(s)$ for all s. Consequently, for any $\epsilon > 0$ we can find

an $N^* \in \mathbb{N}$ such that for any $N \geq N^*$ it holds that $\hat{Q}_t(s,a) \geq \hat{Q}_{t+1}(s,a) - \epsilon$ and $\hat{v}_t(s) \geq \hat{v}_{t+1}(s) - \epsilon$. Hence, if $Q_t(s,a)$ and $v_t(s)$ are nonincreasing in t, then $\hat{Q}_t(s,a)$ and $\hat{v}_t(s)$ are ϵ-nonincreasing with probability 1 for N large enough. □

C.2.4 Proofs of Section 4.6.2.1

Proposition 5 (Proposition 4.5.) *Suppose Assumption 4.1 holds. Then, for any $\alpha \in (0,1)$ we have that:*

$$\mathbb{P}\Big(Q_t(s,a^*) - Q_t(s,a) \in \Theta : \hat{\mathbb{H}}_t\left(\max_{a\in\mathcal{A}}\{\hat{\psi}_t(s,a)\}, \hat{\mathbb{F}}_t(s)\right) \leq 1-\alpha\Big) = 1-\alpha$$

for N large enough, $a^* \in C_t(s)$, and all $a \in \mathcal{A}$.

The proof of this proposition depends on the following two lemmas.

Lemma C.5 *Under Assumption 4.1, (i) $\bar{\sigma}_t^2(s,a) \xrightarrow{a.s.} \sigma_t^2(s,a)$, (ii) $\hat{\sigma}_t^2(s,a) \xrightarrow{a.s.} \sigma_t^2(s,a)$, and (iii) $K\hat{\zeta}_t^2(s,a) \xrightarrow{a.s.} \sigma_t^2(s,a)$.*

Proof.

(i) For simplicity of notation we drop the superscript m in $Q_t^{m,k}(s,a)$ as the batch number m is arbitrary, but fixed. Let $\bar{v}_t^2(s,a) \coloneqq \frac{K-1}{K}\bar{\sigma}_t^2(s,a)$. For $\bar{v}_t^2(s,a)$ we have:

$$\begin{aligned}\bar{v}_t^2(s,a) &= \frac{1}{K}\sum_{k=1}^{K}\left[Q_t^k(s,a) - \bar{Q}_t(s,a)\right]^2 \\ &= \frac{1}{K}\sum_{k=1}^{K}\left[Q_t^k(s,a)\right]^2 - \left[\bar{Q}_t(s,a)\right]^2.\end{aligned}$$

From Theorem 4.1 we have that $\bar{Q}_t(s,a) \xrightarrow{a.s.} Q_t(s,a)$ and by Lemma C.2 it follows that $\left[\bar{Q}_t(s,a)\right]^2 \xrightarrow{a.s.} Q_t^2(s,a)$. It remains to show that:

$$\frac{1}{K}\sum_{k=1}^{K}\left[Q_t^k(s,a)\right]^2 \xrightarrow{a.s.} \mathbb{E}[(r_t(s,a,\omega) + \gamma v_{t+1}(f_{t+1}(s,a,\omega)))^2 | s,a].$$

For $\frac{1}{K}\sum_{k=1}^{K}\left[Q_t^k(s,a)\right]^2$ it holds that:

$$\begin{aligned}\frac{1}{K}\sum_{k=1}^{K}\left[Q_t^k(s,a)\right]^2 &= \frac{1}{K}\sum_{k=1}^{K}\left[r_t(s,a,\omega^k) + \gamma\hat{v}_{t+1}(s^k)\right]^2 \\ &= \frac{1}{K}\sum_{k=1}^{K} r_t^2(s,a,\omega^k) + \gamma r_t(s,a,\omega^k)\hat{v}_{t+1}(s^k) + \gamma^2\hat{v}_{t+1}^2(s^k),\end{aligned}$$

where $s^k = f_{t+1}(s, a, \omega^k)$. Since $r_t(s, a, \omega^k)$ are iid, so are $r_t^2(s, a, \omega^k)$ and we get that:

$$\frac{1}{K}\sum_{k=1}^{K} r_t(s, a, \omega^k) \xrightarrow{a.s.} \mathbb{E}[r_t(s, a, \omega)|s, a] \text{ and } \frac{1}{K}\sum_{k=1}^{K} r_t^2(s, a, \omega^k) \xrightarrow{a.s.} \mathbb{E}[r_t^2(s, a, \omega)|s, a],$$

by the strong law of large numbers. Further, from equation (C.2) it follows that:

$$\frac{1}{K}\sum_{k=1}^{K} \hat{v}_{t+1}(s^k) \xrightarrow{a.s.} \mathbb{E}[v_{t+1}(s')|s, a],$$

and we can deduce that:

$$\frac{1}{K}\sum_{k=1}^{K} r_t(s, a, \omega^k)\hat{v}_{t+1}(s^k) \xrightarrow{a.s.} \mathbb{E}[r_t(s, a, \omega)|s, a]\mathbb{E}[v_{t+1}(s')|s, a],$$

by the strong law of large numbers and Lemma C.2. Lastly, notice that:

$$\frac{1}{K}\sum_{k=1}^{K} \hat{v}_{t+1}^2(s^k) = \frac{1}{K}\sum_{k=1}^{K}\sum_{s'} \mathbb{1}\{s^k = s'|s, a\}\hat{v}_{t+1}^2(s').$$

For a fixed state s' we have:

$$\frac{1}{K}\sum_{k=1}^{K} \mathbb{1}\{s^k = s'|s, a\}\hat{v}_{t+1}^2(s') \xrightarrow{a.s.} p_t(s'|s, a)v_{t+1}^2(s'),$$

also by the strong law of large numbers and Lemma C.2. Adding over all the states:

$$\begin{aligned}\frac{1}{K}\sum_{k=1}^{K}\sum_{s'} \mathbb{1}\{s^k = s'|s, a\}\hat{v}_{t+1}(s') &\xrightarrow{a.s.} \sum_{s'} p_t(s'|s, a)v_{t+1}^2(s') \\ &= \mathbb{E}[v_{t+1}^2(s')|s, a].\end{aligned}$$

Therefore,

$$\frac{1}{K}\sum_{k=1}^{K}\left[Q_t^k(s, a)\right]^2 \xrightarrow{a.s.} \mathbb{E}[r_t^2(s, a, \omega)|s, a] + 2\gamma\mathbb{E}[r_t(s, a, \omega)|s, a]\mathbb{E}[v_{t+1}(s')|s, a] + \gamma\mathbb{E}\left[v_{t+1}^2(s')|s, a\right]. \tag{C.8}$$

Moreover, since

$$\begin{aligned}
&\mathbb{E}[r_t^2(s,a,\omega)|s,a] + 2\gamma\mathbb{E}[r_t(s,a,\omega)|s,a]\mathbb{E}[v_{t+1}(s')|s,a] + \mathbb{E}\left[v_{t+1}^2(s')|s,a\right] \\
&= \mathbb{E}\left[r_t^2(s,a,\omega) + 2\gamma r_t(s,a,\omega)v_{t+1}(f_{t+1}(s,a,\omega)) + \gamma v_{t+1}^2(f_{t+1}(s,a,\omega))|s,a\right] \\
&= \mathbb{E}[(r_t(s,a,\omega) + \gamma v_{t+1}(f_{t+1}(s,a,\omega)))^2|s,a] \qquad \text{(C.9)}
\end{aligned}$$

we have that:

$$\bar{v}_t^2(s,a) \xrightarrow{a.s.} \mathbb{E}[(r_t(s,a,\omega) + \gamma v_{t+1}(f_{t+1}(s,a,\omega)))^2|s,a] - \mathbb{E}[r_t(s,a,\omega) + \gamma v_{t+1}(f_{t+1}(s,a,\omega))|s,a]^2.$$

As $\bar{\sigma}_t^2(s,a) = \frac{K}{K-1}\bar{v}_t^2(s,a)$ and $\frac{K}{K-1} \to 1$ as $K \to \infty$, we can conclude that $\bar{\sigma}_t^2(s,a) \xrightarrow{a.s.} \sigma_t^2(s,a)$.

(ii) This result is a direct consequence of (i) as $\bar{\sigma}_t^2(s,a)$ is equivalent to $\hat{\sigma}_t^2(s,a)$ if we replace K by N.

(iii) Let $\hat{v}_t^2(s,a) \coloneqq \frac{M-1}{M}\hat{\zeta}_t^2(s,a)$. For $\hat{v}_t^2(s,a)$ we have:

$$\begin{aligned}
\hat{v}_t^2(s,a) &= \frac{1}{M}\sum_{m=1}^{M}\left[\bar{Q}_t^m(s,a) - \hat{Q}_t(s,a)\right]^2 \\
&= \frac{1}{M}\sum_{m=1}^{M}\left[\bar{Q}_t^m(s,a)\right]^2 - \left[\hat{Q}_t(s,a)\right]^2.
\end{aligned}$$

From Theorem 4.1 it holds that $\hat{Q}_t(s,a) \xrightarrow{a.s.} Q_t(s,a)$ and by Lemma C.2 it follows that

$\hat{Q}_t^2(s,a) \xrightarrow{a.s.} Q_t^2(s,a)$. For $\frac{1}{M}\sum_{m=1}^{M}\left[\bar{Q}_t^m(s,a)\right]^2$ we have:

$$\begin{aligned}
\frac{1}{M}\sum_{m=1}^{M}\left[\bar{Q}_t^m(s,a)\right]^2 &= \frac{1}{M}\sum_{m=1}^{M}\left[\frac{1}{K}\sum_{k=1}^{K} r_t(s,a,\omega^{m,k}) + \gamma\hat{v}_{t+1}(s^{m,k})\right]^2 \\
&= \frac{1}{M}\sum_{m=1}^{M}\left[\frac{1}{K}\sum_{k=1}^{K} r_t(s,a,\omega^{m,k}) + \frac{\gamma}{K}\sum_{k=1}^{K}\hat{v}_{t+1}(s^{m,k})\right]^2 \\
&= \frac{1}{MK^2}\sum_{m=1}^{M}\left[\sum_{k=1}^{K} r_t(s,a,\omega^{m,k})\right]^2 + 2\gamma\left[\sum_{k=1}^{K} r_t(s,a,\omega^{m,k})\right]\left[\sum_{k=1}^{K}\hat{v}_{t+1}(s^{m,k})\right] \\
&\quad + \gamma^2\left[\sum_{k=1}^{K}\hat{v}_{t+1}(s^{m,k})\right]^2 \\
&= \frac{1}{MK^2}\sum_{m=1}^{M}\sum_{k=1}^{K}\sum_{l\neq k} r_t(s,a,\omega^{m,k})r_t(s,a,\omega^{m,l}) + 2\gamma r_t(s,a,\omega^{m,k})\hat{v}_{t+1}(s^{m,l}) \\
&\quad + \gamma^2\hat{v}_{t+1}(s^{m,k})\hat{v}_{t+1}(s^{m,l}) \\
&\quad + \frac{1}{MK^2}\sum_{m=1}^{M}\sum_{k=1}^{K} r_t^2(s,a,\omega^{m,k}) + 2\gamma r_t(s,a,\omega^{m,k})\hat{v}_{t+1}(s^{m,k}) + \gamma^2\hat{v}_{t+1}^2(s^{m,k}),
\end{aligned}$$

where $s^{m,k} = f_{t+1}(s,a,\omega^{m,k})$. Moreover, since k is independent of l for every $k \neq l$ it holds that:

$$\begin{aligned}
\frac{1}{M}\sum_{m=1}^{M}\left[\bar{Q}_t^m(s,a)\right]^2 &= \frac{1}{MK^2}\sum_{m=1}^{M} K(K-1)\bar{r}_t^m(s,a)\bar{r}_t^m(s,a) \\
&\quad + 2\gamma K(K-1)\bar{r}_t^m(s,a)\frac{1}{K-1}\sum_{l=1}^{K}\hat{v}_{t+1}(s^{m,l}) \\
&\quad + \gamma^2 K(K-1)\frac{1}{K}\sum_{k=1}^{K}\hat{v}_{t+1}(s^{m,k})\frac{1}{K-1}\sum_{l=1}^{K}\hat{v}_{t+1}(s^{m,l}) \\
&\quad + \frac{1}{MK^2}\sum_{m=1}^{M}\sum_{k=1}^{K} r_t^2(s,a,\omega^{m,k}) + 2\gamma r_t(s,a,\omega^{m,k})\hat{v}_{t+1}(s^{m,k}) + \gamma^2\hat{v}_{t+1}^2(s^{m,k})
\end{aligned}$$

where $\bar{r}_t^m(s,a) \coloneqq \frac{1}{K}\sum_{k=1}^{K} r_t(s,a,\omega^{m,k})$. By the strong law of large numbers and Lemma C.2 we can conclude that:

$$\frac{1}{M}\sum_{m=1}^{M}\bar{r}_t^m(s,a) \xrightarrow{a.s.} \mathbb{E}[r_t(s,a,\omega)|s,a], \quad \frac{1}{M}\sum_{m=1}^{M}\left[\bar{r}_t^m(s,a)\right]^2 \xrightarrow{a.s.} \mathbb{E}[r_t(s,a,\omega)|s,a]^2,$$

and

$$\frac{1}{MK}\sum_{m=1}^{M}\sum_{k=1}^{K} r_t^2(s,a,\omega^{m,k}) \xrightarrow{a.s.} \mathbb{E}[r_t^2(s,a,\omega)|s,a].$$

Furthermore, we have that:

$$\frac{1}{MK}\sum_{m=1}^{M}\sum_{k=1}^{K} \hat{v}_{t+1}(s^{m,k}) = \frac{1}{MK}\sum_{m=1}^{M}\sum_{k=1}^{K}\sum_{s'} \mathbb{1}\{s^{m,k}=s'|s,a\}\hat{v}_{t+1}(s'),$$

and it holds that:

$$\begin{aligned}\frac{1}{MK}\sum_{m=1}^{M}\sum_{k=1}^{K}\sum_{s'} \mathbb{1}\{s^{m,k}=s'|s,a\}\hat{v}_{t+1}(s') &\xrightarrow{a.s.} \sum_{s'} p_t(s'|s,a)v_{t+1}(s') \\ &= \mathbb{E}[v_{t+1}(s')|s,a],\end{aligned}$$

where $s' = f_{t+1}(s,a,\omega)$. Thus,

$$\begin{aligned}\frac{1}{M}\sum_{m=1}^{M}[\bar{Q}_t^m(s,a)]^2 \xrightarrow{a.s.} &\mathbb{E}[r_t(s,a,\omega)|s,a]^2 - \frac{1}{K}\mathbb{E}[r_t(s,a,\omega)|s,a]^2 \\ &+ 2\gamma\mathbb{E}[r_t(s,a,\omega)|s,a]\mathbb{E}[v_{t+1}(s')|s,a] - \frac{2\gamma}{K}\mathbb{E}[r_t(s,a,\omega)|s,a]\mathbb{E}[v_{t+1}(s')|s,a] \\ &+ \gamma^2\mathbb{E}[v_{t+1}(s')|s,a]^2 - \frac{\gamma^2}{K}\mathbb{E}[v_{t+1}(s')|s,a]^2 \\ &+ \frac{1}{K}\mathbb{E}[r_t^2(s,a,\omega)|s,a] + \frac{2\gamma}{K}\mathbb{E}[r_t(s,a,\omega)|s,a]\mathbb{E}[v_{t+1}(s')|s,a] \\ &+ \frac{\gamma^2}{K}\mathbb{E}[v_{t+1}^2(s')|s,a].\end{aligned}$$

Combining this results with

$$\hat{Q}_t(s,a) \xrightarrow{a.s.} Q_t(s,a) = \mathbb{E}[r_t(s,a,\omega) + \gamma v_{t+1}(s')|s,a] = \mathbb{E}[r_t(s,a,\omega)|s,a] + \gamma\mathbb{E}[v_{t+1}(s')|s,a]$$

we get:

$$
\begin{aligned}
\hat{v}_t^2(s,a) \xrightarrow{a.s.} & \mathbb{E}[r_t(s,a,\omega)|s,a]^2 - \frac{1}{K}\mathbb{E}[r_t(s,a,\omega)|s,a]^2 \\
& + 2\gamma\mathbb{E}[r_t(s,a,\omega)|s,a]\mathbb{E}[v_{t+1}(s')|s,a] - \frac{2\gamma}{K}\mathbb{E}[r_t(s,a,\omega)|s,a]\mathbb{E}[v_{t+1}(s')|s,a] \\
& + \gamma^2\mathbb{E}[v_{t+1}(s')|s,a]^2 - \frac{\gamma^2}{K}\mathbb{E}[v_{t+1}(s')|s,a]^2 \\
& + \frac{1}{K}\mathbb{E}[r_t^2(s,a,\omega)|s,a] + \frac{2\gamma}{K}\mathbb{E}[r_t(s,a,\omega)|s,a]\mathbb{E}[v_{t+1}(s')|s,a] + \frac{\gamma^2}{K}\mathbb{E}[v_{t+1}^2(s')|s,a] \\
& - \left(\mathbb{E}[r_t(s,a,\omega)|s,a] + \gamma\mathbb{E}[v_{t+1}(s')|s,a]\right)^2 \\
= & \mathbb{E}[r_t(s,a,\omega)|s,a]^2 - \frac{1}{K}\mathbb{E}[r_t(s,a,\omega)|s,a]^2 \\
& + 2\gamma\mathbb{E}[r_t(s,a,\omega)|s,a]\mathbb{E}[v_{t+1}(s')|s,a] - \frac{2\gamma}{K}\mathbb{E}[r_t(s,a,\omega)|s,a]\mathbb{E}[v_{t+1}(s')|s,a] \\
& + \gamma^2\mathbb{E}[v_{t+1}(s')|s,a]^2 - \frac{\gamma^2}{K}\mathbb{E}[v_{t+1}(s')|s,a]^2 \\
& + \frac{1}{K}\mathbb{E}[r_t^2(s,a,\omega)|s,a] + \frac{2\gamma}{K}\mathbb{E}[r_t(s,a,\omega)|s,a]\mathbb{E}[v_{t+1}(s')|s,a] + \frac{\gamma^2}{K}\mathbb{E}[v_{t+1}^2(s')|s,a] \\
& - \mathbb{E}[r_t(s,a,\omega)|s,a]^2 - 2\gamma\mathbb{E}[r_t(s,a,\omega)|s,a]\mathbb{E}[v_{t+1}(s')|s,a] - \gamma^2\mathbb{E}[v_{t+1}(s')|s,a]^2 \\
= & \frac{1}{K}\mathbb{E}[r_t^2(s,a,\omega)|s,a] + \frac{2\gamma}{K}\mathbb{E}[r_t(s,a,\omega)|s,a]\mathbb{E}[v_{t+1}(s')|s,a] + \frac{\gamma^2}{K}\mathbb{E}[v_{t+1}^2(s')|s,a] \\
& - \frac{1}{K}\mathbb{E}[r_t(s,a,\omega)|s,a]^2 - \frac{2\gamma}{K}\mathbb{E}[r_t(s,a,\omega)|s,a]\mathbb{E}[v_{t+1}(s')|s,a] - \frac{\gamma^2}{K}\mathbb{E}[v_{t+1}(s')|s,a]^2.
\end{aligned}
$$

Therefore,

$$
K\hat{v}_t^2(s,a) \xrightarrow{a.s.} \mathbb{E}[(r_t(s,a,\omega) + \gamma v_{t+1}(s'))^2|s,a] - \mathbb{E}[r_t(s,a,\omega) + \gamma v_{t+1}(s')|s,a]^2.
$$

Because $\hat{\zeta}_t^2(s,a) = \frac{M}{M-1}\hat{v}_t^2(s,a)$ and $\frac{M}{M-1} \to 1$ as $M \to \infty$, it holds that $K\hat{\zeta}_t^2(s,a) \xrightarrow{a.s.} \sigma_t^2(s,a)$. □

Lemma C.6 *Under Assumption 4.1 it holds that* $\sup_{x\in\mathbb{R}} |\hat{\mathbb{H}}_t(x, \hat{\mathbb{F}}_t(s)) - \mathbb{H}_t(x, \mathbb{F}_t(s))| \xrightarrow{a.s.} 0$.

Proof. We first show that $\sup_{x\in\mathbb{R}} |\hat{\mathbb{F}}_t(x,s) - \mathbb{F}_t(x,s)| \xrightarrow{a.s.} 0$. Let

$$
Y_{m,k}(x) \coloneqq \mathbb{1}\{Q_t^{m,k}(s,a) \le x\} - \mathbb{F}_t(x,s,a),
$$

for fixed s, a, and t. Note that $\mathbb{E}[Y_{m,k}(x)] = 0$ for any s, a, and t. Moreover, define:

$$
E_{M,K}(\varepsilon) \coloneqq \left\{\omega \in \mathbb{R} : \left|\frac{1}{MK}\sum_m\sum_k Y_{m,k}^{\omega}(x)\right| > \varepsilon\right\}.
$$

We want to show that:

$$\lim_{M\to\infty}\lim_{K\to\infty}\frac{1}{MK}\sum_m\sum_k Y^{\omega}_{m,k}(x)=0,$$

for almost every $\omega\in\mathbb{R}$. By Markov's inequality (Casella and Beroer, 2001, Lemma 3.8.3), it follows that:

$$\begin{aligned}\mathbb{P}\left(\left|\sum_m\sum_k Y_{m,k}(x)\right|>\varepsilon MK\right)&=\mathbb{P}\left(\left|\sum_m\sum_k Y_{m,k}(x)\right|^4>\varepsilon^4M^4N^4\right)\\&\leq\frac{\mathbb{E}\left[\sum_m\sum_k Y_{m,k}(x)\right]^4}{\varepsilon^4M^4N^4}.\end{aligned}$$

Expanding $\mathbb{E}\left[\sum_m\sum_k Y_{m,k}(x)\right]^4$, it can be noticed that only the terms with $\mathbb{E}[Y^4_{m,k}(x)]$ and $\mathbb{E}[Y^2_{m,k}(x)Y^2_{m',k'}(x)]$ for $k\neq k'$ and $m\neq m'$ are non-zero. Furthermore, we have that $\mathbb{E}[Y^2_{m,k}(x)Y^2_{m',k'}(x)]=\mathbb{E}[Y^2_{m,k}(x)]\mathbb{E}[Y^2_{m',k'}(x)]$ by the Markov property. Thus, we get the following bound:

$$\mathbb{P}\left(\left|\sum_m\sum_k Y_{m,k}(x)\right|\right)\leq\frac{\sum_m\sum_k\mathbb{E}[Y^4_{m,k}(x)]+\sum_m\sum_{m'\neq m}\sum_k\sum_{k'\neq k}\mathbb{E}[Y^2_{m,k}(x)]\mathbb{E}[Y^2_{m',k'}(x)]}{\varepsilon^4M^4N^4}.$$

But since $\mathbb{F}_t(\cdot,s,a)$ is a true cdf and $\mathbb{1}\{Q^{m,k}_t(s,a)\leq x\}\in\{0,1\}$, it holds that $Y^2_{m,k}(x)\in[0,1]$ and $Y^4_{m,k}(x)\in[0,1]$, and we have:

$$\mathbb{P}\left(\left|\sum_m\sum_k Y_{m,k}(x)\right|\right)\leq\frac{2}{\varepsilon^4K^2M^2},$$

and it follows that $\sum_m\sum_k\mathbb{P}(E_{M,K}(\varepsilon))<\infty$. By the first Borel-Cantelli Lemma (Billingsley, 1995, Theorem 4.3), we get that the probability that $E_{M,K}(\varepsilon)$ happens infinitely often (i.o.) for any $\varepsilon>0$ is zero, which implies that $\hat{\mathbb{F}}_t(x,s,a)\xrightarrow{a.s.}\mathbb{F}_t(x,s,a)$. Since $\hat{\mathbb{F}}_t(x,s,a)\xrightarrow{a.s.}\mathbb{F}_t(x,s,a)$ independently of $x\in\mathbb{R}$, we have that:

$$\sup_{x\in\mathbb{R}}|\hat{\mathbb{F}}_t(x,s,a)-\mathbb{F}_t(x,s,a)|\xrightarrow{a.s.}0.$$

Because $Q_t(s,a)$ is independent from $Q_t(s,a')$ for $a\neq a'$, it follows that:

$$\mathbb{F}_t(\boldsymbol{x},s)=\prod_{a\in\mathcal{A}}\mathbb{F}_t(x_a,s,a)\quad\text{and}\quad\hat{\mathbb{F}}_t(\boldsymbol{x},s,a)=\prod_{a\in\mathcal{A}}\hat{\mathbb{F}}_t(x_a,s,a),$$

for $\boldsymbol{x}=(x_1,x_2,\ldots,x_A)\in\mathbb{R}^A$. Since $\hat{\mathbb{F}}_t(x,s,a)\xrightarrow{a.s.}\mathbb{F}_t(x,s,a)$ uniformly on $x\in\mathbb{R}$ we also have

that $\hat{\mathbb{F}}_t(\boldsymbol{x}, s) \xrightarrow{a.s.} \mathbb{F}_t(\boldsymbol{x}, s)$ uniformly on $\boldsymbol{x} \in \mathbb{R}^A$. We now show that $\lim_{M\to\infty}\lim_{K\to\infty} \hat{\mathbb{H}}_t(\cdot, \hat{\mathbb{F}}_t(s)) = \lim_{M\to\infty}\lim_{K\to\infty} \mathbb{H}_t(\cdot, \mathbb{F}_t(s))$. First, notice that:

$$\begin{aligned}
\mathbb{H}_t(x, \mathbb{F}_t(s)) &= \mathbb{P}\left(\max_{a\in\mathcal{A}}\{\psi_t(s,a)\} \le x\right) \\
&= \prod_{a\in\mathcal{A}} \mathbb{P}\left(\psi_t(s,a) \le x\right) \\
&= \prod_{a\in\mathcal{A}} \mathbb{P}\left(\frac{\hat{Q}_t(s,a^*) - \hat{Q}_t(s,a) - (Q_t(s,a^*) - Q_t(s,a))}{\sqrt{N^{-1}\left[\hat{\sigma}_t^2(s,a^*) + \hat{\sigma}_t^2(s,a)\right]}} \le x\right),
\end{aligned}$$

where the second equality holds because $\psi_t(s,a)$ is independent from $\psi_t(s,a')$ for $a \neq a'$. Since $\hat{Q}_t(s,a) \xrightarrow{a.s.} Q_t(s,a)$ from Theorem 4.1 and $\hat{\sigma}_t^2(s,a) \xrightarrow{a.s.} \sigma_t^2(s,a)$ from Lemma C.5, by the Central Limit Theorem (Casella and Beroer, 2001, Theorem 5.5.15) we have that:

$$\lim_{N\to\infty} \mathbb{P}\left(\frac{\hat{Q}_t(s,a^*) - \hat{Q}_t(s,a) - (Q_t(s,a^*) - Q_t(s,a))}{\sqrt{N^{-1}\left[\hat{\sigma}_t^2(s,a^*) + \hat{\sigma}_t^2(s,a)\right]}} \le x\right) = \Phi(x),$$

where $\Phi(\cdot)$ is the cdf of a standard normal random variable. Thus, $\lim_{N\to\infty} \mathbb{H}_t(x, \mathbb{F}_t(s)) = \Phi(x)^A$.

Second, note that:

$$\begin{aligned}
\hat{\mathbb{H}}_t(x, \hat{\mathbb{F}}_t(s)) &= \mathbb{P}\left(\max_{a\in\mathcal{A}}\{\hat{\psi}_t(s,a)\} \le x\right) \\
&= \prod_{a\in\mathcal{A}} \mathbb{P}\left(\hat{\psi}_t(s,a) \le x\right) \\
&= \prod_{a\in\mathcal{A}} \mathbb{P}\left(\frac{\hat{Q}_t(s,a^*) - \hat{Q}_t(s,a) - (Q_t(s,a^*) - Q_t(s,a))}{\sqrt{M^{-1}\left[\hat{\zeta}_t^2(s,a^*) + \hat{\zeta}_t^2(s,a)\right]}} \le x\right),
\end{aligned}$$

where the second equality holds because $\hat{\psi}_t(s,a)$ is independent from $\hat{\psi}_t(s,a')$ for $a \neq a'$. Since $\hat{Q}_t(s,a) \xrightarrow{a.s.} Q_t(s,a)$ from Theorem 4.1 and $K\hat{\zeta}_t^2(s,a) \xrightarrow{a.s.} \sigma_t^2(s,a)$ from Lemma C.5, applying the Central Limit Theorem once again we get:

$$\lim_{M\to\infty}\lim_{K\to\infty} \mathbb{P}\left(\frac{\hat{Q}_t(s,a^*) - \hat{Q}_t(s,a) - (Q_t(s,a^*) - Q_t(s,a))}{\sqrt{\frac{K}{MK}\left[\hat{\zeta}_t^2(s,a^*) + \hat{\zeta}_t^2(s,a)\right]}} \le x\right) = \Phi(x),$$

where $\Phi(\cdot)$ is the cdf of a standard normal random variable. Hence, $\lim_{M\to\infty}\lim_{K\to\infty} \hat{\mathbb{H}}_t(x, \hat{\mathbb{F}}_t(s)) =$

$\lim_{N\to\infty} \mathbb{I}\hat{\mathrm{H}}_t(x, \hat{\mathbb{F}}_t(s)) = \Phi(x)^A$.

Therefore, we can deduce that $\lim_{N\to\infty} \mathbb{I}\hat{\mathrm{H}}_t(\cdot, \hat{\mathbb{F}}_t(s)) = \lim_{N\to\infty} \mathbb{H}_t(\cdot, \mathbb{F}_t(s))$. Moreover, by Pólya's Theorem (Serfling, 1980, Theorem 1.5.3) it follows that:

$$\lim_{N\to\infty} \sup_{x\in\mathbb{R}} \left| \mathbb{I}\hat{\mathrm{H}}_t(x, \hat{\mathbb{F}}_t(s)) - \mathbb{H}_t(x, \mathbb{F}_t(s)) \right| = 0,$$

because of the continuity of $\lim_{N\to\infty} \mathbb{H}_t(x, \mathbb{F}_t(s)) = \Phi(x)^A$ for any $x \in \mathbb{R}$. Combining this result with the previous conclusion that $\hat{\mathbb{F}}_t(\boldsymbol{x}, s) \xrightarrow{a.s.} \mathbb{F}_t(\boldsymbol{x}, s)$ uniformly on $\boldsymbol{x} \in \mathbb{R}^A$, we get that:

$$\sup_{x\in\mathbb{R}} \left| \mathbb{I}\hat{\mathrm{H}}_t(x, \hat{\mathbb{F}}_t(s)) - \mathbb{H}_t(x, \mathbb{F}_t(s)) \right| \xrightarrow{a.s.} 0.$$

□

Proof of Proposition 4.5. From Lemma C.6 we get that:

$$\sup_{\tau\in\mathbb{R}} |\mathbb{I}\hat{\mathrm{H}}_t(\tau, \hat{\mathbb{F}}_t(s)) - \mathbb{H}_t(\tau, \mathbb{F}_t(s))| \xrightarrow{a.s.} 0.$$

Since a.s. convergence implies convergence in distribution, it follows that:

$$\sup_{\tau\in\mathbb{R}} |\mathbb{I}\hat{\mathrm{H}}_t(\tau, \hat{\mathbb{F}}_t(s)) - \mathbb{H}_t(\tau, \mathbb{F}_t(s))| \xrightarrow{\mathcal{D}} 0.$$

Because $\mathbb{H}_t(\cdot, \mathbb{F}_t(s))$ is a true cdf, it is uniformly distributed on $[0, 1]$. Therefore, $\mathbb{I}\hat{\mathrm{H}}_t(\cdot, \hat{\mathbb{F}}_t(s))$ must also follow a $\mathcal{U}(0, 1)$ distribution asymptotically and we get that:

$$\mathbb{P}(\mathbb{I}\hat{\mathrm{H}}_t(\cdot, \hat{\mathbb{F}}_t(s)) \leq 1 - \alpha) = \mathbb{P}(\mathcal{U}(0, 1) \leq 1 - \alpha) = 1 - \alpha,$$

for N large enough. □

Theorem 2 (Theorem 4.2.) *Under Assumption 4.1 we have that:*

$$\Pi_t(s, \alpha) = \left\{ a \in \mathcal{A} : \hat{Q}_t(s, a^*) - \hat{Q}_t(s, a) \leq \hat{d}_t(s, \alpha) \sqrt{M^{-1} \left[\hat{\zeta}_t^2(s, a^*) + \hat{\zeta}_t^2(s, a) \right]} \right\}$$

with $\hat{d}_t(s, \alpha) = \mathbb{I}\hat{\mathrm{H}}_t^{-1}(1 - \alpha, \hat{\mathbb{F}}_t(s))$ *is a set of* α*-nonsignificant actions with probability 1 for* N *large enough and* $a^* \in C_t(s)$.

Proof. From Proposition 4.5 we have that:

$$\mathbb{P}\Big(Q_t(s, a^*) - Q_t(s, a) \in \Theta : \mathbb{I}\hat{\mathrm{H}}_t \left(\max_{a\in\mathcal{A}} \{\hat{\psi}_t(s, a)\}, \hat{\mathbb{F}}_t(s) \right) \leq 1 - \alpha \Big) = 1 - \alpha,$$

for N large enough. Thus, it follows that:

$$
\begin{aligned}
&\mathbb{P}\Big(Q_t(s,a^*) - Q_t(s,a) \in \Theta : \hat{\mathbb{H}}_t\left(\max_{a\in\mathcal{A}}\{\hat{\psi}_t(s,a)\}, \hat{\mathbb{F}}_t(s)\right) \leq 1-\alpha\Big) \\
&= \mathbb{P}\Big(\hat{\mathbb{H}}_t\left(\max_{a\in\mathcal{A}}\{\hat{\psi}_t(s,a)\}, \hat{\mathbb{F}}_t(s)\right) \leq 1-\alpha\Big) \\
&= \mathbb{P}\Big(\max_{a\in\mathcal{A}}\{\hat{\psi}_t(s,a)\} \leq \hat{\mathbb{H}}_t^{-1}(1-\alpha, \hat{\mathbb{F}}_t(s))\Big) \\
&= \mathbb{P}\Big(\max_{a\in\mathcal{A}}\{\hat{\psi}_t(s,a)\} \leq \hat{d}_t(s,\alpha)\Big) \\
&= \mathbb{P}\Big(\hat{\psi}_t(s,a^1) \leq \hat{d}_t(s,\alpha), \ldots, \hat{\psi}_t(s,a^A) \leq \hat{d}_t(s,\alpha)\Big) \\
&= \mathbb{P}\Big(\hat{Q}_t(s,a^*) - \hat{Q}_t(s,a^1) - \hat{d}_t(s,\alpha)\sqrt{M^{-1}\left[\hat{\zeta}_t(s,a^*) + \hat{\zeta}_t^2(s,a)\right]} \leq Q_t(s,a^*) - Q_t(s,a^1), \\
&\ldots, \hat{Q}_t(s,a^*) - \hat{Q}_t(s,a^A) - \hat{d}_t(s,\alpha)\sqrt{M^{-1}\left[\hat{\zeta}_t(s,a^*) + \hat{\zeta}_t^2(s,a)\right]} \leq Q_t(s,a^*) - Q_t(s,a^A)\Big) \\
&= 1-\alpha,
\end{aligned}
$$

where $a^1, \ldots, a^A \in \mathcal{A}$ for N large enough. Moreover, from Lemma C.6 it follows that $\hat{\mathbb{H}}_t(1-\alpha, \hat{\mathbb{F}}_t(s)) \xrightarrow{a.s.} \mathbb{H}_t(1-\alpha, \mathbb{F}_t(s))$ and hence, $\hat{d}_t(s,\alpha) = \hat{\mathbb{H}}_t^{-1}(1-\alpha, \hat{\mathbb{F}}_t(s)) \xrightarrow{a.s.} \mathbb{H}_t^{-1}(1-\alpha, \mathbb{F}_t(s)) = d_t(s,\alpha)$ by Theorem 2.3.1 in Serfling (1980). Consequently, the asymptotic confidence intervals simultaneously contain $Q_t(s,a^*) - Q_t(s,a^1), \ldots, Q_t(s,a^*) - Q_t(s,a^A)$ with probability exactly $1-\alpha$. Furthermore, under the null hypothesis (i.e. all actions have the same performance), any action a such that $\hat{Q}_t(s,a^*) - \hat{Q}_t(s,a) \leq d_t(s,\alpha)\sqrt{M^{-1}\left[\hat{\zeta}_t(s,a^*) + \hat{\zeta}_t^2(s,a)\right]}$ is not statistically significant from a^* at state s. □

Proposition 6 (Proposition 4.6.) *Suppose that Assumption 4.1 is satisfied. Then,*

$$
\lim_{M\to\infty}\lim_{K\to\infty} \sqrt{MK}\sup_{x\in\mathbb{R}}\left|\hat{\mathbb{H}}_t(x, \hat{\mathbb{F}}_t(s)) - \mathbb{H}_t(x, \mathbb{F}_t(s))\right| \leq CA^{5/4}\sqrt{2\kappa_t^3},
$$

where C is the constant appearing in the multivariate Berry-Esseen bound.

The proof of this proposition depends on the following lemma.

Lemma C.7 *Under Assumption 4.1 we have that* $\lim_{M\to\infty}\lim_{K\to\infty} \mathbb{E}[Q_t^{m,k}(s,a)] = Q_t(s,a)$ *and* $\lim_{M\to\infty}\lim_{K\to\infty} Var\left(Q_t^{m,k}(s,a)\right) = \sigma_t(s,a)$ *for all s, a, and t.*

Proof. We first show that $\lim_{M\to\infty}\lim_{K\to\infty} \mathbb{E}[Q_t^{m,k}(s,a)] = Q_t(s,a)$. Due to the assumption of bounded rewards, we have that $\hat{Q}_t(s,a)$ are also bounded for all s and a. Since by Theorem 4.1 we have that $\hat{Q}_t(s,a) \xrightarrow{a.s.} Q_t(s,a)$, from the Bounded Convergence Theorem (Billingsley, 1995, Theorem 16.5) it holds that $\lim_{M\to\infty}\lim_{K\to\infty} \mathbb{E}[\hat{Q}_t(s,a)] = Q_t(s,a)$. Moreover, as $Q_t^{m,k}(s,a)$ are

iid random variables we get that:

$$\begin{aligned}\lim_{M\to\infty}\lim_{K\to\infty}\mathbb{E}[\hat{Q}_t(s,a)] &= \lim_{M\to\infty}\lim_{K\to\infty}\mathbb{E}\left[\frac{1}{MK}\sum_m\sum_k Q_t^{m,k}(s,a)\right]\\ &= \lim_{M\to\infty}\lim_{K\to\infty}\frac{1}{MK}\sum_m\sum_k\mathbb{E}\left[Q_t^{m,k}(s,a)\right]\\ &= \lim_{M\to\infty}\lim_{K\to\infty}\frac{MK}{MK}\mathbb{E}\left[Q_t^{m,k}(s,a)\right]\\ &= Q_t(s,a).\end{aligned}$$

Thus, $\lim_{M\to\infty}\lim_{K\to\infty}\mathbb{E}[Q_t^{m,k}(s,a)] = Q_t(s,a)$. We now show that $\lim_{M\to\infty}\lim_{K\to\infty}\mathrm{Var}\left(Q_t^{m,k}(s,a)\right) = \sigma_t(s,a)$. By definition, we have that:

$$\begin{aligned}\lim_{M\to\infty}\lim_{K\to\infty}\mathrm{Var}\left(Q_t^{m,k}(s,a)\right) &= \lim_{M\to\infty}\lim_{K\to\infty}\mathbb{E}\left[(Q_t^{m,k}(s,a))^2\right] - \left(\mathbb{E}[Q_t^{m,k}(s,a)]\right)^2\\ &= \lim_{M\to\infty}\lim_{K\to\infty}\mathbb{E}\left[(Q_t^{m,k}(s,a))^2\right] - Q_t^2(s,a).\end{aligned}$$

From equations (C.8) and (C.9) it can be deduced that:

$$\frac{1}{MK}\sum_m\sum_k\left(Q_t^{m,k}(s,a)\right)^2 \xrightarrow{a.s.} \mathbb{E}[(r_t(s,a,\omega)+\gamma v_{t+1}(f_{t+1}(s,a,\omega)))^2|s,a].$$

Since $Q_t^{m,k}(s,a)$ are bounded random variables, so are $(Q_t^{m,k}(s,a))^2$. Applying the Bounded Convergence Theorem to $\frac{1}{MK}\sum_m\sum_k\left(Q_t^{m,k}(s,a)\right)^2$ we get that:

$$\lim_{M\to\infty}\lim_{K\to\infty}\mathbb{E}\left[\frac{1}{MK}\sum_m\sum_k\left(Q_t^{m,k}(s,a)\right)^2\right] = \mathbb{E}[(r_t(s,a,\omega)+\gamma v_{t+1}(f_{t+1}(s,a,\omega)))^2|s,a].$$

Furthermore, as $(Q_t^{m,k}(s,a))^2$ are iid random variables it follows that:

$$\lim_{M\to\infty}\lim_{K\to\infty}\mathbb{E}\left[\left(Q_t^{m,k}(s,a)\right)^2\right] = \mathbb{E}[(r_t(s,a,\omega)+\gamma v_{t+1}(f_{t+1}(s,a,\omega)))^2|s,a]$$

and we can conclude that:

$$\begin{aligned}\lim_{M\to\infty}\lim_{K\to\infty}\mathrm{Var}\left(Q_t^{m,k}(s,a)\right) &= \mathbb{E}[(r_t(s,a,\omega)+\gamma v_{t+1}(f_{t+1}(s,a,\omega)))^2|s,a]\\ &\quad - \mathbb{E}[r_t(s,a,\omega)+\gamma v_{t+1}(f_{t+1}(s,a,\omega))|s,a]^2\\ &= \sigma_t(s,a).\end{aligned}$$

Proof of Proposition 4.6. Let $\Phi_A(x)$ denote the standard normal cdf in $\mathbb{R}^A$. By the triangle inequality we have that:

$$\begin{aligned}\lim_{N\to\infty}\sup_{x\in\mathbb{R}}\left|\hat{\mathbb{H}}_t(x,\hat{\mathbb{F}}_t(s))-\mathbb{H}_t(x,\mathbb{F}_t(s))\right| &= \lim_{N\to\infty}\sup_{x\in\mathbb{R}}\left|\hat{\mathbb{H}}_t(x,\hat{\mathbb{F}}_t(s))-\Phi_A(x)+\Phi_A(x)-\mathbb{H}_t(x,\mathbb{F}_t(s))\right| \\ &\leq \lim_{N\to\infty}\sup_{x\in\mathbb{R}}\left|\hat{\mathbb{H}}_t(x,\hat{\mathbb{F}}_t(s))-\Phi_A(x)\right| \\ &\quad+\lim_{N\to\infty}\sup_{x\in\mathbb{R}}\left|\Phi_A(x)-\mathbb{H}_t(x,\mathbb{F}_t(s))\right|. \end{aligned} \tag{C.10}$$

We show that each component of the right-hand side of equation (C.10) is bounded by $CA^{5/4}\sqrt{\frac{\kappa_t^3}{2N}}$. By definition, we have that:

$$\begin{aligned}\mathbb{H}_t(x,\mathbb{F}_t(s)) &= \mathbb{P}\left(\max_{a\in\mathcal{A}}\{\psi_t(s,a)\}\leq x\right) \\ &= \mathbb{P}\left(\psi_t(s,a^1)\leq x,\ldots,\psi_t(s,a^A)\leq x\right) \\ &= \mathbb{P}\Bigg(\frac{\hat{Q}_t(s,a^*)-\hat{Q}_t(s,a^1)-\left(Q_t(s,a^*)-Q_t(s,a^1)\right)}{\sqrt{N^{-1}\left[\hat{\sigma}_t^2(s,a^*)+\hat{\sigma}_t^2(s,a^1)\right]}}\leq x,\ldots, \\ &\quad\frac{\hat{Q}_t(s,a^*)-\hat{Q}_t(s,a^A)-\left(Q_t(s,a^*)-Q_t(s,a^A)\right)}{\sqrt{N^{-1}\left[\hat{\sigma}_t^2(s,a^*)+\hat{\sigma}_t^2(s,a^A)\right]}}\leq x\Bigg) \\ &= \mathbb{P}\Bigg(\frac{\frac{1}{N}\sum_n Q_t^n(s,a^*)-Q_t^n(s,a^1)-\left(Q_t(s,a^*)-Q_t(s,a^1)\right)}{\sqrt{N^{-1}\left[\hat{\sigma}_t^2(s,a^*)+\hat{\sigma}_t^2(s,a^1)\right]}}\leq x,\ldots, \\ &\quad\frac{\frac{1}{N}\sum_n Q_t^n(s,a^*)-Q_t^n(s,a^A)-\left(Q_t(s,a^*)-Q_t(s,a^A)\right)}{\sqrt{N^{-1}\left[\hat{\sigma}_t^2(s,a^*)+\hat{\sigma}_t^2(s,a^A)\right]}}\leq x\Bigg) \\ &= \mathbb{P}\Bigg(\frac{\frac{1}{\sqrt{N}}\sum_n Q_t^n(s,a^*)-Q_t^n(s,a^1)-\left(Q_t(s,a^*)-Q_t(s,a^1)\right)}{\sqrt{\hat{\sigma}_t^2(s,a^*)+\hat{\sigma}_t^2(s,a^1)}}\leq x,\ldots, \\ &\quad\frac{\frac{1}{\sqrt{N}}\sum_n Q_t^n(s,a^*)-Q_t^n(s,a^A)-\left(Q_t(s,a^*)-Q_t(s,a^A)\right)}{\sqrt{\hat{\sigma}_t^2(s,a^*)+\hat{\sigma}_t^2(s,a^A)}}\leq x\Bigg) \\ &= \mathbb{P}\left(\frac{1}{\sqrt{N}}\sum_n\hat{Z}_n(a^1)\leq x,\ldots,\frac{1}{\sqrt{N}}\sum_n\hat{Z}_n(a^A)\leq x\right),\end{aligned}$$

where $a^1, \ldots, a^A \in \mathcal{A}$ and

$$\hat{Z}_n(a) := \frac{Q_t^n(s,a^*) - Q_t^n(s,a) - (Q_t(s,a^*) - Q_t(s,a))}{\sqrt{\hat{\sigma}_t^2(s,a^*) + \hat{\sigma}_t^2(s,a)}},$$

for all a and a fixed s. Since $\hat{\sigma}_t^2(s,a) \xrightarrow{a.s.} \sigma_t^2(s,a)$ from Lemma C.5, by Lemma C.2 we get that:

$$\hat{Z}_n(a) \xrightarrow{a.s.} Z_n(a) := \frac{Q_t^n(s,a^*) - Q_t^n(s,a) - (Q_t(s,a^*) - Q_t(s,a))}{\sqrt{\sigma_t^2(s,a^*) + \sigma_t^2(s,a)}}.$$

From Lemma C.7 it holds that $\lim_{N\to\infty} \mathbb{E}[Q_t^n(s_t,a_t)] = Q_t(s_t,a_t)$ and $\lim_{N\to\infty} \mathrm{Var}\left(Q_t^n(s,a)\right) = \sigma_t(s,a)$. Thus, we have that $Z_n(a)$ are iid random variables with $\mathbb{E}\left[Z_n(a)\right] = 0$ and $\mathrm{Var}\left(Z_n(a)\right) = 1$ for all n and a. Since the variance of $Z_n(a)/\sqrt{N}$ can be linearly transformed to 1, by the Multivariate Berry-Esseen Theorem (Bentkus, 2005) we get that:

$$\lim_{N\to\infty} \sup_{x\in\mathbb{R}} |\Phi_A(x) - \mathbb{H}_t(x, \mathbb{F}_t(s))| \leq \frac{CA^{1/4}}{\sqrt{N}} \mathbb{E}\left[\left(Z_n^2(a^1) + Z_n^2(a^2) \ldots + Z_n^2(a^A)\right)^{3/2}\right].$$

We now show that $\mathbb{E}\left[\left(Z_n^2(a^1) + Z_n^2(a^2) \ldots + Z_n^2(a^A)\right)^{3/2}\right] \leq A\sqrt{\frac{1}{2}\kappa_t^3}$. As $\sqrt{x+y} \leq \sqrt{x} + \sqrt{y}$ for any $x, y \in \mathbb{R}_+$, it follows that:

$$\begin{aligned}
\mathbb{E}\left[\left(Z_n^2(a^1) + Z_n^2(a^2) + \ldots + Z_n^2(a^A)\right)^{3/2}\right] &= \mathbb{E}\left[\left(\sqrt{Z_n^2(a^1) + Z_n^2(a^2) + \ldots + Z_n^2(a^A)}\right)^3\right] \\
&\leq \mathbb{E}\left[\left(\sqrt{Z_n^2(a^1)} + \sqrt{Z_n^2(a^2)} + \ldots + \sqrt{Z_n^2(a^A)}\right)^3\right] \\
&= \mathbb{E}\left[\left(Z_n(a^1) + Z_n(a^2) + \ldots + Z_n(a^A)\right)^3\right].
\end{aligned}$$

Expanding $\left(Z_n(a^1) + Z_n(a^2) + \ldots + Z_n(a^A)\right)^3$, we get a summation of terms of the following form: $\mathbb{E}[Z_n^3(a)]$, $\mathbb{E}[Z_n^2(a)Z_n(a')]$, and $\mathbb{E}[Z_n(a)Z_n(a')Z_n(a'')]$ for $a \neq a' \neq a'' \in \mathcal{A}_t(s)$. Thus,

$$\begin{aligned}
\mathbb{E}\left[\left(Z_n(a^1) + Z_n(a^2) + \ldots + Z_n(a^A)\right)^3\right] &= \mathbb{E}\left[Z_n^3(a^1) + Z_n^2(a^1)Z_n(a^2) + \ldots + Z_n^3(a^A)\right] \\
&= \mathbb{E}\left[Z_n^3(a^1)\right] + \mathbb{E}\left[Z_n^2(a^1)Z_n(a^2)\right] + \ldots + \mathbb{E}\left[Z_n^3(a^A)\right].
\end{aligned}$$

Because of the Markov property, we have that $\mathbb{E}[Z_n^2(a)Z_n(a')] = \mathbb{E}[Z_n^2(a)]\mathbb{E}[Z_n(a')]$ and

$\mathbb{E}[Z_n(a)Z_n(a')Z_n(a'')] = \mathbb{E}[Z_n(a)]\mathbb{E}[Z_n(a')]\mathbb{E}[Z_n(a'')]$. Hence,

$$\mathbb{E}\left[\left(Z_n^2(a^1) + Z_n^2(a^2) + \ldots + Z_n^2(a^A)\right)^{3/2}\right] \leq \sum_{a\in\mathcal{A}} \mathbb{E}\left[Z_n^3(a)\right].$$

The proof proceeds by showing that $Z_n(a) \leq \sqrt{\frac{1}{2}\kappa_t^3}$ for all $a \in \mathcal{A}$. Since $Q_t(s,a^*) - Q_t(s,a) > 0$, we have that:

$$\begin{aligned} Z_n(a) &= \frac{Q_t^n(s,a^*) - Q_t^n(s,a) - (Q_t(s,a^*) - Q_t(s,a))}{\sqrt{\sigma_t^2(s,a^*) + \sigma_t^2(s,a)}} \\ &\leq \frac{Q_t^n(s,a^*) - Q_t^n(s,a)}{\sqrt{\sigma_t^2(s,a^*) + \sigma_t^2(s,a)}}. \end{aligned}$$

Further, by Lemma C.1 it holds that $\hat{Q}_t(s,a^*) - \hat{Q}_t(s,a) \leq \kappa_t$, which implies that:

$$Z_n(a) \leq \frac{\kappa_t}{\sqrt{\sigma_t^2(s,a^*) + \sigma_t^2(s,a)}}.$$

From Theorem 2.2 in Sharma et al. (2010) we have that $\sigma_t^2(s,a) \geq \kappa_t^{-1}$, for $\sigma_t^2(s,a) > 0$ and it follows that:

$$Z_n(a) \leq \frac{\kappa_t}{\sqrt{2\kappa_t^{-1}}} = \sqrt{\frac{1}{2}\kappa_t^3}.$$

Since this bound is valid for any a, $\mathbb{E}\left[Z_n(a)\right] = 0$, and $\mathrm{Var}\left(Z_n(a)\right) = 1$, we have that:

$$\mathbb{E}\left[Z_n^3(a)\right] \leq \mathbb{E}\left[Z_n^2(a)\sqrt{\frac{\kappa_t^3}{2}}\right] = \mathbb{E}\left[Z_n^2(a)\right]\sqrt{\frac{1}{2}\kappa_t^3} = \sqrt{\frac{1}{2}\kappa_t^3}.$$

Consequently,

$$\mathbb{E}\left[\left(Z_n^2(a^1) + Z_n^2(a^2)\ldots + Z_n^2(a^A)\right)^{3/2}\right] \leq A\sqrt{\frac{1}{2}\kappa_t^3}, \tag{C.11}$$

and it follows that:

$$\lim_{N\to\infty}\sup_{x\in\mathbb{R}} |\Phi_A(x) - \mathbb{H}_t(x, \mathbb{F}_t(s))| \leq CA^{5/4}\sqrt{\frac{\kappa_t^3}{2N}}.$$

In a similar way,

$$
\begin{aligned}
\hat{\mathbb{H}}_t(x, \hat{\mathbb{F}}_t(s)) &= \mathbb{P}\left(\max_{a\in\mathcal{A}}\{\hat{\psi}_t(s,a)\} \le x\right) \\
&= \mathbb{P}\left(\hat{\psi}_t(s,a^1) \le x, \ldots, \hat{\psi}_t(s,a^A) \le x\right) \\
&= \mathbb{P}\Bigg(\frac{\hat{Q}_t(s,a^*) - \hat{Q}_t(s,a^1) - \left(Q_t(s,a^*) - Q_t(s,a^1)\right)}{\sqrt{M^{-1}\left[\hat{\zeta}_t^2(s,a^*) + \hat{\zeta}_t^2(s,a^1)\right]}} \le x, \ldots, \\
&\quad \frac{\hat{Q}_t(s,a^*) - \hat{Q}_t(s,a^A) - \left(Q_t(s,a^*) - Q_t(s,a^A)\right)}{\sqrt{M^{-1}\left[\hat{\zeta}_t^2(s,a^*) + \hat{\zeta}_t^2(s,a^A)\right]}} \le x\Bigg) \\
&= \mathbb{P}\Bigg(\frac{\frac{1}{MK}\sum_m\sum_k Q_t^{m,k}(s,a^*) - Q_t^{m,k}(s,a^1) - \left(Q_t(s,a^*) - Q_t(s,a^1)\right)}{\sqrt{\frac{K}{MK}\left[\hat{\zeta}_t^2(s,a^*) + \hat{\zeta}_t^2(s,a^1)\right]}} \le x, \ldots, \\
&\quad \frac{\frac{1}{MK}\sum_m\sum_k Q_t^{m,k}(s,a^*) - Q_t^{m,k}(s,a^A) - \left(Q_t(s,a^*) - Q_t(s,a^A)\right)}{\sqrt{\frac{K}{MK}\left[\hat{\zeta}_t^2(s,a^*) + \hat{\zeta}_t^2(s,a^A)\right]}} \le x\Bigg) \\
&= \mathbb{P}\Bigg(\frac{\frac{1}{\sqrt{MK}}\sum_m\sum_k Q_t^{m,k}(s,a^*) - Q_t^{m,k}(s,a^1) - \left(Q_t(s,a^*) - Q_t(s,a^1)\right)}{\sqrt{K\left[\hat{\zeta}_t^2(s,a^*) + \hat{\zeta}_t^2(s,a^1)\right]}} \le x, \ldots, \\
&\quad \frac{\frac{1}{\sqrt{MK}}\sum_m\sum_k Q_t^{m,k}(s,a^*) - Q_t^{m,k}(s,a^A) - \left(Q_t(s,a^*) - Q_t(s,a^A)\right)}{\sqrt{K\left[\hat{\zeta}_t^2(s,a^*) + \hat{\zeta}_t^2(s,a^A)\right]}} \le x\Bigg) \\
&= \mathbb{P}\left(\frac{1}{\sqrt{MK}}\sum_m\sum_k \bar{Z}_{m,k}(a^1) \le x, \ldots, \frac{1}{\sqrt{MK}}\sum_m\sum_k \bar{Z}_{m,k}(a^A) \le x\right),
\end{aligned}
$$

where $a^1, \ldots, a^A \in \mathcal{A}_t(s)$ and

$$
\bar{Z}_{m,k}(a) := \frac{Q_t^{m,k}(s,a^*) - Q_t^{m,k}(s,a) - (Q_t(s,a^*) - Q_t(s,a))}{\sqrt{K\left[\hat{\zeta}_t^2(s,a^*) + \hat{\zeta}_t^2(s,a)\right]}},
$$

for all a and a fixed s. Since $K\hat{\zeta}_t^2(s,a) \xrightarrow{a.s.} \sigma_t^2(s,a)$ from Lemma C.5, by Lemma C.2 we get that:

$$
\bar{Z}_{m,k}(a) \xrightarrow{a.s.} Z_{m,k}(a) := \frac{Q_t^{m,k}(s,a^*) - Q_t^{m,k}(s,a) - (Q_t(s,a^*) - Q_t(s,a))}{\sqrt{\sigma_t^2(s,a^*) + \sigma_t^2(s,a)}}.
$$

As the variance of $Z_{m,k}(a)/\sqrt{MK}$ can be linearly transformed to 1, by the Multivariate

Berry-Esseen Theorem (Bentkus, 2005) we get that:

$$\lim_{M\to\infty}\lim_{K\to\infty}\sup_{x\in\mathbb{R}}\left|\hat{\mathbb{H}}_t(x,\hat{\mathbb{F}}_t(s))-\Phi_A(x)\right| \leq \frac{CA^{1/4}}{\sqrt{MK}}\mathbb{E}\left[\left(Z^2_{m,k}(a^1)+Z^2_{m,k}(a^2)\ldots+Z^2_{m,k}(a^A)\right)^{3/2}\right],$$

and by (C.11) it holds that:

$$\lim_{M\to\infty}\lim_{K\to\infty}\sup_{x\in\mathbb{R}}\left|\hat{\mathbb{H}}_t(x,\hat{\mathbb{F}}_t(s))-\Phi_A(x)\right| \leq CA^{5/4}\sqrt{\frac{\kappa_t^3}{2MK}} = CA^{5/4}\sqrt{\frac{\kappa_t^3}{2N}}.$$

Summing both components of the right-hand side of equation (C.10) it follows that:

$$\begin{aligned}\lim_{M\to\infty}\lim_{K\to\infty}\sup_{x\in\mathbb{R}}\left|\hat{\mathbb{H}}_t(x,\hat{\mathbb{F}}_t(s))-\mathbb{H}_t(x,\mathbb{F}_t(s))\right| &\leq 2CA^{5/4}\sqrt{\frac{\kappa_t^3}{2N}}\\ &= CA^{5/4}\sqrt{\frac{2\kappa_t^3}{N}}.\end{aligned}$$

□

C.2.5 Proofs of Section 4.6.2.2

Proposition 7 (Proposition 4.7.) *Suppose that Assumption 4.1 holds. Then, (i) $|\Pi_t(s,\alpha)|$ is nonincreasing in $\alpha \in (0,1)$. Moreover, (ii) there exist an α such that $\Pi_t(s,\alpha) \subseteq \mathcal{A}_t^*(s)$ with probability 1 for N large enough.*

Proof.

(i) We first show that $|\Pi_t(s,\alpha)|$ is nonincreasing in $\alpha \in (0,1)$. By definition, the quantile function of $\hat{\psi}_t(s,a)$ for α is given by:

$$\begin{aligned}\hat{d}_t(s,\alpha) &= \inf\left\{x\in\mathbb{R} : \hat{\mathbb{H}}_t(x,\hat{\mathbb{F}}_t(s)) \geq 1-\alpha\right\}\\ &= \inf\left\{x\in\mathbb{R} : \frac{1}{M}\sum_{m=1}^{M}\mathbb{1}\{\bar{\psi}_t^m(s,a)\leq x\} \geq 1-\alpha\right\}\end{aligned}$$

Suppose $\alpha_1 < \alpha_2$, then we have:

$$\begin{aligned}\hat{d}_t(\alpha_1, s) &= \inf\left\{x \in \mathbb{R} : \frac{1}{M}\sum_{m=1}^{M} \mathbb{1}\{\bar{\psi}_t^m(s,a) \leq x\} \geq 1-\alpha_1\right\} \\ &\geq \inf\left\{x \in \mathbb{R} : \frac{1}{M}\sum_{m=1}^{M} \mathbb{1}\{\bar{\psi}_t^m(s,a) \leq x\} \geq 1-\alpha_2\right\} \\ &= \hat{d}_t(\alpha_2, s).\end{aligned}$$

Thus, $\hat{d}_t(s,\alpha)$ is nonincreasing in α, which implies that $|\Pi_t(s,\alpha)|$ is nonincreasing in α.

(ii) We now show that there exist an α such that $\Pi_t(s,\alpha) \subseteq \mathcal{A}_t^*(s)$ with probability 1 for N large enough. Suppose that $C_t(s) = \hat{\mathcal{A}}_t^*(s)$, otherwise the claim follows trivially. Since $\mathbb{H}_t$ follows a $\mathcal{U}(0,1)$ distribution we have that $d_t(s,\alpha)$ is continuous and strictly increasing by the Continuous Inverse Theorem (Bartle and Sherbert, 2010). The proof proceeds by contradiction. Suppose that there is no $\alpha \in (0,1)$ such that $\Pi_t(s,\alpha) = \hat{\mathcal{A}}_t^*(s)$. Because of the continuity of $\hat{d}_t(s,\alpha)$ there exists a point α such that $\hat{d}_t(s,1) < \hat{d}_t(s,\alpha) < \hat{d}_t(s,0)$ by the Intermediate Value Theorem (Bartle and Sherbert, 2010). Thus, for every K, $\hat{\zeta}_t^2(s,a^*) > 0$, and $\hat{\zeta}_t^2(s,a) > 0$ we can find an α such that:

$$\hat{Q}_t(s,a^*) - \hat{Q}_t(s,a) > \hat{d}_t(s,\alpha)\sqrt{M^{-1}\left[\hat{\zeta}_t^2(s,a^*) + \hat{\zeta}_t^2(s,a)\right]},$$

for any $a^* \in \hat{\mathcal{A}}_t^*(s)$ and all $a \in \hat{\mathcal{A}}_t(s)$, contradicting the supposition that there is no $\alpha \in (0,1)$ such that $\Pi_t(s,\alpha) = \hat{\mathcal{A}}_t^*(s)$. As from Corollary 4.1 it holds that $\hat{\mathcal{A}}_t^*(s) \subseteq \mathcal{A}_t^*(s)$, we get that $\Pi_t(s,\alpha) \subseteq \mathcal{A}_t^*(s)$ with probability 1 for N large enough. □

The proofs of the remaining results of this subsection rely in the following lemma.

Lemma C.8 *Under Assumption 4.1 the following holds:*

$$\begin{aligned}\sigma_t^2(s,a) = \mathbb{E}[r_t^2(s,a,\omega)|s,a] - \mathbb{E}[r_t(s,a,\omega)|s,a]^2 \\ + \gamma^2\left(\mathbb{E}[v_{t+1}^2(f_{t+1}(s,a,\omega))|s,a] - \mathbb{E}[v_{t+1}(f_{t+1}(s,a,\omega))|s,a]^2\right).\end{aligned}$$

Proof. By definition, we have that:

$$\begin{aligned}\sigma_t^2(s,a) &= \mathbb{E}\left[(r_t(s,a,\omega) + \gamma v_{t+1}(f_{t+1}(s,a,\omega)))^2|s,a\right] - \mathbb{E}\left[r_t(s,a,\omega) + \gamma v_{t+1}(f_{t+1}(s,a,\omega))|s,a\right]^2 \\ &= \mathbb{E}[r_t^2(s,a,\omega)|s,a] + \gamma^2\mathbb{E}[v_{t+1}^2(f_{t+1}(s,a,\omega))|s,a] - \mathbb{E}[r_t(s,a,\omega)|s,a]^2 \\ &\quad - \gamma^2\mathbb{E}[v_{t+1}(f_{t+1}(s,a,\omega))|s,a]^2,\end{aligned}$$

where the second equality follows because:

$$\begin{aligned}\mathbb{E}[r_t(s,a,\omega)v_{t+1}(f_{t+1}(s,a,\omega))|s,a] &= \sum_{s'} p_t(s'|s,a)\mathbb{E}[r_t(s,a,\omega)|s,a]v_{t+1}(s') \\ &= \mathbb{E}[r_t(s,a,\omega)|s,a]\sum_{s'} p_t(s'|s,a)v_{t+1}(s') \\ &= \mathbb{E}[r_t(s,a,\omega)|s,a]\mathbb{E}[v_{t+1}(f_{t+1}(s,a,\omega))|s,a].\end{aligned}$$

□

Proposition 8 (Proposition 4.8.) *Suppose Assumptions 4.1 through 4.4 hold. Then,* $|\Pi_t(s,\alpha)|$ *is* ϵ*-nonincreasing in* s *with probability 1 for* N *large enough.*

The proof of this proposition depends on the following lemma.

Lemma C.9 *Let* X*,* Y*, and* Z *be partially ordered finite sets,* $g : X \mapsto \mathbb{R}$ *be a nonincreasing function of* X*, and* $h : X \times Y \times Z \mapsto [0,1]$ *be a function satisfying:*

$$\sum_{x\in X} h(x,y^+,z^+) + h(x,y^-,z^-) = \sum_{x\in X} h(x,y^+,z^-) + h(x,y^-,z^+).$$

Then, we have (i):

$$\sum_{x\in X} \left[h(x,y^+,z^+) + h(x,y^-,z^-)\right] g(x) \leq \sum_{x\in X} \left[h(x,y^+,z^-) + h(x,y^-,z^+)\right] g(x),$$

if h *is a superadditive function and (ii):*

$$\sum_{x\in X} \left[h(x,y^+,z^+) + h(x,y^-,z^-)\right] g(x) \geq \sum_{x\in X} \left[h(x,y^+,z^-) + h(x,y^-,z^+)\right] g(x),$$

if h *is a subadditive function for* $y^+ \geq y^- \in Y$*, and* $z^+ \geq z^- \in Z$*.*

Proof of Lemma C.9.

(i) By definition we have:

$$\sum_{x'\geq x} \left[h(x',y^+,z^+) + h(x',y^-,z^-)\right] \geq \sum_{x'\geq x} \left[h(x',y^+,z^-) + h(x',y^-,z^+)\right] \tag{C.12}$$

Moreover, we have:

$$\sum_{x\in X} h(x,y^+,z^+) + \sum_{x\in X} h(x,y^-,z^-) - \sum_{x\in X} h(x,y^+,z^-) - \sum_{x\in X} h(x,y^-,z^+) = 0.$$

Since $g(x)$ is nonincreasing on $x \in X$ it holds that:

$$
\begin{aligned}
0 &= \sum_{x\in X} h(x, y^+, z^+) + \sum_{x\in X} h(x, y^-, z^-) - \sum_{x\in X} h(x, y^+, z^-) - \sum_{x\in X} h(x, y^-, z^+) \\
&= \left[\sum_{x} h(x, y^+, z^+) + \sum_{x} h(x, y^-, z^-) - \sum_{x} h(x, y^+, z^-) - \sum_{x} h(x, y^-, z^+)\right] g(x^1) \\
&\geq h(x^1, y^+, z^+)g(x^1) + h(x^1, y^-, z^-)g(x^1) - h(x^1, y^+, z^-)g(x^1) - h(x^1, y^-, z^+)g(x^1) \\
&\quad + \Bigg[\sum_{x\in X\setminus\{x^1\}} h(x, y^+, z^+) + \sum_{x\in X\setminus\{x^1\}} h(x, y^-, z^-) - \sum_{x\in X\setminus\{x^1\}} h(x, y^+, z^-) \\
&\qquad - \sum_{x\in X\setminus\{x^1\}} h(x, y^-, z^+)\Bigg] g(x^2) \\
&\geq \sum_{x'\in\{x^1,x^2\}} h(x', y^+, z^+)g(x') + \sum_{x'\in\{x^1,x^2\}} h(x', y^-, z^-)g(x') - \sum_{x'\in\{x^1,x^2\}} h(x', y^+, z^-)g(x') \\
&\quad - \sum_{x'\in\{x^1,x^2\}} h(x', y^-, z^+)g(x') + \Bigg[\sum_{x\in X\setminus\{x^1,x^2\}} h(x, y^+, z^+) + \sum_{x\in X\setminus\{x^1,x^2\}} h(x, y^-, z^-) \\
&\quad - \sum_{x\in X\setminus\{x^1,x^2\}} h(x, y^+, z^-) - \sum_{x\in X\setminus\{x^1,x^2\}} h(x, y^-, z^+)\Bigg] g(x^3),
\end{aligned}
$$

where $x^1 \leq x^2 \leq x^3 \in X$, $y^+ \geq y^- \in Y$, and $z^+ \geq z^- \in Z$. This pattern implies that:

$$
0 \geq \sum_{x'\in X} h(x', y^+, z^+)g(x') + \sum_{x'\in X} h(x', y^-, z^-)g(x') - \sum_{x'\in X} h(x', y^+, z^-)g(x') - \sum_{x'\in X} h(x', y^-, z^+)g(x').
$$

Thus,

$$
\sum_{x'\in X} h(x', y^+, z^+)g(x') + h(x', y^-, z^-)g(x') \leq \sum_{x'\in X} h(x', y^+, z^-)g(x') + h(x', y^-, z^+)g(x'),
$$

which implies that:

$$
\sum_{x'\in X} \left[h(x', y^+, z^+) + h(x', y^-, z^-)\right] g(x') \leq \sum_{x'\in X} \left[h(x', y^+, z^-) + h(x', y^-, z^+)\right] g(x'),
$$

for $y^+ \geq y^- \in Y$ and $z^+ \geq z^- \in Z$.

(ii) For the subadditive case, note that we have:

$$
\sum_{x'\geq x} \left[h(x', y^+, z^+) + h(x', y^-, z^-)\right] \leq \sum_{x'\geq x} \left[h(x', y^+, z^-) + h(x', y^-, z^+)\right],
$$

by definition and that:

$$\sum_{x\in X} h(x, y^+, z^-) + \sum_{x\in X} h(x, y^-, z^+) - \sum_{x\in X} h(x, y^+, z^+) - \sum_{x\in X} h(x, y^-, z^-) = 0.$$

The rest of the proof proceeds in the same way as the superadditive case and we get:

$$\sum_{x'\in X} \left[h(x', y^+, z^+) + h(x', y^-, z^-)\right] g(x') \geq \sum_{x'\in X} \left[h(x', y^+, z^-) + h(x', y^-, z^+)\right] g(x'),$$

for $y^+ \geq y^- \in Y$ and $z^+ \geq z^- \in Z$. □

Proof of Proposition 4.8. By definition, we have that:

$$\Pi_t(s,\alpha) = \left\{a \in \mathcal{A} : \hat{Q}_t(s,a^*) - \hat{Q}_t(s,a) \leq d_t(s,\alpha)\sqrt{M^{-1}\left[\hat{\zeta}_t^2(s,a^*) + \hat{\zeta}_t^2(s,a)\right]}\right\}.$$

To show that $|\Pi_t(s,\alpha)|$ is ϵ-nonincreasing in s, it suffices to demonstrate that $\hat{Q}_t(s,a^*) - \hat{Q}_t(s,a)$ is $\frac{1}{4}\epsilon$-nondecreasing when $\hat{d}_t(s,\alpha)\sqrt{M^{-1}\left[\hat{\zeta}_t^2(s,a^*) + \hat{\zeta}_t^2(s,a)\right]}$ is $\frac{3}{4}\epsilon$-nonincreasing. We first prove that $\hat{Q}_t(s,a^*) - \hat{Q}_t(s,a)$ is $\frac{1}{4}\epsilon$-nondecreasing in s with probability 1 for N large enough.

Suppose that $p_t(s'|s,a)$ and $\mathbb{E}[r_t(s,a,\omega)|s,a]$ are known for all s, a, and t. Then, for any $s \in \mathcal{S}$ we have:

$$\begin{aligned}
Q_t(s,a^*) - Q_t(s,a) &= \mathbb{E}[r_t(s,a^*,\omega) + \gamma v_{t+1}(s')|s,a^*] - \mathbb{E}[r_t(s,a,\omega) + \gamma v_{t+1}(s')|s,a] \\
&= \mathbb{E}[r_t(s,a^*,\omega)|s,a^*] + \gamma\mathbb{E}[v_{t+1}(s')|s,a^*] - \mathbb{E}[r_t(s,a,\omega)|s,a] - \gamma\mathbb{E}[v_{t+1}(s'))|s,a] \\
&= \mathbb{E}[r_t(s,a^*,\omega)|s,a^*] - \mathbb{E}[r_t(s,a,\omega)|s,a] \\
&\quad + \gamma\left(\sum_{s'} p_t(s'|s,a^*)v_{t+1}(s') - \sum_{s'} p_t(s'|s,a)v_{t+1}(s')\right) \\
&= \mathbb{E}[r_t(s,a^*,\omega)|s,a^*] - \mathbb{E}[r_t(s,a,\omega)|s,a] \\
&\quad + \gamma\left(\sum_{s'} \left[p_t(s'|s,a^*) - p_t(s'|s,a)\right] v_{t+1}(s')\right).
\end{aligned}$$

By Assumption 4.4, we have that $\mathbb{E}[r_t(s,a^*,\omega)|s,a^*] - \mathbb{E}[r_t(s,a,\omega)|s,a]$ is nondecreasing in s and it follows that:

$$Q_t(s,a^*) - Q_t(s,a) \leq \mathbb{E}[r_t(\bar{s},a^*,\omega)|\bar{s},a^*] - \mathbb{E}[r_t(\bar{s},a,\omega)|\bar{s},a] + \gamma\left(\sum_{s'} \left[p_t(s'|s,a^*) - p_t(s'|s,a)\right] v_{t+1}(s')\right),$$

for $s \leq \bar{s}$. From Assumption 4.2 we get that $v_{t+1}(s')$ is nonincreasing in s by Proposition 4.7.3 in Puterman (2014). Since $\bar{p}_t(s'|s,a)$ is subadditive on $\mathcal{S}\times\mathcal{A}$ by Assumption 4.4, applying

Lemma C.9 with $h(x, y^+, z^+) = p_t(s'|\bar{s}, a^*)$, $h(x, y^-, z^-) = p_t(s'|s, a)$, $h(x, y^+, z^-) = p_t(s'|\bar{s}, a)$, $h(x, y^-, z^+) = p_t(s'|s, a^*)$, and $g(x) = v_{t+1}(s')$ we get:

$$\sum_{s'} p_t(s'|\bar{s}, a)v_{t+1}(s') + p_t(s'|s, a^*)v_{t+1}(s') \leq \sum_{s'} p_t(s'|\bar{s}, a^*)v_{t+1}(s') + p_t(s'|s, a)v_{t+1}(s')$$

which implies:

$$\sum_{s'} \left[p_t(s'|s, a^*) - p_t(s'|s, a)\right] v_{t+1}(s') \leq \sum_{s'} \left[p_t(s'|\bar{s}, a^*) - p_t(s'|\bar{s}, a)\right] v_{t+1}(s')$$

for $s \leq \bar{s}$ and $a \leq a^*$. Therefore, it follows that:

$$\begin{aligned}
Q_t(s, a^*) - Q_t(s, a) &\leq \mathbb{E}[r_t(\bar{s}, a^*, \omega)|s, a^*] - \mathbb{E}[r_t(\bar{s}, a, \omega)|s, a] \\
&\quad + \gamma \left(\sum_{s'} \left[p_t(s'|\bar{s}, a^*) - p_t(s'|\bar{s}, a)\right] v_{t+1}(s') \right) \\
&= \mathbb{E}[r_t(\bar{s}, a^*, \omega)|\bar{s}, a^*] - \mathbb{E}[r_t(\bar{s}, a, \omega)|\bar{s}, a] \\
&\quad + \gamma \left(\sum_{s'} p_t(s'|\bar{s}, a^*)v_{t+1}(s') - \sum_{s'} p_t(s'|\bar{s}, a)v_{t+1}(s') \right) \\
&= \mathbb{E}[r_t(\bar{s}, a^*, \omega)|\bar{s}, a^*] - \mathbb{E}[r_t(\bar{s}, a, \omega)|\bar{s}, a] + \gamma \mathbb{E}[v_{t+1}(s')|\bar{s}, a^*] - \gamma \mathbb{E}[v_{t+1}(s')|\bar{s}, a] \\
&= \mathbb{E}[r_t(\bar{s}, a^*, \omega) + \gamma v_{t+1}(s')|\bar{s}, a^*] - \mathbb{E}[r_t(\bar{s}, a, \omega) + \gamma v_{t+1}(s')|\bar{s}, a] \\
&= Q_t(\bar{s}, a^*) - Q_t(\bar{s}, a). \qquad \text{(C.13)}
\end{aligned}$$

Hence, $Q_t(s, a^*) - Q_t(s, a)$ is nondecreasing in s. From Theorem 4.1 and Lemma C.2 we get that $\hat{Q}_t(s, a^*) - \hat{Q}_t(s, a) \xrightarrow{a.s.} Q_t(s, a^*) - Q_t(s, a)$, and it holds that $\hat{Q}_t(s, a^*) - \hat{Q}_t(s, a)$ is $\frac{1}{4}\epsilon$-nondecreasing in s for N large enough.

We now show that $\hat{d}_t(s, \alpha)\sqrt{M^{-1}\left[\hat{\zeta}_t^2(s, a^*) + \hat{\zeta}_t^2(s, a)\right]}$ is $\frac{3}{4}\epsilon$-nondecreasing in s with probability 1 for N large enough. Combining Lemma C.6 and Theorem 2.3.1 in Serfling (1980) we get that $\hat{d}_t(s, \alpha) \xrightarrow{a.s.} d_t(s, \alpha)$. Since $\hat{\psi}_t(s, a)$ is a pivotal statistic, its distribution does not depend on s (Casella and Beroer, 2001, Section 9.2.2). Thus, $d_t(s, \alpha)$ is constant in s and is $\hat{d}_t(s, \alpha)$ $\frac{1}{4}\epsilon$-constant with probability 1 for large enough N. By Lemma C.8, it holds that:

$$\sigma_t^2(s, a) = \mathbb{E}[r_t^2(s, a, \omega)|s, a] - \mathbb{E}[r_t(s, a, \omega)|s, a]^2 + \gamma^2 \left(\mathbb{E}[v_{t+1}^2(s')|s, a] - \mathbb{E}[v_{t+1}(s')|s, a]^2\right).$$

Further, from Assumption 4.3 we have that $\mathbb{E}[r_t^2(s, a, \omega)|s, a] - \mathbb{E}[r_t(s, a, \omega)|s, a]^2$ and $\mathbb{E}[v_{t+1}^2(s')|s, a] - \mathbb{E}[v_{t+1}(s')|s, a]^2$ are nonincreasing in s. Since $K\hat{\zeta}_t^2(s, a) \xrightarrow{a.s.} \sigma_t^2(s, a)$ by Lemma C.5, it follows that $K\hat{\zeta}_t^2(s, a)$ is $\frac{1}{4}\epsilon$-nonincreasing in s for all a and large enough N. Conse-

quently, $\hat{d}_t(s,\alpha)\sqrt{M^{-1}\left[\hat{\zeta}_t^2(s,a^*)+\hat{\zeta}_t^2(s,a)\right]}$ is $\frac{3}{4}\epsilon$-nonincreasing in s with probability 1 for N large enough.

Therefore, we get that:

$$\hat{Q}_t(s,a^*)-\hat{Q}_t(s,a)-\frac{1}{4}\epsilon \leq \hat{Q}_t(\bar{s},a^*)-\hat{Q}_t(\bar{s},a)$$

and

$$\hat{d}_t(\bar{s},\alpha)\sqrt{M^{-1}\left[\hat{\zeta}_t^2(\bar{s},a^*)+\hat{\zeta}_t^2(\bar{s},a)\right]} \leq \hat{d}_t(s,\alpha)\sqrt{M^{-1}\left[\hat{\zeta}_t^2(s,a^*)+\hat{\zeta}_t^2(s,a)\right]}+\frac{3}{4}\epsilon,$$

for $\epsilon > 0$ and $s \leq \bar{s}$ with probability 1 for N large enough, which completes the proof. □

Proposition 9 (Proposition 4.9.) *Suppose Assumptions 4.1 through 4.5 hold. Furthermore, assume that $v_t(s) - v_{t+1}(s)$ is nondecreasing in s'. Then, $|\Pi_t(s,\alpha)|$ is ϵ-nonincreasing in t with probability 1 for N large enough.*

Proof. By definition, we have that:

$$\Pi_t(s,\alpha)=\left\{a\in\mathcal{A}_t(s):\hat{Q}_t(s,a^*)-\hat{Q}_t(s,a)\leq d_t(s,\alpha)\sqrt{M^{-1}\left[\hat{\zeta}_t^2(s,a^*)+\hat{\zeta}_t^2(s,a)\right]}\right\}.$$

To show that $|\Pi_t(s,\alpha)|$ is ϵ-nonincreasing in t, it suffices to demonstrate that $\hat{Q}_t(s,a^*)-\hat{Q}_t(s,a)$ is $\frac{1}{4}\epsilon$-nondecreasing when $\hat{d}_t(s,\alpha)\sqrt{M^{-1}\left[\hat{\zeta}_t^2(s,a^*)+\hat{\zeta}_t^2(s,a)\right]}$ is $\frac{3}{4}\epsilon$-nonincreasing. We first prove that $\hat{Q}_t(s,a^*)-\hat{Q}_t(s,a)$ is ϵ-nondecreasing in t with probability 1 for N large enough.

Suppose that $p_t(s'|s,a)$ and $\mathbb{E}[r_t(s,a,\omega)|s,a]$ are known for all s, a, and t. Then, for any $t\in\mathcal{T}\setminus\{T\}$ we have:

$$\begin{aligned}
Q_t(s,a^*)-Q_t(s,a) &= \mathbb{E}[r_t(s,a^*,\omega)+\gamma v_{t+1}(s')|s,a^*]-\mathbb{E}[r_t(s,a,\omega)+\gamma v_{t+1}(s')|s,a]\\
&= \mathbb{E}[r_t(s,a^*,\omega)|s,a^*]+\gamma\mathbb{E}[v_{t+1}(s')|s,a^*]-\mathbb{E}[r_t(s,a,\omega)|s,a]-\gamma\mathbb{E}[v_{t+1}(s'))|s,a]\\
&= \mathbb{E}[r_t(s,a^*,\omega)|s,a^*]-\mathbb{E}[r_t(s,a,\omega)|s,a]\\
&\quad+\gamma\left(\sum_{s'}p_t(s'|s,a^*)v_{t+1}(s')-\sum_{s'}p_t(s'|s,a)v_{t+1}(s')\right)\\
&= \mathbb{E}[r_t(s,a^*,\omega)|s,a^*]-\mathbb{E}[r_t(s,a,\omega)|s,a]\\
&\quad+\gamma\left(\sum_{s'}\left[p_t(s'|s,a^*)-p_t(s'|s,a)\right]v_{t+1}(s')\right).
\end{aligned}$$

By Assumption 4.3, we have that $\mathbb{E}[r_t(s,a^*,\omega)|s,a^*]-\mathbb{E}[r_t(s,a,\omega)|s,a]$ is nondecreasing in t

and it follows that:

$$Q_t(s,a^*) - Q_t(s,a) \leq \mathbb{E}[r_{t+1}(s,a^*,\omega)|s,a^*] - \mathbb{E}[r_{t+1}(s,a,\omega)|s,a] + \gamma\left(\sum_{s'}\left[p_t(s'|s,a^*) - p_t(s'|s,a)\right]v_{t+1}(s')\right).$$

Since $v_{t+1}(s')$ is nonincreasing in s by Assumption 4.2 from Proposition 4.7.3 in Puterman (2014) and $\bar{p}_t(s'|s,a)$ is a subadditive function on $\mathcal{T}\times\mathcal{A}$ by Assumption 4.4, we get:

$$\sum_{s'}\left[p_t(s'|s,a^*) - p_t(s'|s,a)\right]v_{t+1}(s') \leq \sum_{s'}\left[p_{t+1}(s'|s,a^*) - p_{t+1}(s'|s,a)\right]v_{t+1}(s')$$

for $a \leq a^*$ by Lemma C.9 with $h(x,y^+,z^+) = p_{t+1}(s'|s,a^*)$, $h(x,y^-,z^-) = p_t(s'|s,a)$, $h(x,y^+,z^-) = p_{t+1}(s'|s,a)$, $h(x,y^-,z^+) = p_t(s'|s,a^*)$, and $g(x) = v_{t+1}(s')$. Since $v_{t+1}(s') - v_{t+2}(s')$ is nondecreasing in s' by assumption and $\bar{p}_t(s'|s,a^*) \leq \bar{p}_t(s'|s,a)$ from Assumption 4.5, by Lemma C.4 it holds that:

$$\sum_{s'} p_{t+1}(s'|s,a^*)\left[v_{t+1}(s') - v_{t+2}(s')\right] \leq \sum_{s'} p_{t+1}(s'|s,a)\left[v_{t+1}(s') - v_{t+2}(s')\right],$$

which implies:

$$\sum_{s'}\left[p_{t+1}(s'|s,a^*) - p_{t+1}(s'|s,a)\right]v_{t+1}(s') \leq \sum_{s'}\left[p_{t+1}(s'|s,a^*) - p_{t+1}(s'|s,a)\right]v_{t+2}(s').$$

Therefore, we get that:

$$\sum_{s'}\left[p_t(s'|s,a^*) - p_t(s'|s,a)\right]v_{t+1}(s') \leq \sum_{s'}\left[p_{t+1}(s'|s,a^*) - p_{t+1}(s'|s,a)\right]v_{t+2}(s')$$

and it follows that:

$$\begin{aligned}
Q_t(s,a^*) - Q_t(s,a) &\leq \mathbb{E}[r_{t+1}(s,a^*,\omega)|s,a^*] - \mathbb{E}[r_{t+1}(s,a,\omega)|s,a] \\
&\quad + \gamma\left(\sum_{s'}\left[p_{t+1}(s'|s,a^*) - p_{t+1}(s'|s,a)\right]v_{t+2}(s')\right) \\
&= \mathbb{E}[r_{t+1}(s,a^*,\omega)|s,a^*] - \mathbb{E}[r_{t+1}(s,a,\omega)|s,a] \\
&\quad + \gamma\left(\sum_{s'} p_{t+1}(s'|s,a^*)v_{t+2}(s') - \sum_{s'} p_{t+1}(s'|s,a)v_{t+2}(s')\right) \\
&= \mathbb{E}[r_{t+1}(s,a^*,\omega) + \gamma v_{t+2}(s')|s,a^*] - \mathbb{E}[r_{t+1}(s,a,\omega) + \gamma v_{t+2}(s')|s,a] \\
&= Q_{t+1}(s,a^*) - Q_{t+1}(s,a). \qquad \text{(C.14)}
\end{aligned}$$

Hence, $Q_t(s,a^*) - Q_t(s,a)$ is nondecreasing in t. From Theorem 4.1 and Lemma C.2 we get that $\hat{Q}_t(s,a^*) - \hat{Q}_t(s,a) \xrightarrow{a.s.} Q_t(s,a^*) - Q_t(s,a)$, and it holds that $\hat{Q}_t(s,a^*) - \hat{Q}_t(s,a)$ is $\frac{1}{4}\epsilon$-nondecreasing in t for N large enough.

We now show that $\hat{d}_t(s,\alpha)\sqrt{M^{-1}\left[\hat{\zeta}_t^2(s,a^*) + \hat{\zeta}_t^2(s,a)\right]}$ is $\frac{3}{4}\epsilon$-nonincreasing in t. Combining Lemma C.6 and Theorem 2.3.1 in Serfling (1980) we get that $\hat{d}_t(s,\alpha) \xrightarrow{a.s.} d_t(s,\alpha)$. Because $\hat{\psi}_t(s,a)$ is a pivotal statistic, its distribution does not depend on t. Thus, $d_t(s,\alpha)$ is constant in t and $\hat{d}_t(s,\alpha)$ is $\frac{1}{4}\epsilon$-constant in t with probability 1 for large enough N. By Lemma C.8, it holds that:

$$\sigma_t^2(s,a) = \mathbb{E}[r_t^2(s,a,\omega)|s,a] - \mathbb{E}[r_t(s,a,\omega)|s,a]^2 + \gamma^2\left(\mathbb{E}[v_{t+1}^2(s')|s,a] - \mathbb{E}[v_{t+1}(s')|s,a]^2\right).$$

Further, from Assumption 4.3 we have that $\mathbb{E}[r_t^2(s,a,\omega)|s,a] - \mathbb{E}[r_t(s,a,\omega)|s,a]^2$ and $\mathbb{E}[v_{t+1}^2(s')|s,a] - \mathbb{E}[v_{t+1}(s')|s,a]^2$ are nonincreasing in t. Since $K\hat{\zeta}_t^2(s,a) \xrightarrow{a.s.} \sigma_t^2(s,a)$ by Lemma C.5, it follows that $K\hat{\zeta}_t^2(s,a)$ is $\frac{1}{4}\epsilon$-nonincreasing in t for all a and large enough N. Consequently, $\hat{d}_t(s,\alpha)\sqrt{M^{-1}\left[\hat{\zeta}_t^2(s,a^*) + \hat{\zeta}_t^2(s,a)\right]}$ is $\frac{3}{4}\epsilon$-nonincreasing in t with probability 1 for N large enough. Combining this result with equation (C.14) we get:

$$\begin{aligned}
\hat{Q}_t(s,a^*) - \hat{Q}_t(s,a) - \frac{1}{4}\epsilon &\leq \hat{Q}_{t+1}(s,a^*) - \hat{Q}_{t+1}(s,a) \\
&\leq \hat{d}_{t+1}(s,\alpha)\sqrt{M^{-1}\left[\hat{\zeta}_{t+1}^2(s,a^*) + \hat{\zeta}_{t+1}^2(s,a)\right]} \\
&\leq \hat{d}_t(s,\alpha)\sqrt{M^{-1}\left[\hat{\zeta}_t^2(s,a^*) + \hat{\zeta}_t^2(s,a)\right]} + \frac{3}{4}\epsilon,
\end{aligned}$$

with probability 1 for N large enough which completes the proof. □

Remark 1 (Remark 4.2) *The conditions in Proposition 4.8 and Proposition 4.9 are sufficient to*

prove that there exist approximately optimal decision rules $\hat{\pi}_t^*(s)$ that are ϵ-monotone on (i) s and (ii) t with probability 1 for N large enough, respectively.

Proof.

(i) From Assumption 4.2 and Proposition 4.7.3 in Puterman (2014) we have that $v_{t+1}(s)$ is nonincreasing in s. Since $\bar{p}_t(s'|s,a)$ is subadditive on $\mathcal{S} \times \mathcal{A}$ by Assumption 4.4, applying Lemma C.9 with $h(x, y^+, z^+) = p_t(s'|s^+, a^+)$, $h(x, y^-, z^-) = p_t(s'|s^-, a^-)$, $h(x, y^+, z^-) = p_t(s'|s^+, a^-)$, $h(x, y^-, z^+) = p_t(s'|s^-, a^+)$, and $g(x) = v_{t+1}(s')$ we conclude that $\sum_{s'} p_t(s'|s,a)v_{t+1}(s')$ is superadditive on $\mathcal{S} \times \mathcal{A}$. Because $\mathbb{E}[r_t(s,a,\omega)|s,a]$ is superadditive on $\mathcal{S} \times \mathcal{A}$ by Assumption 4.4 and the sum of subadditive functions is superadditive, it follows that $Q_t(s,a)$ is superadditive on $\mathcal{S} \times \mathcal{A}$. We get that $\pi_t^*(s)$ are nondecreasing in s from Lemma 4.7.1 in Puterman (2014). The convergence result follows from Theorem 4.1 and Definition 4.3.

(ii) By Assumption 4.4 we have that $\bar{p}_t(s'|s,a)$ is subadditive on $\mathcal{T} \times \mathcal{A}$. Further, by Assumption 4.2 and Proposition 4.7.3 in Puterman (2014) we have that $v_{t+1}(s)$ is nonincreasing in s. By Lemma C.9 with $h(x, y^+, z^+) = p_{t+1}(s'|s, a^+)$, $h(x, y^-, z^-) = p_t(s'|s, a^-)$, $h(x, y^+, z^-) = p_{t+1}(s'|s, a^-)$, $h(x, y^-, z^+) = p_t(s'|s, a^+)$, and $g(x) = v_{t+1}(s')$, we get that:

$$\sum_{s' \in \mathcal{S}} \left[p_{t+1}(s'|s,a^+) + p_t(s'|s,a^-)\right] v_{t+1}(s') \geq \sum_{s' \in \mathcal{S}} \left[p_{t+1}(s'|s,a^-) + p_t(s'|s,a^+)\right] v_{t+1}(s'),$$

for $a^+ \geq a^-$. Since $v_{t+1}(s') - v_{t+2}(s')$ is nondecreasing in s' by assumption and $\bar{p}_t(s'|s,a^-) \geq \bar{p}_t(s'|s,a^+)$ from Assumption 4.5, by Lemma C.4 it holds that $\sum_{s'} p_t(s'|s,a)v_{t+1}(s')$ is a superadditive function on $\mathcal{T} \times \mathcal{A}$. Because the sum of superadditive functions is superadditive, it follows that $Q_t(s,a)$ is superadditive on $\mathcal{T} \times \mathcal{A}$. We get that $\pi_t^*(s)$ are nondecreasing in t from Lemma 4.7.1 in Puterman (2014). The convergence result follows from Theorem 4.1 and Definition 4.3. □

Theorem 3 (Theorem 4.3.) *Suppose that assumptions 4.1, 4.2, 4.3, and 4.5 are satisfied. Then, $\Pi_t(s,\alpha)$ is an α-nonsignificant range of actions at state s and decision epoch t with probability 1 for N large enough.*

Proof. To show that $\Pi_t(s,\alpha)$ is a range of α-nonsignificant actions for a fixed state s, it suffices to prove that if:

$$\hat{Q}_t(s,a^*) - \hat{Q}_t(s,a^-) \leq \hat{d}_t(s,\alpha)\sqrt{M^{-1}\left[\hat{\zeta}_t^2(s,a^*) + \hat{\zeta}_t^2(s,a^-)\right]},$$

and

$$\hat{Q}_t(s,a^*) - \hat{Q}_t(s,a^+) \leq \hat{d}_t(s,\alpha)\sqrt{M^{-1}\left[\hat{\zeta}_t^2(s,a^*) + \hat{\zeta}_t^2(s,a^+)\right]},$$

then

$$\hat{Q}_t(s,a^*) - \hat{Q}_t(s,a') \leq \hat{d}_t(s,\alpha)\sqrt{M^{-1}\left[\hat{\zeta}_t^2(s,a^*) + \hat{\zeta}_t^2(s,a')\right]},$$

for $a^- \leq a' \leq a^+$. We demonstrate that the above will happen with probability 1 for N large enough. The proof proceeds by contradiction.

Suppose that $a^-, a^+ \in \Pi_t(s,\alpha)$ but $a' \notin \Pi_t(s,\alpha)$ with $a^- \leq a' \leq a^+$ for a fixed state s. We first show that $Q_t(s,a)$ is nondecreasing in a. By Assumption 4.2, we have that $v_t(s)$ is nonincreasing in s from Proposition 4.7.3 in Puterman (2014). Combining this result with the assumption that $\bar{p}_t(s'|s,a)$ is nonincreasing in a we can deduce that:

$$\begin{aligned}
\mathbb{E}\Big[v_{t+1}(s')|s,a^-\Big] &= \sum_{s'} p_t(s'|s,a^-)v_{t+1}(s') \\
&\leq \sum_{s'} p_t(s'|s,a')v_{t+1}(s') \\
&\leq \sum_{s'} p_t(s'|s,a^+)v_{t+1}(s') \\
&= \mathbb{E}\Big[v_{t+1}(s')|s,a^+\Big],
\end{aligned}$$

indicating that $\mathbb{E}\Big[v_{t+1}(s')|s,a\Big]$ is nondecreasing in a. Since $\mathbb{E}[r_t(s,a,\omega)|s,a]$ is nondecreasing a by Assumption 4.5, we then get that:

$$\begin{aligned}
Q_t(s,a^-) &= \mathbb{E}[r_t(s,a^-,\omega)|s,a^-] + \gamma\mathbb{E}\Big[v_{t+1}(s')|s,a^-\Big] \\
&\leq \mathbb{E}[r_t(s,a',\omega)|s,a'] + \gamma\mathbb{E}\Big[v_{t+1}(s')|s,a'\Big] \\
&\leq \mathbb{E}[r_t(s,a^+,\omega)|s,a^+] + \gamma\mathbb{E}\Big[v_{t+1}(s')|s,a^+\Big] \\
&= Q_t(s,a^+).
\end{aligned}$$

Thus, $Q_t(s,a)$ is nondecreasing in a. Moreover, by Theorem 4.1 we have that $\hat{Q}_t(s,a) \xrightarrow{a.s.} Q_t(s,a)$ uniformly on $\mathcal{A}$. Hence, we can find an N large enough such that $\hat{Q}_t(s,a^-) - \frac{1}{4}\epsilon \leq \hat{Q}_t(s,a') \leq \hat{Q}_t(s,a^+) + \frac{1}{4}\epsilon$ for $\epsilon > 0$.

We now show that $\sigma_t(s,a^-)^2 \geq \sigma_t(s,a')^2 \geq \sigma_t(s,a^+)^2$. By Lemma C.8, it holds that:

$$\sigma_t^2(s,a) = \mathbb{E}[r_t^2(s,a,\omega)|s,a] - \mathbb{E}[r_t(s,a,\omega)|s,a]^2 + \gamma^2\left(\mathbb{E}[v_{t+1}^2(s')|s,a] - \mathbb{E}[v_{t+1}(s')|s,a]^2\right).$$

From Assumption 4.3, it follows that $\mathbb{E}[r_t^2(s,a,\omega)|s,a] - \mathbb{E}[r_t(s,a,\omega)|s,a]^2$ and $\mathbb{E}[v_{t+1}^2(s')|s,a] - \mathbb{E}[v_{t+1}(s')|s,a]^2$ are nonincreasing in a. Thus, $\sigma_t^2(s,a)$ is nonincreasing in a and it holds that

$\sigma_t^2(s,a^-) \geq \sigma_t^2(s,a') \geq \sigma_t^2(s,a^+)$. Since $K\hat{\zeta}_t^2(s,a) \xrightarrow{a.s.} \sigma_t^2(s,a)$ by Lemma C.5, it follows that $K\hat{\zeta}_t^2(s,a)$ is $\frac{1}{4}\epsilon$-nondecreasing in a for large enough N. Hence, we have that $K\hat{\zeta}_t^2(s,a^-) + \frac{1}{4}\epsilon \geq K\hat{\zeta}_t^2(s,a') \geq K\hat{\zeta}_t^2(s,a^+) - \frac{1}{4}\epsilon$ for $\epsilon > 0$ and it follows that:

$$\begin{aligned}\hat{Q}_t(s,a^*) - \hat{Q}_t(s,a^-) - \frac{1}{4}\epsilon &\leq \hat{Q}_t(s,a^*) - \hat{Q}_t(s,a') \\ &\leq \hat{d}_t(s,\alpha)\sqrt{M^{-1}\left[\hat{\zeta}_t^2(s,a^*) + \hat{\zeta}_t^2(s,a')\right]} \\ &\leq \hat{d}_t(s,\alpha)\sqrt{M^{-1}\left[\hat{\zeta}_t^2(s,a^*) + \hat{\zeta}_t^2(s,a^-)\right]} + \frac{1}{4}\epsilon,\end{aligned}$$

and

$$\begin{aligned}\hat{Q}_t(s,a^*) - \hat{Q}_t(s,a^+) + \frac{1}{4}\epsilon &\geq \hat{Q}_t(s,a^*) - \hat{Q}_t(s,a') \\ &\geq \hat{d}_t(s,\alpha)\sqrt{M^{-1}\left[\hat{\zeta}_t^2(s,a^*) + \hat{\zeta}_t^2(s,a')\right]} \\ &\geq \hat{d}_t(s,\alpha)\sqrt{M^{-1}\left[\hat{\zeta}_t^2(s,a^*) + \hat{\zeta}_t^2(s,a^+)\right]} - \frac{1}{4}\epsilon,\end{aligned}$$

for any $\epsilon > 0$. Combining these results we find that $a' \in \Pi_t(s,\alpha)$, a contradiction. Consequently, it must hold that $a^-, a', a^+ \in \Pi_t(s,\alpha)$ and $\Pi_t(s,\alpha)$ is a range of actions. □

Proposition 10 (Proposition 4.10.) *Suppose assumptions 4.1, 4.2, 4.4 and 4.5 hold. Then, we have that (i) $\mathscr{A}_t^*(s,\tilde{a}) \subseteq \mathcal{A}_t^*(s)$ and (ii) $\Pi_t(s,\alpha,\tilde{a}) \subseteq \Pi_t(s,\alpha)$ for N large enough.*

The proof of this result depends on the following notation. Let

$$Q_t(s,a,\tilde{a}) \coloneqq \mathbb{E}\left[r_t(s,a,\omega) + \gamma Q_{t+1}(f_{t+1}(s,a,\omega),\tilde{a})|s,a,\tilde{a}\right]$$

denote the action-value function associated with state s and action a at decision epoch t, assuming that action $\tilde{a} \in \mathcal{A}_{t+1}(f_{t+1}(s,a,\omega))$ is taken at decision epoch $t+1$, and

$$\hat{Q}_t(s,a,\tilde{a}) \coloneqq \frac{1}{MK}\sum_{m=1}^{M} r_t(s,a,\omega^{m,k}) + \gamma\hat{Q}_{t+1}(f_{t+1}(s,a,\omega^{m,k}),\tilde{a})$$

denote its empirical estimate. Moreover, let $\mathscr{A}_t^*(s,\tilde{a}) \coloneqq \operatorname{argmax}_{a\in\mathcal{A}} Q_t(s,a,\tilde{a})$ and $\hat{\mathscr{A}}_t^*(s,\tilde{a}) \coloneqq \operatorname{argmax}_{a\in\mathcal{A}} \hat{Q}_t(s,a,\tilde{a})$.

In a similar way, we define the range of α-nonsignificant actions given that action $\tilde{a} \in \mathcal{A}_{t+1}(f_{t+1}(s,a,\omega))$ will be taken at $t+1$ as:

$$\Pi_t(s,\alpha,\tilde{a}) \coloneqq \left\{a \in \mathcal{A} : \hat{Q}_t(s,a^*,\tilde{a}) - \hat{Q}_t(s,a,\tilde{a}) \leq \hat{d}_t(s,\alpha,\tilde{a})\sqrt{M^{-1}\left[\hat{\zeta}_t^2(s,a^*,\tilde{a}) + \hat{\zeta}_t^2(s,a,\tilde{a})\right]}\right\},$$

where $\hat{d}_t(s, \alpha, \tilde{a})$ is the $1 - \alpha$ empirical quantile of the distribution of the maximum of the root statistics and

$$\hat{\varsigma}_t^2(s, a, \tilde{a}) := \frac{1}{M-1} \sum_{m=1}^{M} \left(\bar{Q}_t^m(s, a, \tilde{a}) - \hat{Q}_t(s, a, \tilde{a}) \right)^2,$$

with $\bar{Q}_t^m(s, a) := \frac{1}{K} \sum_{k=1}^{K} r_t(s, a, \omega^{m,k}) + \gamma Q_{t+1}(f_{t+1}(s, a, \omega^{m,k}), \tilde{a})$.

Proof of Proposition 4.10.

(i) We first show that $\mathscr{A}_t^*(s, \tilde{a}) \subseteq \mathcal{A}_t^*(s)$. Let $\bar{a} \in \mathcal{A}_t^-(s)$ and $a^* \in \mathcal{A}_t^*(s)$. We then have that:

$$\mathbb{E}\left[r_t(s, \bar{a}, \omega) + \gamma v_{t+1}(s')|s, \bar{a}\right] < \mathbb{E}\left[r_t(s, a^*, \omega) + \gamma v_{t+1}(s')|s, a^*\right],$$

for $s' = f_{t+1}(s, a, \omega)$. Since $\mathbb{E}\left[r_t(s, a, \omega) + \gamma v_{t+1}(s')|s, a\right]$ is equal to $\mathbb{E}\left[r_t(s, a, \omega)|s, a\right] + \gamma \mathbb{E}\left[v_{t+1}(s')|s, a\right]$ for all a, then we must have one of these cases: (1) $\mathbb{E}\left[r_t(s, \bar{a}, \omega)|s, \bar{a}\right] < \mathbb{E}\left[r_t(s, a^*, \omega)|s, a^*\right]$, (2) $\mathbb{E}\left[v_{t+1}(s')|s, \bar{a}\right] < \mathbb{E}\left[v_{t+1}(s')|s, a^*\right]$, or (3) $\mathbb{E}\left[r_t(s, \bar{a}, \omega)|s, \bar{a}\right] < \mathbb{E}\left[r_t(s, a^*, \omega)|s, a^*\right]$ and $\mathbb{E}\left[v_{t+1}(s')|s, \bar{a}\right] < \mathbb{E}\left[v_{t+1}(s')|s, a^*\right]$. We want to show that if:

$$\mathbb{E}\left[r_t(s, \bar{a}, \omega) + \gamma v_{t+1}(s')|s, \bar{a}\right] < \mathbb{E}\left[r_t(s, a^*, \omega) + \gamma v_{t+1}(s')|s, a^*\right],$$

then:

$$\mathbb{E}\left[r_t(s, \bar{a}, \omega) + \gamma Q_{t+1}(s', \tilde{a})|s, \bar{a}\right] < \mathbb{E}\left[r_t(s, a^*, \omega) + \gamma Q_{t+1}(s', \tilde{a})|s, a^*\right],$$

for $\tilde{a} \in \mathcal{A}_{t+1}(s')$. The proof proceeds by showing that $\mathscr{A}_t^*(s, \tilde{a}) \subseteq \mathcal{A}_t^*(s)$ in each case via contrapositive arguments.

Case (1): $\mathbb{E}\left[r_t(s, \bar{a}, \omega)|s, \bar{a}\right] < \mathbb{E}\left[r_t(s, a^*, \omega)|s, a^*\right]$. For any $\tilde{a} \in \mathcal{A}_{t+1}(s')$ we have:

$$\begin{aligned}
\mathbb{E}\left[r_t(s, \bar{a}, \omega) + \gamma Q_{t+1}(s', \tilde{a})|s, \bar{a}\right] &= \mathbb{E}\left[r_t(s, \bar{a}, \omega)|s, \bar{a}\right] + \gamma \mathbb{E}\left[Q_{t+1}(s', \tilde{a})|s, \bar{a}\right] \\
&< \mathbb{E}\left[r_t(s, a^*, \omega)|s, a^*\right] + \gamma \mathbb{E}\left[Q_{t+1}(s', \tilde{a})|s, \bar{a}\right] \\
&\leq \mathbb{E}\left[r_t(s, a^*, \omega)|s, a^*\right] + \gamma \mathbb{E}\left[Q_{t+1}(s', \tilde{a})|s, a^*\right] \\
&= \mathbb{E}\left[r_t(s, a^*, \omega) + \gamma Q_{t+1}(s', \tilde{a})|s, a^*\right], \qquad \text{(C.15)}
\end{aligned}$$

where the strict inequality holds since $\mathbb{E}\left[r_t(s, \bar{a}, \omega)|s, \bar{a}\right] < \mathbb{E}\left[r_t(s, a^*, \omega)|s, a^*\right]$ in this case and the not strict inequality follows because $\bar{p}_t(s'|s, a)$ is nonincreasing in a by Assumption 4.5 and $Q_{t+1}(s', \tilde{a})$ is nonincreasing in s' by Assumption 4.2 and Proposition 4.3. Thus, $\bar{a} \in \mathcal{A}_t^-(s)$ implies that $\bar{a} \in \mathscr{A}_t^-(s, \tilde{a})$. The inclusion $\mathscr{A}_t^*(s, \tilde{a}) \subseteq \mathcal{A}_t^*(s)$ follows.

Case (2): $\mathbb{E}\left[v_{t+1}(s')|s,\bar{a}\right] < \mathbb{E}\left[v_{t+1}(s')|s,a^*\right]$. Note that:

$$\mathbb{E}\left[v_{t+1}(s')|s,\bar{a}\right] < \mathbb{E}\left[v_{t+1}(s')|s,a^*\right]$$

implies that:

$$\mathbb{E}\left[v_{t+1}(s')|s,\bar{a}\right] - \xi < \mathbb{E}\left[v_{t+1}(s')|s,a^*\right] - \xi,$$

for any $\xi \in \mathbb{R}$. Let $\xi_{t+1}(s') \coloneqq v_{t+1}(s') - Q_{t+1}(s',\tilde{a})$ for all s'. We then have:

$$\begin{aligned}
&\mathbb{E}\left[v_{t+1}(s')|s,\bar{a}\right] < \mathbb{E}\left[v_{t+1}(s')|s,a^*\right] \\
&\Rightarrow \mathbb{E}\left[v_{t+1}(s')|s,\bar{a}\right] - \mathbb{E}\left[\xi_{t+1}(s')|s,\bar{a}\right] < \mathbb{E}\left[v_{t+1}(s')|s,a^*\right] - \mathbb{E}\left[\xi_{t+1}(s')|s,\bar{a}\right] \\
&\Rightarrow \sum_{s'\in\mathcal{S}} p_t(s'|s,\bar{a})[v_{t+1}(s') - \xi_{t+1}(s')] < \sum_{s'\in\mathcal{S}} p_t(s'|s,a^*)v_{t+1}(s') - \sum_{s'\in\mathcal{S}} p_t(s'|s,\bar{a})\xi_{t+1}(s') \\
&\Rightarrow \sum_{s'\in\mathcal{S}} p_t(s'|s,\bar{a})Q_{t+1}(s',\tilde{a}) < \sum_{s'\in\mathcal{S}} p_t(s'|s,a^*)v_{t+1}(s') - \sum_{s'\in\mathcal{S}} p_t(s'|s,\bar{a})\xi_{t+1}(s'). \qquad \text{(C.16)}
\end{aligned}$$

Since $\mathbb{E}[r_t(s,a^*,\omega)|s,a^*] - \mathbb{E}[r_t(s,a,\omega)|s,a]$ is nondecreasing in s and $\bar{p}_t(s'|s,a)$ is subadditive on $\mathcal{S}\times\mathcal{A}$ by Assumption 4.4, from Lemma C.9 we then get that $\xi_{t+1}(s')$ is nondecreasing in s' (see equation (C.13) in the proof of Proposition 4.8). Since $\bar{p}_t(s'|s,a)$ is nonincreasing in a by Assumption 4.5, applying Lemma C.4 with $h(x,y) = p_t(s'|s,a^*)$, $h(x,\bar{y}) = p_t(s'|s,\bar{a})$, and $g(x) = \xi_{t+1}(s')$ it follows that:

$$\sum_{s'\in\mathcal{S}} p_t(s'|s,a^*)\xi_{t+1}(s') \leq \sum_{s'\in\mathcal{S}} p_t(s'|s,\bar{a})\xi_{t+1}(s').$$

Hence, from equation (C.16) we have:

$$\begin{aligned}
\sum_{s'\in\mathcal{S}} p_t(s'|s,\bar{a})Q_{t+1}(s',\tilde{a}) &< \sum_{s'\in\mathcal{S}} p_t(s'|s,a^*)v_{t+1}(s') - \sum_{s'\in\mathcal{S}} p_t(s'|s,\bar{a})\xi_{t+1}(s') \\
&\leq \sum_{s'\in\mathcal{S}} p_t(s'|s,a^*)v_{t+1}(s') - \sum_{s'\in\mathcal{S}} p_t(s'|s,a^*)\xi_{t+1}(s') \\
&= \sum_{s'\in\mathcal{S}} p_t(s'|s,a^*)[v_{t+1}(s') - \xi_{t+1}(s')] \\
&= \sum_{s'\in\mathcal{S}} p_t(s'|s,a^*)Q_{t+1}(s',\tilde{a})
\end{aligned}$$

which implies that:

$$\mathbb{E}\left[Q_{t+1}(s',\tilde{a})|s,\bar{a}\right] < \mathbb{E}\left[Q_{t+1}(s',\tilde{a})|s,a^*\right]. \qquad \text{(C.17)}$$

We then get that:

$$\begin{aligned}\mathbb{E}\left[r_t(s,\bar{a},\omega)+\gamma Q_{t+1}(s',\tilde{a})|s,\bar{a}\right] &= \mathbb{E}\left[r_t(s,\bar{a},\omega)|s,\bar{a}\right]+\gamma\mathbb{E}\left[Q_{t+1}(s',\tilde{a})|s,\bar{a}\right]\\ &\leq \mathbb{E}\left[r_t(s,a^*,\omega)|s,a^*\right]+\gamma\mathbb{E}\left[Q_{t+1}(s',\tilde{a})|s,\bar{a}\right]\\ &< \mathbb{E}\left[r_t(s,a^*,\omega)|s,a^*\right]+\gamma\mathbb{E}\left[Q_{t+1}(s',\tilde{a})|s,a^*\right]\\ &= \mathbb{E}\left[r_t(s,a^*,\omega)+\gamma Q_{t+1}(s',\tilde{a})|s,a^*\right], \end{aligned} \tag{C.18}$$

where the not strict inequality follows because $\mathbb{E}\left[r_t(s,a,\omega)|s,a\right]$ is nondecreasing in a by Assumption 4.5 and the strict inequality follows from equation (C.17). Hence, $\bar{a}\in\mathcal{A}_t^{\cdot}(s)$ implies that $\bar{a}\in\mathscr{A}_t^{\cdot}(s,\tilde{a})$ and the inclusion $\mathscr{A}_t^*(s,\tilde{a})\subseteq\mathcal{A}_t^*(s)$ holds.

Case (3): $\mathbb{E}\left[r_t(s,\bar{a},\omega)|s,\bar{a}\right] < \mathbb{E}\left[r_t(s,a^*,\omega)|s,a^*\right]$ and $\mathbb{E}\left[v_{t+1}(s')|s,\bar{a}\right] < \mathbb{E}\left[v_{t+1}(s')|s,a^*\right]$. This case follows directly from cases 1 and 2.

Since $\mathscr{A}_t^*(s,\tilde{a})\subseteq\mathcal{A}_t^*(s)$ in all 3 cases, we have that $\mathcal{A}_t^*(s)\subseteq\mathscr{A}_t^*(s)$.

(ii) We now show that $\Pi_t(s,\alpha,\tilde{a})\subseteq\Pi_t(s,\alpha)$ via another contrapositive argument. Fix a realization of the sequence of the stochastic process $\boldsymbol{\omega}=(\omega^n : n\in\{1,\ldots,N\})$ and let $\bar{a}\notin\Pi_t(s,\alpha)$. Suppose that $\tilde{a}\in\mathcal{A}_{t+1}^{\cdot}(f_{t+1}(s,a,\omega))$. If $\tilde{a}\in\mathcal{A}_{t+1}^*(f_{t+1}(s,a,\omega))$ the result is trivially true. We want to prove that:

$$\hat{Q}_t(s,a^*)-\hat{Q}_t(s,\bar{a})>\hat{d}_t(s,\alpha)\sqrt{M^{-1}\left[\hat{\zeta}_t^2(s,a^*)+\hat{\zeta}_t^2(s,\bar{a})\right]}, \tag{C.19}$$

suggests that:

$$\hat{\boldsymbol{Q}}_t(s,a^*,\tilde{a})-\hat{\boldsymbol{Q}}_t(s,\bar{a},\tilde{a})>\hat{d}_t(s,\alpha,\tilde{a})\sqrt{M^{-1}\left[\hat{\varsigma}_t^2(s,a^*,\tilde{a})+\hat{\varsigma}_t^2(s,\bar{a},\tilde{a})\right]},$$

for any $\tilde{a}\in\mathcal{A}_{t+1}^{\cdot}(f_{t+1}(s,\bar{a},\omega))$. Suppose that $\mathbb{E}\left[r_t(s,a,\omega)|s,a\right]$ and $p_t(s'|s,a)$ are known for all s', s, a, and t. From the left-hand side of equation (C.19) we want to show that:

$$\boldsymbol{Q}_t(s,a^*,\tilde{a})-\boldsymbol{Q}_t(s,\bar{a},\tilde{a})\geq Q_t(s,a^*)-Q_t(s,\bar{a}).$$

Since

$$\boldsymbol{Q}_t(s,a,\tilde{a})=\mathbb{E}\left[r_t(s,a,\omega)+\gamma Q_{t+1}(s',\tilde{a})|s,a\right]=\mathbb{E}\left[r_t(s,a,\omega)|s,a\right]+\gamma\mathbb{E}\left[Q_{t+1}(s',\tilde{a})|s,a\right],$$

and

$$Q_t(s,a)=\mathbb{E}\left[r_t(s,a,\omega)+\gamma v_{t+1}(s')|s,a\right]=\mathbb{E}\left[r_t(s,a,\omega)|s,a\right]+\gamma\mathbb{E}\left[v_{t+1}(s')|s,a\right],$$

for $s' = f_{t+1}(s, a, \omega)$, $\tilde{a} \in \mathcal{A}_{t+1}(s')$, and all $a \in \mathcal{A}_t(s)$, it suffices to show that:

$$\mathbb{E}\left[Q_{t+1}(s', \tilde{a})|s, a^*\right] - \mathbb{E}\left[Q_{t+1}(s', \tilde{a})|s, \bar{a}\right] \geq \mathbb{E}\left[v_{t+1}(s')|s, a^*\right] - \mathbb{E}\left[v_{t+1}(s')|s, \bar{a}\right],$$

which implies:

$$\mathbb{E}\left[v_{t+1}(s')|s, \bar{a}\right] - \mathbb{E}\left[Q_{t+1}(s', \tilde{a})|s, \bar{a}\right] \geq \mathbb{E}\left[v_{t+1}(s')|s, a^*\right] - \mathbb{E}\left[Q_{t+1}(s', \tilde{a})|s, a^*\right],$$

and

$$\sum_{s'} p_t(s'|s, \bar{a}) \left[v_{t+1}(s') - Q_{t+1}(s', \tilde{a})\right] \geq \sum_{s'} p_t(s'|s, a^*) \left[v_{t+1}(s') - Q_{t+1}(s', \tilde{a})\right].$$

As $\mathbb{E}[r_t(s, a^*, \omega)|s, a^*] - \mathbb{E}[r_t(s, a, \omega)|s, a]$ is nondecreasing in s and $\bar{p}_t(s'|s, a)$ is subadditive on $\mathcal{S} \times \mathcal{A}$ by Assumption 4.4, from Lemma C.9 we get that $v_{t+1}(s') - Q_{t+1}(s', \tilde{a})$ is nondecreasing in s' (see equation (C.13) in the proof of Proposition 4.8). Because $\bar{p}_{t+1}(s'|s, a)$ is nonincreasing in a by Assumption 4.5, from Lemma C.4 it holds that:

$$\sum_{s'} p_t(s'|s, \bar{a}) \left[v_{t+1}(s') - Q_{t+1}(s', \tilde{a})\right] \geq \sum_{s'} p_t(s'|s, a^*) \left[v_{t+1}(s') - Q_{t+1}(s', \tilde{a})\right],$$

indicating that:

$$Q_t(s, a^*, \tilde{a}) - Q_t(s, \bar{a}, \tilde{a}) \geq Q_t(s, a^*) - Q_t(s, \bar{a}). \tag{C.20}$$

The convergence result follows from Theorem 4.1.

For the right-hand side of (C.19), note that $\hat{\zeta}_t^2(s, a) \geq \hat{\varsigma}_t^2(s, a)$ and $\hat{d}_t(s, \alpha) \geq \hat{d}_t(s, \alpha, a)$ for all a. Both inequalities follow because $\boldsymbol{\omega}$ is fixed, which implies that

$$r_t(s, a, \omega^{m,k}) + \gamma \hat{v}_{t+1}(f_{t+1}(s, a, \omega^{m,k})) \geq r_t(s, a, \omega^{m,k}) + \gamma \hat{Q}_{t+1}(f_{t+1}(s, a, \omega^{m,k}), \tilde{a}),$$

for every $\omega^{m,k} \in \boldsymbol{\omega}$. Thus, we get that:

$$\hat{d}_t(s, \alpha) \sqrt{M^{-1}\left[\hat{\zeta}_t^2(s, a^*) + \hat{\zeta}_t^2(s, \bar{a})\right]} \geq \hat{d}_t(s, \alpha, \tilde{a}) \sqrt{M^{-1}\left[\hat{\varsigma}_t^2(s, a^*) + \hat{\varsigma}_t^2(s, \bar{a})\right]}. \tag{C.21}$$

Note that this inequality holds for any $N \in \mathbb{N}$. Combining the results from (C.20) and (C.21) it follows that:

$$\hat{Q}_t(s, a^*, \tilde{a}) - \hat{Q}_t(s, \bar{a}, \tilde{a}) > \hat{d}_t(s, \alpha, \tilde{a}) \sqrt{M^{-1}\left[\hat{\varsigma}_t^2(s, a^*, \tilde{a}) + \hat{\varsigma}_t^2(s, \bar{a}, \tilde{a})\right]}.$$

□

C.3 Case Study Details

In this section, we provide additional methods or results from our hypertension treatment for primary prevention case study.

C.3.1 Progression of Risk Factors Over Time

We regress patients' untreated SBP, HDL, LDL, and TC on their age, sex, race, smoking status, and diabetes status (Table C.1). The intercept term of each regression model is adjusted by applying the difference between the linear regression fitted value and the observed value in the NHANES data. We select the models using stepwise variable selection.

Table C.1: Linear regression models coefficients.

Risk Factor	SBP	DBP	HDL	TC
Intercept	86.33	31.96	39.14	74.41
Age	0.85	1.72	0.153	4.50
Age Squared	0.00	-0.02	11.34	-0.04
Sex (Female)	-3.87	-2.94	1.21	6.93
Race (Black)	7.64	1.97	-3.11	-5.15
Smoking	1.82	-0.38	-8.82	6.44
Diabetes	4.07	-1.60	0.00	-20.90

C.3.2 Additional Policies

In this subsection, we describe the policies we considered as standards in our case study. Both policies are included for the purposes of comparing treatment strategies contained in our ranges of near-optimal treatment choices.

C.3.2.1 Optimal Treatment Plans

We use the standard backwards induction algorithm as described in Section 4.5 of Puterman (2014) to determine optimal hypertension treatment plans based on the standard formulation of our MDP $(\mathcal{T}, \mathcal{S}, \mathcal{A}, \boldsymbol{P}, \rho, \gamma)$ with transition probabilities $\boldsymbol{P}$ and reward function ρ. The standard backwards induction algorithm uses the following set of equations:

$$v_t(s) = \max_{a \in \mathcal{A}} \left\{ \sum_{s' \in \mathcal{S}} p_t(s'|s,a) \left[\rho_t(s',a) + \gamma v_{t+1}(s') \right] \right\},$$

for $s \in \mathcal{S}$ and $t \in \{0, 1, \ldots, 9\}$. Given the terminal condition:

$$V_{10}(s) = \begin{cases} \mathcal{L}(s) & \text{if } s \in \mathcal{H}, \\ 0 & \text{otherwise,} \end{cases}$$

we can find the optimal treatment strategy by recursively computing the value functions $v_t(s)$.

C.3.2.2 Current Clinical Guidelines Treatment Strategy

Our current clinical guidelines treatment strategy is based on the 2017 Hypertension Clinical Practice Guidelines (Whelton et al., 2018). In this strategy, any patient with a 10-year ASCVD risk of at least 10% and stage 1 hypertension receives an antihypertensive medication. Any patient with stage 2 hypertension receive at least two antihypertensive medications. If the patient does not reach a SBP below 130 and a DBP below 80 with the current treatment, we intensify the treatment up to 11 times in a year. This represent a clinician reassessing the BP levels of a patient after a month of prescribing a medication.

C.3.3 Convergence Analysis

To select the number of batches for each patient in our population, we first fixed the number of observations per batch to satisfy the conditions in Proposition 4.2 with $\beta = 0.001$. That is, each batch has $K = \left\lceil 2\kappa_t^2 \log(21/0.001) \right\rceil$ number of observations, where $\lceil x \rceil := \min\{y \in \mathbb{Z} | y \geq x\}$. Recall that at each year t and healthy state $s \in \mathcal{H}$ there is a total of $A = 21$ treatment choices. Using this approach, the sets of approximately optimal treatment choices identified in each batch are contained in the true sets of optimal actions with a probability of at least 99.9% for each patient. For the case that there is only one optimal treatment, the approximately optimal treatment is equal to the optimal with a probability of at least 99.9%. Note that this approach results in different number of observations K for each patient in the population. We then increase the number of batches (or simulation replicates) iteratively until the maximum width of the simultaneous confidence intervals reaches convergence.

By iteratively increasing the number of batches using $K = \left\lceil 2\kappa_t^2 \log(21/0.001) \right\rceil$ observations, we find that $M = 300$ batches (or independent simulations) may be enough to obtain a maximum confidence interval width close to the maximum width attained with 1,000 batches (Figure C.1). We also note that using $M = 300$ batches we achieve a maximum confidence interval width of 0.02 life-years. This width implies that any treatment choice

that results in less than 0.02 life-years than the control will be excluded from the range of actions.

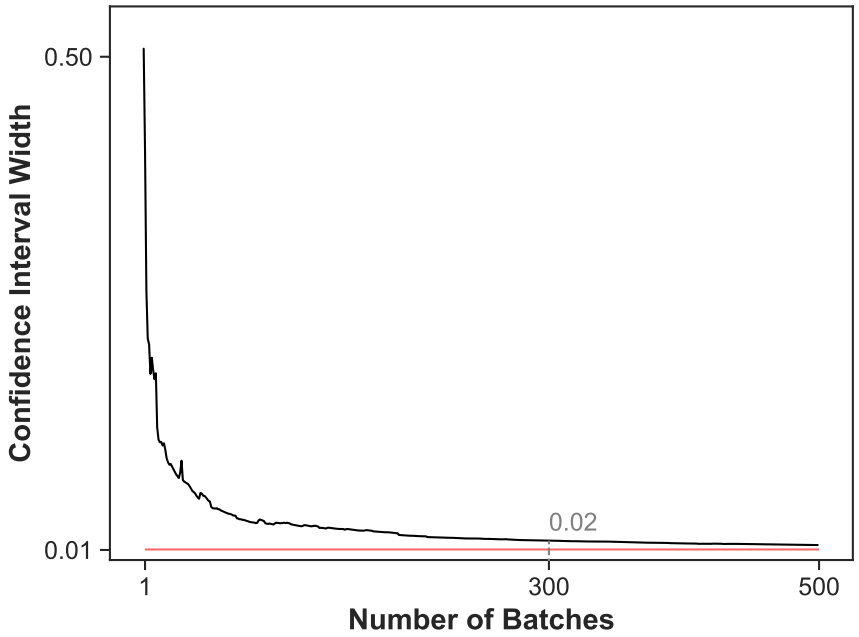


Figure C.1: Convergence of confidence interval width over the number of batches. Red line represents the confidence interval width using 1,000 batches (0.01).

C.3.4 Additional Results

This subsection presents additional results of our case study. All of the results included in this subsection have been described in Section 4.7 in the main body of the chapter.

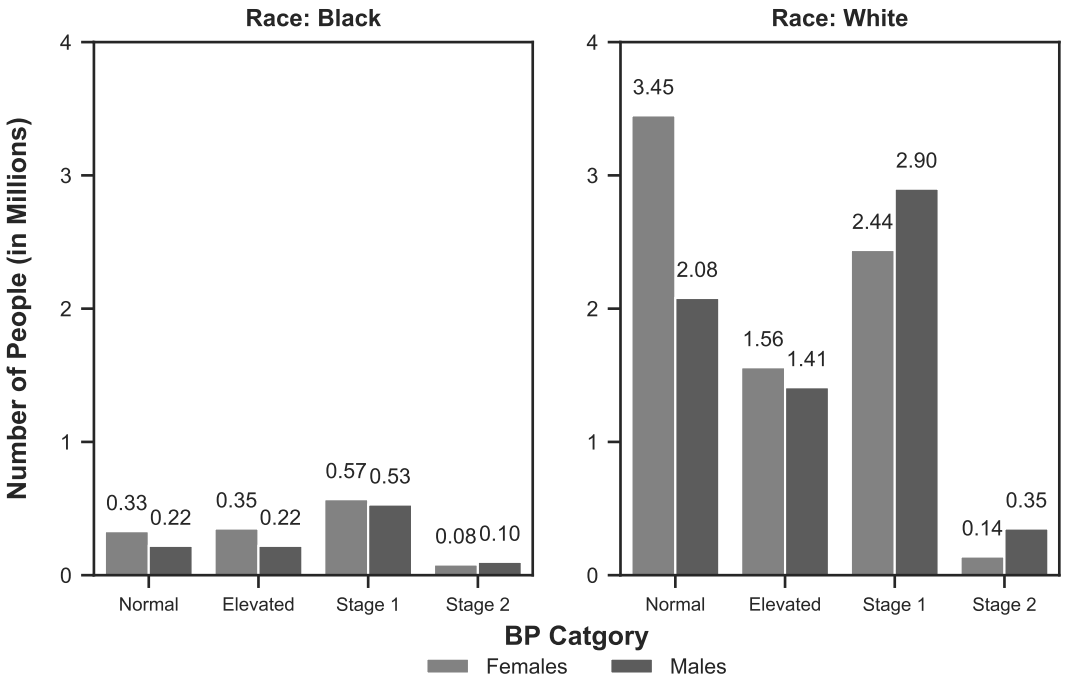


Figure C.2: Number of people by race, race, and BP category. BP groups are consistent with the BP categories of the 2017 Hypertension Clinical Practice Guidelines. The label "Elevated" denotes elevated BP, "Stage 1" denotes stage 1 hypertension, and the label "Stage 2" denotes stage 2 hypertension.

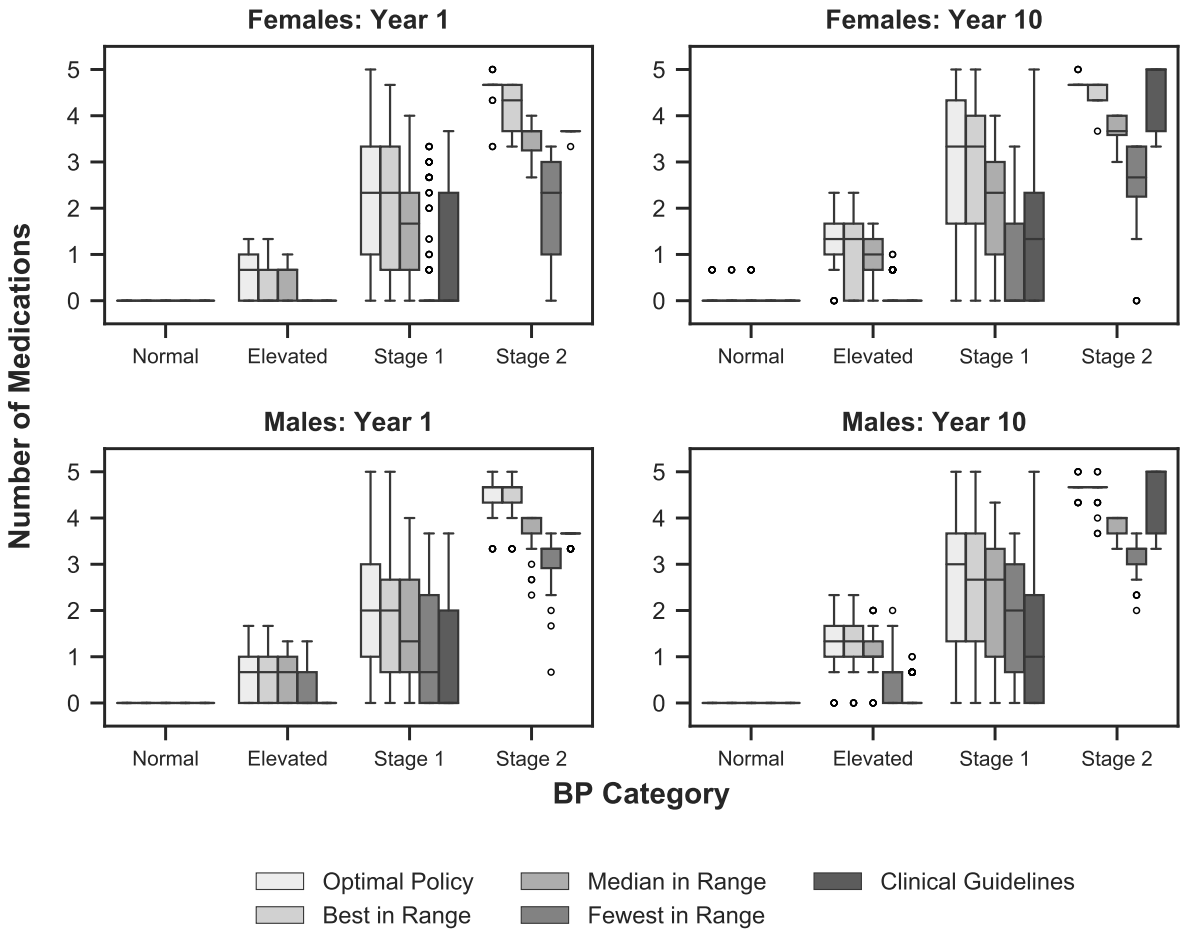


Figure C.3: Distribution of treatment at year 1 and year 10 of the study by sex. BP groups are consistent with the BP categories of the 2017 Hypertension Clinical Practice Guidelines. The label "Elevated" denotes elevated BP, "Stage 1" denotes stage 1 hypertension, and the label "Stage 2" denotes stage 2 hypertension.

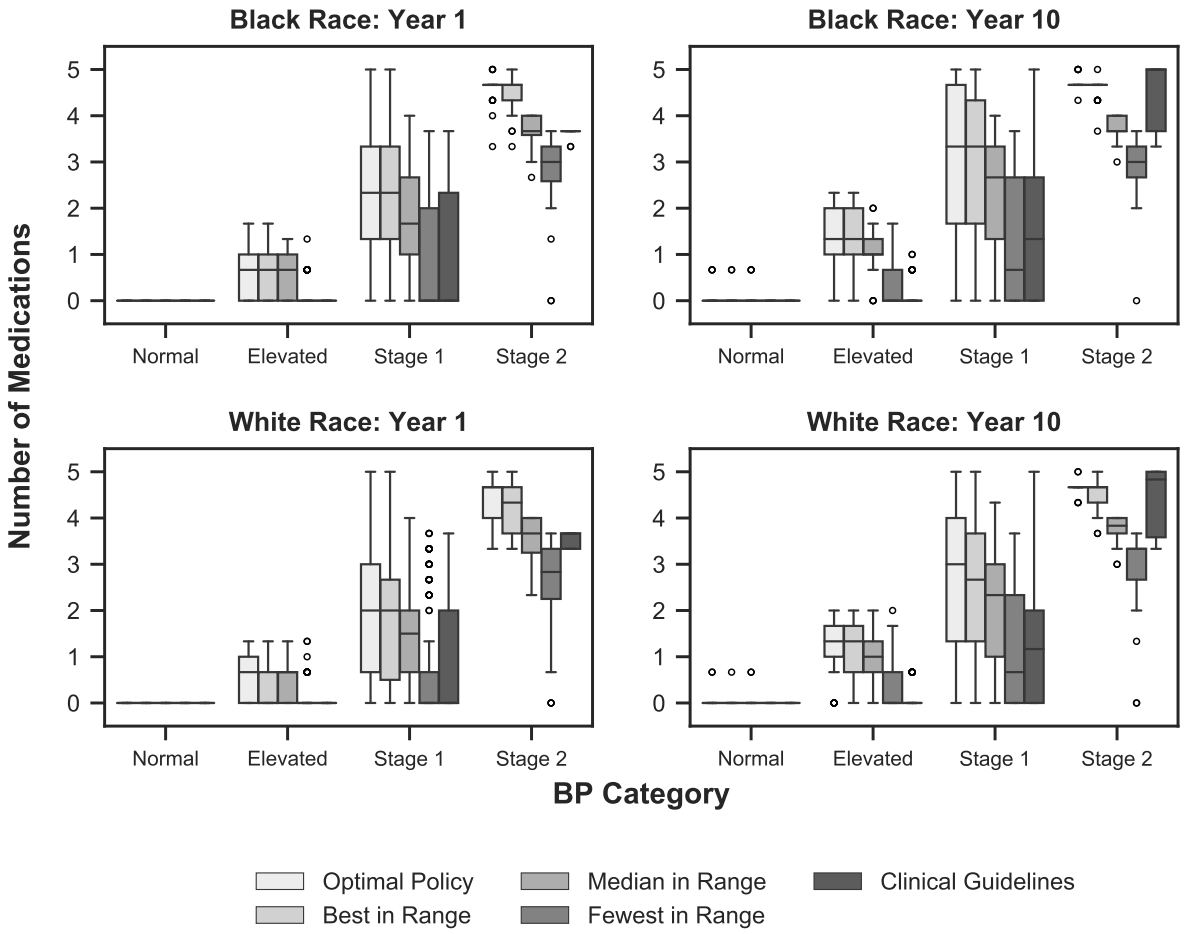


Figure C.4: Distribution of treatment at year 1 and year 10 of the study by race. BP groups are consistent with the BP categories of the 2017 Hypertension Clinical Practice Guidelines. The label "Elevated" denotes elevated BP, "Stage 1" denotes stage 1 hypertension, and the label "Stage 2" denotes stage 2 hypertension.

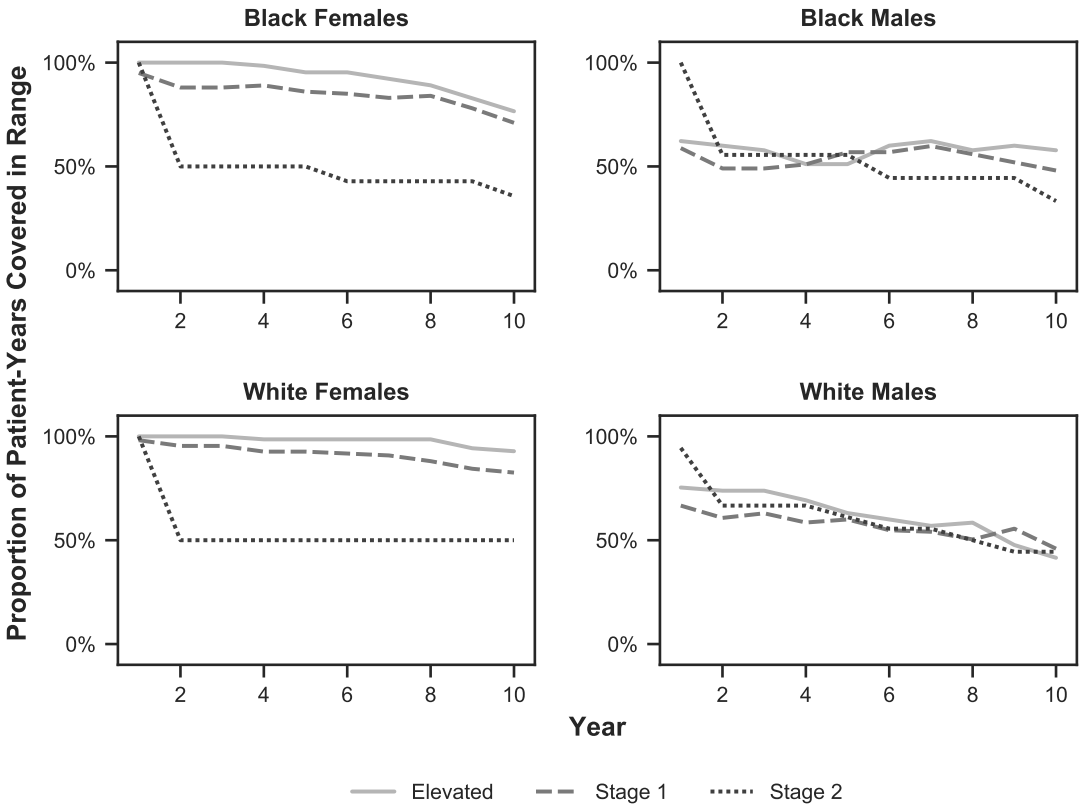


Figure C.5: Proportion of patients of treatment recommendations made by clinical guidelines contained in the ranges of near-optimal actions. BP groups are consistent with the BP categories of the 2017 Hypertension Clinical Practice Guidelines. The label "Elevated" denotes elevated BP, "Stage 1" denotes stage 1 hypertension, and the label "Stage 2" denotes stage 2 hypertension.

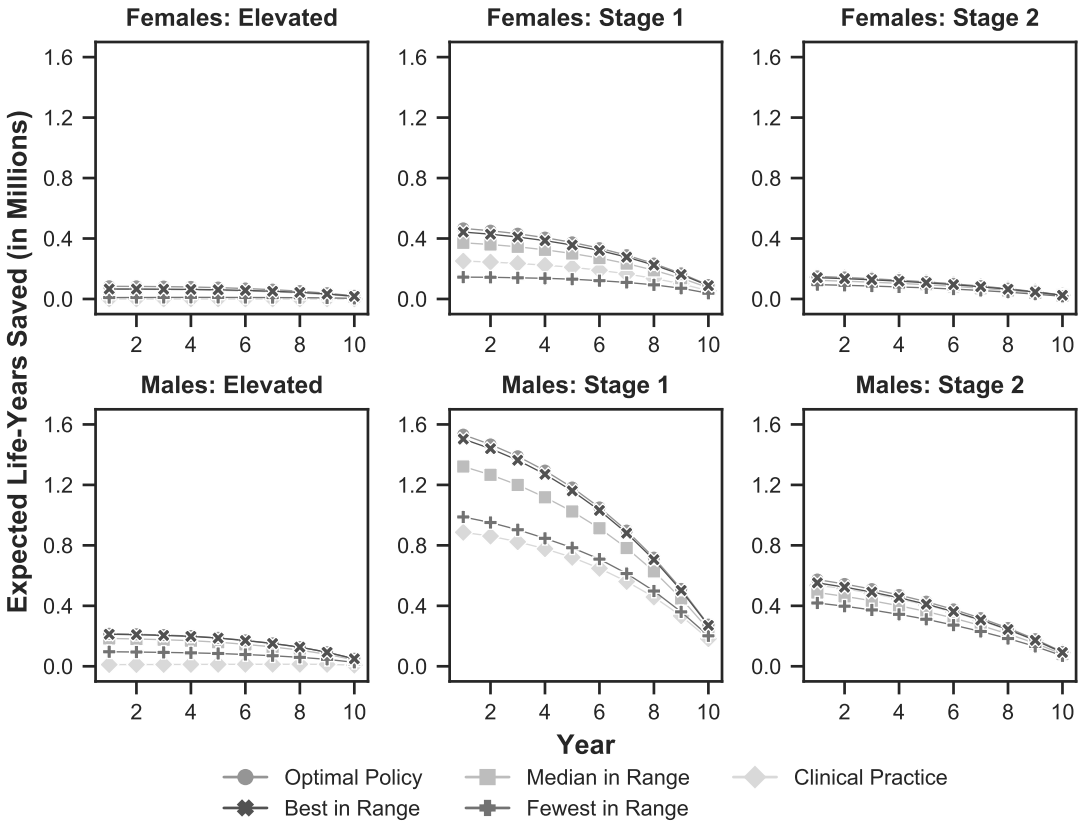


Figure C.6: Life-years saved by each treatment policy compared to no treatment per sex and BP group over the planning horizon. BP groups are consistent with the BP categories of the 2017 Hypertension Clinical Practice Guidelines. The label "Elevated" denotes elevated BP, "Stage 1" denotes stage 1 hypertension, and the label "Stage 2" denotes stage 2 hypertension.

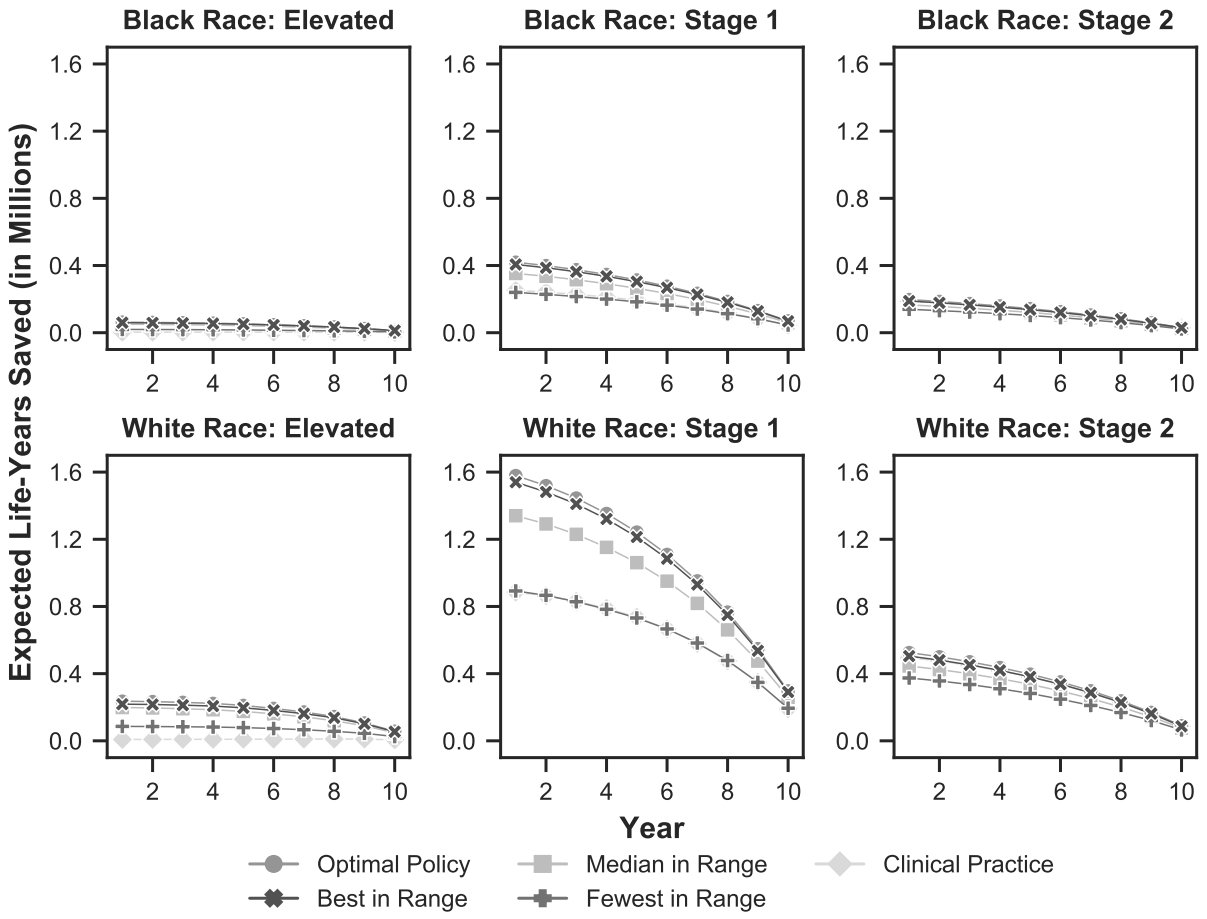


Figure C.7: Life-years saved by each treatment policy compared to no treatment per race and BP group over the planning horizon. BP groups are consistent with the BP categories of the 2017 Hypertension Clinical Practice Guidelines. The label "Elevated" denotes elevated BP, "Stage 1" denotes stage 1 hypertension, and the label "Stage 2" denotes stage 2 hypertension.

C.3.4.1 Results of Sensitivity Analyses

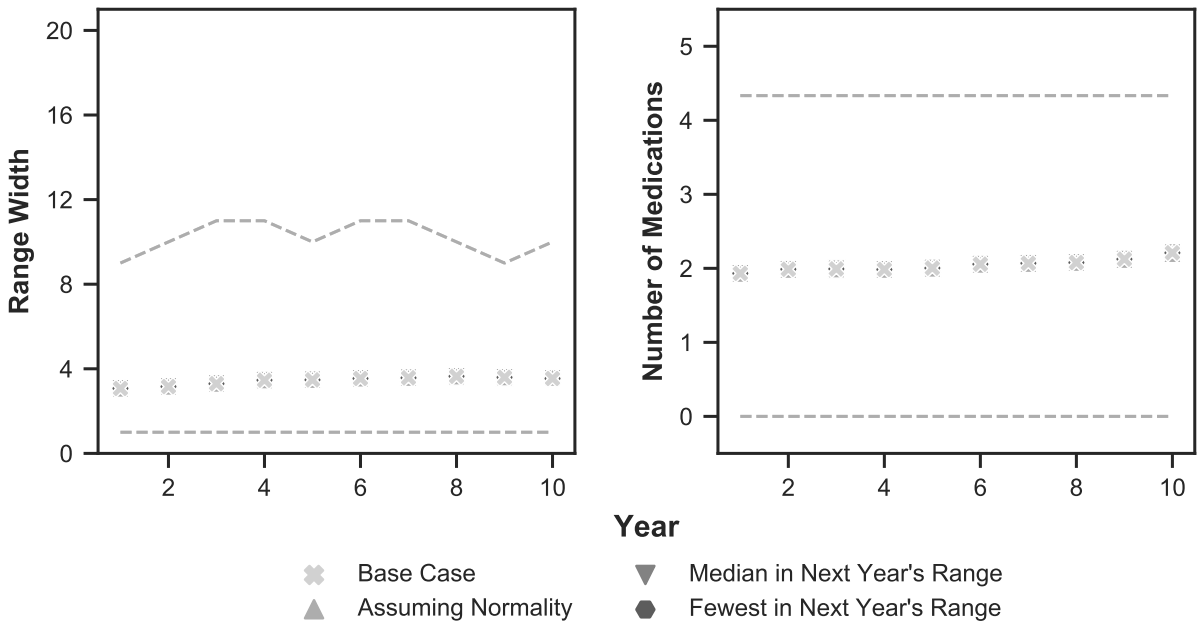


Figure C.8: Average range width and number of medications in base case, assuming normality in the action-value functions, using the action that corresponds to the median number of medications in next year's range, and using the action to corresponds to the fewest number of medications in next year's range. The label "'Median in Next Year's Range" denotes the median number of medications in next year's range and the label "Fewest in Next Year's Range" denotes the fewest number of medications in next year's range. Dashed lines represent the 5^{th} and 95^{th} quantile across the population of adults in the US with ages between 50 and 54.

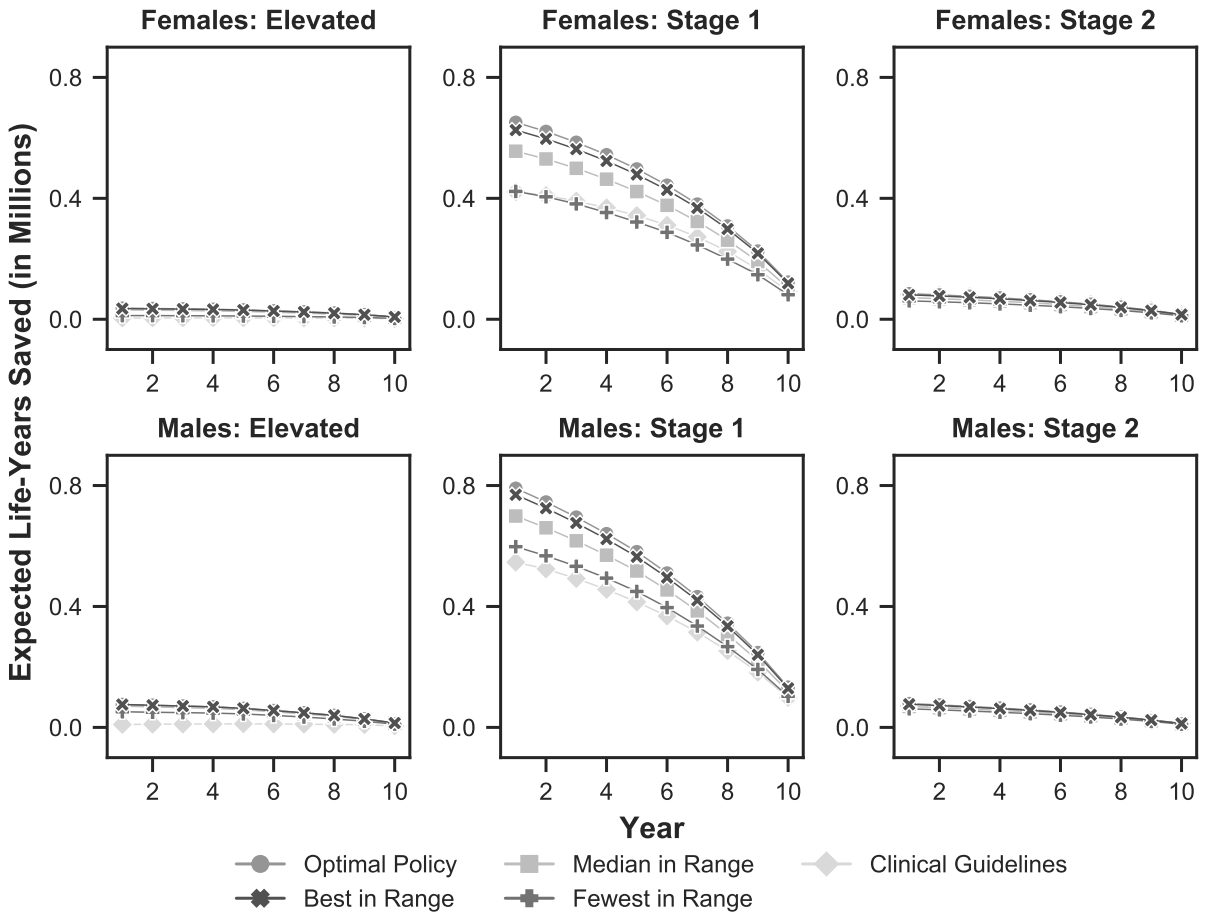


Figure C.9: Life-years saved by each treatment policy compared to no treatment per sex and BP group over the planning horizon in secondary population of adults with ages between 70 and 74. BP groups are consistent with the BP categories of the 2017 Hypertension Clinical Practice Guidelines. The label "Elevated" denotes elevated BP, "Stage 1" denotes stage 1 hypertension, and the label "Stage 2" denotes stage 2 hypertension.

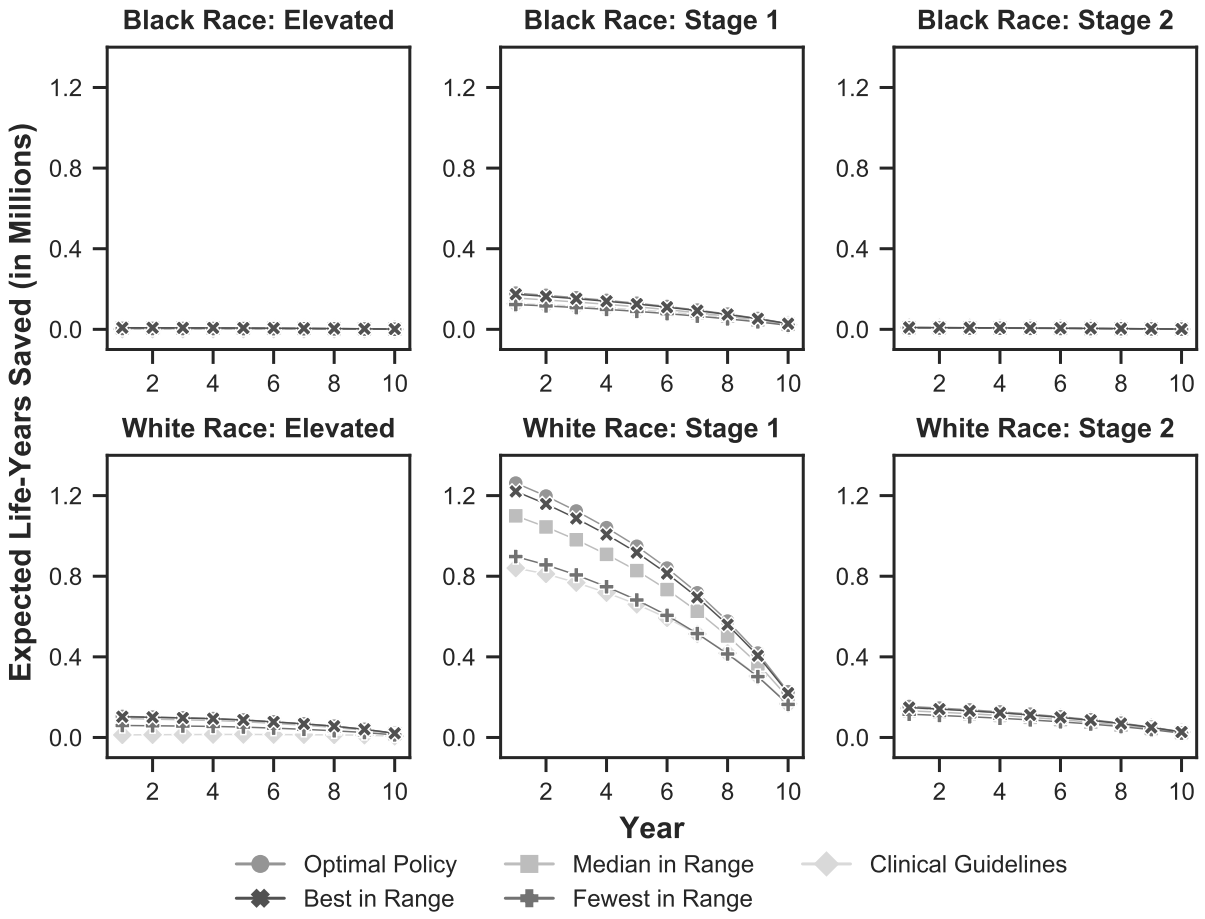


Figure C.10: Life-years saved by each treatment policy compared to no treatment per race and BP group over the planning horizon in secondary population of adults with ages between 70 and 74. BP groups are consistent with the BP categories of the 2017 Hypertension Clinical Practice Guidelines. The label "Elevated" denotes elevated BP, "Stage 1" denotes stage 1 hypertension, and the label "Stage 2" denotes stage 2 hypertension.

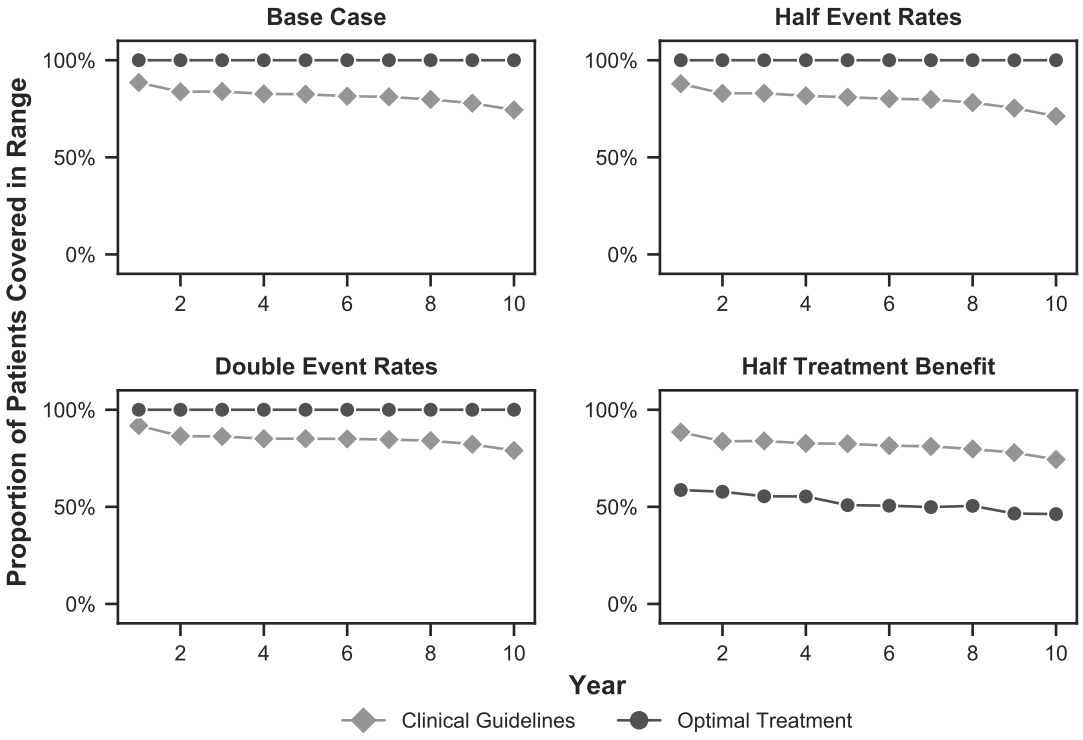


Figure C.11: Proportion of patients whose treatment recommendations are contained in the ranges of near-optimal actions despite parameter misestimation. Each panel represents a different misestimation scenario: no misestimation (top left), patients' true risk for ASCVD events is half the estimated risk (top right), patients' true risk for ASCVD events is double the estimated risk (bottom left), and patients' true benefit from treatment is half the estimated benefit (bottom right).

Bibliography

Abecassis, M. M., R. A. Fisher, K. M. Olthoff, C. E. Freise, D. R. Rodrigo, B. Samstein, I. Kam, and R. M. Merion (2012). Complications of living donor hepatic lobectomy-A comprehensive report. *American Journal of Transplantation 12*(5), 1208–1217.

Abellan, J., C. Armero, D. Conesa, J. Perez-Panades, M. Mart?nez-Beneito, O. Zurriaga, M. Garcia-Blasco, and H. Vanaclocha (2004). Predicting the Behaviour of the Renal Transplant Waiting List in the Pais Valencia (Spain) Using Simulation Modeling. In *Proceedings of the 2004 Winter Simulation Conference*, Volume 2, pp. 883–888. IEEE.

Abraham, G., A. S. Havulinna, O. G. Bhalala, S. G. Byars, A. M. De Livera, L. Yetukuri, E. Tikkanen, M. Perola, H. Schunkert, E. J. Sijbrands, A. Palotie, N. J. Samani, V. Salomaa, S. Ripatti, and M. Inouye (2016, nov). Genomic prediction of coronary heart disease. *European Heart Journal 37*(43), 3267–3278.

Abraham, G., A. Kowalczyk, J. Zobel, and M. Inouye (2013, feb). Performance and Robustness of Penalized and Unpenalized Methods for Genetic Prediction of Complex Human Disease. *Genetic Epidemiology 37*(2), 184–195.

Abraham, G., J. A. Tye-Din, O. G. Bhalala, A. Kowalczyk, J. Zobel, and M. Inouye (2014, feb). Accurate and Robust Genomic Prediction of Celiac Disease Using Statistical Learning. *PLoS Genetics 10*(2), e1004137.

Afdhal, N., S. Zeuzem, P. Kwo, M. Chojkier, N. Gitlin, M. Puoti, M. Romero-Gomez, J. P. Zarski, K. Agarwal, P. Buggisch, G. R. Foster, N. Bräu, M. Buti, I. M. Jacobson, G. M. Subramanian, X. Ding, H. Mo, J. C. Yang, P. S. Pang, W. T. Symonds, J. G. McHutchison, A. J. Muir, A. Mangia, and P. Marcellin (2014). Ledipasvir and sofosbuvir for untreated HCV genotype 1 infection. *New England Journal of Medicine 370*(20), 1889–1898.

Agnihothri, S., L. Cui, M. Delasay, and B. Rajan (2018, oct). The value of mHealth for managing chronic conditions. *Health Care Management Science*.

Agopian, V. G., F. M. Kaldas, J. C. Hong, M. Whittaker, C. Holt, A. Rana, A. Zarrinpar, H. Petrowsky, D. Farmer, H. Yersiz, V. Xia, J. R. Hiatt, and R. W. Busuttil (2012). Liver transplantation for nonalcoholic steatohepatitis: The new epidemic. *Annals of Surgery 256*(4), 624–633.

Ajmera, V., E. R. Perito, N. M. Bass, N. A. Terrault, K. P. Yates, R. Gill, R. Loomba, A. M. Diehl, and B. E. Aouizerat (2017). Novel plasma biomarkers associated with liver disease severity in adults with nonalcoholic fatty liver disease. *Hepatology 65*(1), 65–77.

Akan, M., O. Alagoz, B. Ata, F. S. Erenay, and A. Said (2012). A Broader View of Designing the Liver Allocation System. *Operations Research 60*(4), 757–770.

Alagoz, O., L. M. Maillart, A. J. Schaefer, and M. S. Roberts (2004). The Optimal Timing of Living-Donor Liver Transplantation. *Management Science 50*(10), 1420–1430.

Alagoz, O., L. M. Maillart, a. J. Schaefer, and M. S. Roberts (2007a). Choosing Among Living-Donor and Cadaveric Livers. *Management Science 53*(11), 1702–1715.

Alagoz, O., L. M. Maillart, a. J. Schaefer, and M. S. Roberts (2007b). Determining the Acceptance of Cadaveric Livers Using an Implicit Model of the Waiting List. *Operations Research 55*(1), 24–36.

Alagoz, O., A. J. Schaefer, and M. S. Roberts (2009). Optimizing Organ Allocation and Acceptance. In *Handbook of Optimization in Medicine*, Number C, pp. 1–24.

Alexopoulos, C. and D. Goldsman (2004). To batch or not to batch? *ACM Transactions on Modeling and Computer Simulation 14*(1), 76–114.

Alexopoulos, C. and A. F. Seila (1996). Implementing the batch means method in simulation experiments. *Winter Simulation Conference Proceedings*, 214–221.

Aprahamian, H., D. R. Bish, and E. K. Bish (2019, sep). Optimal Risk-Based Group Testing. *Management Science 65*(9), 4365–4384.

Aprahamian, H., D. R. Bish, and E. K. Bish (2020). Optimal Group Testing: Structural Properties and Robust Solutions, with Application to Public Health Screening. *INFORMS Journal on Computing* (October).

Arias, E., M. Heron, and J. Xu (2017). United States Life Tables, 2013. *National Vital Statistics Reports 66*(3), 1–64.

Arias, E. and J. Xu (2019). United States Life Tables, 2017. *National Vital Statistics Reports 68*(7).

Axelrod, D. A., M. K. Guidinger, S. Finlayson, D. E. Schaubel, D. C. Goodman, M. Chobanian, and R. M. Merion (2008). Rates of solid-organ wait-listing, transplantation, and survival among residents of rural and urban areas. *JAMA - Journal of the American Medical Association 299*(2), 202–207.

Axelrod, D. A., D. Millman, and M. M. Abecassis (2010). US health care reform and transplantation. Part I: Overview and impact on access and reimbursement in the private sector: Personal viewpoint. *American Journal of Transplantation 10*(10), 2197–2202.

Ayer, T., O. Alagoz, N. K. Stout, and E. S. Burnside (2016). Heterogeneity in women's adherence and its role in optimal breast cancer screening policies. *Management Science 62*(5), 1339–1362.

Ayer, T., C. Zhang, A. Bonifonte, A. C. Spaulding, and J. Chhatwal (2019). Prioritizing hepatitis C treatment in U.S. Prisons. *Operations Research 67*(3), 853–873.

Bardach, N. S., J. J. Wang, S. F. De Leon, S. C. Shih, W. J. Boscardin, L. E. Goldman, and R. A. Dudley (2013). Effect of pay-for-performance incentives on quality of care in small practices with electronic health records: A randomized trial. *JAMA - Journal of the American Medical Association 310*(10), 1051–1059.

Bartle, R. G. and D. R. Sherbert (2010). *Introduction to Real Analysis.* (4 ed.). John Wiley & Sons.

Belli, L. S., M. Berenguer, P. A. Cortesi, M. Strazzabosco, S. R. Rockenschaub, S. Martini, C. Morelli, F. Donato, R. Volpes, G. P. Pageaux, A. Coilly, S. Fagiuoli, G. Amaddeo, G. Perricone, C. Vinaixa, G. Berlakovich, R. Facchetti, W. Polak, P. Muiesan, and C. Duvoux (2016). Delisting of liver transplant candidates with chronic hepatitis C after viral eradication: A European study. *Journal of Hepatology 65*(3), 524–531.

Ben-Tal, A., L. E. Ghaoui, and A. Nemirovski (2009). *Robust optimization*. Princeton University Press.

Benjamin, E. J., M. J. Blaha, S. E. Chiuve, M. Cushman, S. R. Das, R. Deo, S. D. de Ferranti, J. Floyd, M. Fornage, C. Gillespie, C. R. Isasi, M. C. Jiménez, L. C. Jordan, S. E. Judd, D. Lackland, J. H. Lichtman, L. Lisabeth, S. Liu, C. T. Longenecker, R. H. Mackey, K. Matsushita, D. Mozaffarian, M. E. Mussolino, K. Nasir, R. W. Neumar, L. Palaniappan, D. K. Pandey, R. R. Thiagarajan, M. J. Reeves, M. Ritchey, C. J. Rodriguez, G. A. Roth, W. D. Rosamond, C. Sasson, A. Towfighi, C. W. Tsao, M. B. Turner, S. S. Virani, J. H. Voeks, J. Z. Willey, J. T. Wilkins, J. H. Wu, H. M. Alger, S. S. Wong, and P. Muntner (2017, mar). Heart Disease and Stroke Statistics—2017 Update: A Report From the American Heart Association. *Circulation 135*(10), e146–e603.

Benjamin, E. J., S. S. Virani, C. W. Callaway, A. M. Chamberlain, A. R. Chang, S. Cheng, S. E. Chiuve, M. Cushman, F. N. Delling, R. Deo, S. D. de Ferranti, J. F. Ferguson, M. Fornage, C. Gillespie, C. R. Isasi, M. C. Jiménez, L. C. Jordan, S. E. Judd, D. Lackland, J. H. Lichtman, L. Lisabeth, S. Liu, C. T. Longenecker, P. L. Lutsey, J. S. Mackey, D. B. Matchar, K. Matsushita, M. E. Mussolino, K. Nasir, M. O'Flaherty, L. P. Palaniappan, A. Pandey, D. K. Pandey, M. J. Reeves, M. D. Ritchey, C. J. Rodriguez, G. A. Roth, W. D. Rosamond, U. K. Sampson, G. M. Satou, S. H. Shah, N. L. Spartano, D. L. Tirschwell, C. W. Tsao, J. H. Voeks, J. Z. Willey, J. T. Wilkins, J. H. Wu, H. M. Alger, S. S. Wong, and P. Muntner (2018, mar). Heart disease and stroke statistics - 2018 update: a report from the American Heart Association. *Circulation 137*(12), E67–E492.

Bentkus, V. (2005, jan). A Lyapunov-type Bound in Rd. *Theory of Probability & Its Applications 49*(2), 311–323.

Beran, R. (1988). Balanced simultaneous confidence sets. *Journal of the American Statistical Association 83*(403), 679–686.

Bertsekas, D. P. (2012). *Dynamic programming and optimal control* (4 ed.), Volume 2. Belmont, MA: Athena scientific.

Bibbins-Domingo, K., D. C. Grossman, S. J. Curry, K. W. Davidson, J. W. Epling, F. A. R. García, M. W. Gillman, A. R. Kemper, A. H. Krist, A. E. Kurth, C. S. Landefeld, M. L. LeFevre, C. M. Mangione, W. R. Phillips, D. K. Owens, M. G. Phipps, and M. P. Pignone (2016). Statin use for the primary prevention of cardiovascular disease in adults. *Journal of the American Medical Association 316*(19), 1997–2007.

Billingsley, P. (1995). *Probability and measure theory*. John Wiley and Sons.

Bodner, T. E. (2008). What improves with increased missing data imputations? *Structural Equation Modeling 15*(4), 651–675.

Borovicka, T., M. J. Jr, P. Kordik, and M. Jirina (2012). Selecting Representative Data Sets. In *Advances in Data Mining Knowledge Discovery and Applications*, pp. 43–70.

Boucherie, R. J. and N. M. van Dijk (2017). *Markov Decision Processes in Practice*, Volume 248 of *International Series in Operations Research & Management Science*. Cham: Springer International Publishing.

Breiman, L. (1996). Bagging Predictors. *Machine Learning 24*(421), 123–140.

Breiman, L., J. Friedman, C. Stone, and R. Olshen (1984). Classification and regression trees.

Bretz, F. and L. A. Hothorn (2003, jan). Comparison of Exact and Resampling Based Multiple Testing Procedures. *Communications in Statistics - Simulation and Computation 32*(2), 461–473.

Brønnnum-Hansen, H., T. Jørgensen, M. Davidsen, M. Madsen, M. Osler, L. U. Gerdes, and M. Schroll (2001, dec). Survival and cause of death after myocardial infarction: the Danish MONICA study. *Journal of Clinical Epidemiology 54*(12), 1244–1250.

Burn, J., M. Dennis, J. Bamford, P. Sandercock, D. Wade, and C. Warlow (1994, feb). Long-term risk of recurrent stroke after a first-ever stroke. The Oxfordshire Community Stroke Project. *Stroke 25*(2), 333–7.

Cabana, M. D., C. S. Rand, N. R. Powe, A. W. Wu, M. H. Wilson, P.-A. C. Abboud, and H. R. Rubin (1999, oct). Why Don't Physicians Follow Clinical Practice Guidelines? *Journal of the American Medical Association 282*(15), 1458.

Calikoglu, S., R. Murray, and D. Feeney (2012, dec). Hospital Pay-For-Performance Programs In Maryland Produced Strong Results, Including Reduced Hospital-Acquired Conditions. *Health Affairs 31*(12), 2649–2658.

Cameron, A. C. and P. K. Trivedi (1990). Regression-based tests for overdispersion in the Poisson model. *Journal of Econometrics 46*(3), 347–364.

Capan, M., A. Khojandi, B. T. Denton, K. D. Williams, T. Ayer, J. Chhatwal, M. Kurt, J. M. Lobo, M. S. Roberts, G. Zaric, S. Zhang, and J. S. Schwartz (2017). From data to improved decisions: operations research in healthcare delivery. *Medical Decision Making 37*(8), 849–859.

Caruana, R. and A. Niculescu-Mizil (2006). An empirical comparison of supervised learning algorithms. *ACM International Conference Proceeding Series 148*, 161–168.

Casella, G. and R. Beroer (2001). *Statistical lnference* (2 ed.). Belmont , California: Duxbury Press.

Centers for Disease Control and Prevention (2020). National Health and Nutrition Examination Survey Data.

Centers for Medicare & Medicaid Services and United States Department of Health and Human Services (2006). Medicare and Medicaid Programs; Conditions for Coverage for Organ Procurement Organizations (OPOs); Final Rule. *Federal Register 71*(104), 30981–31054.

Centers for Medicare & Medicaid Services and United States Department of Health and Human Services (2019). Medicare and Medicaid Programs; Organ Procurement Organizations Conditions for Coverage: Revisions to the Outcome Measure Requirements for Organ Procurement Organization. *Federal Register 84*(246), 70628–70710.

Chan, T., C. Narasimhan, and Y. Xie (2013, jun). Treatment effectiveness and side effects: a model of physician learning. *Management Science 59*(6), 1309–1325.

Chanchaichujit, J., A. Tan, F. Meng, and S. Eaimkhong (2019). *Optimization, Simulation and Predictive Analytics in Healthcare*, pp. 95–121. Singapore: Springer Singapore.

Chang, H. S., J. Hu, M. C. Fu, and S. I. Marcus (2013). *Simulation-Based Algorithms for Markov Decision Processes*. Communications and Control Engineering. Springer London.

Chehrazi, N., L. E. Cipriano, and E. A. Enns (2019). Dynamics of drug resistance: Optimal control of an infectious disease. *Operations Research 67*(3), 619–650.

Chen, Q., T. Ayer, and J. Chhatwal (2018, jun). Optimal M-switch surveillance policies for liver cancer in a hepatitis C–infected population. *Operations Research 66*(3), 673–696.

Chhatwal, J., O. Alagoz, and E. S. Burnside (2010). Optimal Breast Biopsy Decision-Making Based on Mammographic Features and Demographic Factors. *Operations Research 58*(6), 1577–1591.

Chinneck, J. W. and K. Ramadan (2000). Linear Programming with Interval Coefficients. *The Journal of the Operational Research Society 51*(2), 209–220.

Chipman, H. A., E. I. George, and R. E. McCulloch (2012). BART: Bayesian additive regression trees. *Annals of Applied Statistics 6*(1), 266–298.

Chirikos, T. N. (2003). Appraising the economic efficiency of cancer treatment: an exploratory analysis of lung cancer. *Health Care Management Science 6*(2), 87–95.

Chobanian, A. V., G. L. Bakris, H. R. Black, W. C. Cushman, L. A. Green, J. L. Izzo, D. W. Jones, B. J. Materson, S. Oparil, J. T. Wright, and E. J. Roccella (2003). Seventh report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure. *Hypertension 42*(6), 1206–1252.

Cipriano, L. E. and T. A. Weber (2018). Population-level intervention and information collection in dynamic healthcare policy. *Health Care Management Science 21*(4), 604–631.

Classen, D. C. and L. A. Mermel (2015). Specialty Society Clinical Practice Guidelines Time for Evolution or Revolution? *The Journal of the American Medical Association 314*(9), 871–872.

Claxton, A. J., J. Cramer, and C. Pierce (2001). A systematic review of the associations between dose regimens and medication compliance. *Clinical Therapeutics 23*(8), 1296–1310.

Claxton, K., P. J. Neumann, S. Araki, and M. C. Weinstein (2001). Bayesian value-of-information analysis: an application to a policy model of Alzheimer's disease. *International Journal of Technology Assessment in Health Care 17*(1), 38–55.

Cohen, J. B. and R. R. Townsend (2018, feb). The ACC/AHA 2017 Hypertension Guidelines: Both Too Much and Not Enough of a Good Thing? *Annals of Internal Medicine 168*(4), 287.

Collins, R., C. Reith, J. Emberson, J. Armitage, C. Baigent, L. Blackwell, R. Blumenthal, J. Danesh, G. D. Smith, D. DeMets, S. Evans, M. Law, S. MacMahon, S. Martin, B. Neal, N. Poulter, D. Preiss, P. Ridker, I. Roberts, A. Rodgers, P. Sandercock, K. Schulz, P. Sever, J. Simes, L. Smeeth, N. Wald, S. Yusuf, and R. Peto (2016). Interpretation of the evidence for the efficacy and safety of statin therapy. *The Lancet 388*(10059), 2532–2561.

Color Genomics (2018). Genetic testing for inherited heart conditions.

Connor, M. J.-o., M. B. Bch, and P. Natarajan (2020). Current Clinical Implications of Coronary Artery Disease Polygenic Risk Scoring. pp. 4–11.

Cooper, K., S. C. Brailsford, R. Davies, and J. Raftery (2006). A review of health care models for coronary heart disease interventions. *Health Care Management Science 9*(4), 311–324.

Cruz-Ramírez, M., C. Hervás-Martínez, J. C. Fernández, J. Briceño, and M. de la Mata (2013). Predicting patient survival after liver transplantation using evolutionary multi-objective artificial neural networks. *Artificial Intelligence in Medicine 58*(1), 37–49.

Decruyenaere, A., P. Decruyenaere, P. Peeters, F. Vermassen, T. Dhaene, and I. Couckuyt (2015). Prediction of delayed graft function after kidney transplantation: comparison between logistic regression and machine learning methods. *BMC medical informatics and decision making 15*(1), 83.

Dennis, M. S., J. P. Burn, P. A. Sandercock, J. M. Bamford, D. T. Wade, and C. P. Warlow (1993). Long-term survival after first-ever stroke: the Oxfordshire Community Stroke Project. *Stroke 24*(6), 796–800.

Denton, B. T., O. Alagoz, A. Holder, and E. K. Lee (2011). Medical decision making: open research challenges.

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

Deo, S., K. Rajaram, S. Rath, U. S. Karmarkar, and M. B. Goetz (2015, apr). Planning for HIV screening, testing, and care at the veterans health administration. *Operations Research 63*(2), 287–304.

DeRoos, L. J., W. J. Marrero, E. B. Tapper, C. J. Sonnenday, M. S. Lavieri, D. W. Hutton, and N. D. Parikh (2019). Estimated Association Between Organ Availability and Presumed Consent in Solid Organ Transplant. *JAMA Network Open 2*(10), e1912431.

Dienstag, J. L. and A. B. Cosimi (2012, oct). Liver Transplantation — A Vision Realized. *New England Journal of Medicine 367*(16), 1483–1485.

Dong, H., D. Coyle, and M. Buxton (2007). Value of information analysis for a new technology: computer-assisted total knee replacement. *International Journal of Technology Assessment in Health Care, 3*(3), 337–342.

Downs, J. R. and P. G. O'Malley (2015). Management of dyslipidemia for cardiovascular disease risk reduction: Synopsis of the 2014 U.S. department of veterans affairs and U.S. department of defense clinical practice guideline. *Annals of Internal Medicine 163*(4), 291–297.

Doycheva, I., K. D. Watt, and N. Alkhouri (2017). Nonalcoholic fatty liver disease in adolescents and young adults: The next frontier in the epidemic. *Hepatology 65*(6), 2100–2109.

Doyle, M. B., N. Vachharajani, J. R. Wellen, J. A. Lowell, S. Shenoy, G. Ridolfi, M. D. Jendrisak, J. Coleman, M. Maher, D. Brockmeier, D. Kappel, and W. C. Chapman (2014). A novel organ donor facility: A decade of experience with liver donors. *American Journal of Transplantation 14*(3), 615–620.

DuBay, D. A., P. A. MacLennan, R. D. Reed, M. Fouad, M. Martin, C. B. Meeks, G. Taylor, M. L. Kilgore, M. Tankersley, S. H. Gray, J. A. White, D. E. Eckhoff, and J. E. Locke (2015). The impact of proposed changes in liver allocation policy on cold ischemia times and organ transportation costs. *American Journal of Transplantation 15*(2), 541–546.

Dulai, P. S., S. Singh, J. Patel, M. Soni, L. J. Prokop, Z. Younossi, G. Sebastiani, M. Ekstedt, H. Hagstrom, P. Nasr, P. Stal, V. W. S. Wong, S. Kechagias, R. Hultcrantz, and R. Loomba (2017). Increased risk of mortality by fibrosis stage in nonalcoholic fatty liver disease: Systematic review and meta-analysis. *Hepatology 65*(5), 1557–1565.

Dunnett, C. W. (1955). A Multiple Comparison Procedure for Comparing Several Treatments with a Control. *Journal of the American Statistical Association 50*(272), 1096–1121.

Eckermann, S. and A. R. Willan (2008, may). Time and expected value of sample information wait for no patient. *Value in Health 11*(3), 522–526.

Erenay, F. S., O. Alagoz, and A. Said (2014). Optimizing colonoscopy screening for colorectal cancer prevention and surveillance. *Manufacturing & Service Operations Management 16*(3), 381–400.

Ertefaie, A., T. Wu, K. G. Lynch, and I. Nahum-Shani (2016). Identifying a set that contains the best dynamic treatment regimes. *Biostatistics 17*(1), 135–148.

Faraway, J. (2014). *Linear models with R* (2 ed.). CRC Press.

Faraway, J. J. (2006). *Extending the linear model with R: generalized linear, mixed effects and nonparametric regression models*.

Faraway, J. J. (2009). *Linear Models with R*.

Fard, M. M. and J. Pineau (2011). Non-deterministic policies in markovian decision processes. *Journal of Artificial Intelligence Research 40*, 1–24.

Felli, J. C. and G. B. Hazen (1998, jan). Sensitivity analysis and the expected value of perfect information. *Medical Decision Making 18*(1), 95–109.

Felli, J. C. and G. B. Hazen (1999). A Bayesian approach to sensitivity analysis. *Health Economics 8*(3), 263–268.

Fishman, G. S. (1978, jan). Grouping Observations in Digital Simulation. *Management Science 24*(5), 510–521.

Flegal, K. M., M. D. Carroll, B. K. Kit, and C. L. Ogden (2012). Prevalence of obesity and trends in the distribution of body mass index among US adults, 1999-2010. *Journal of the American Medical Association 307*(5), 491–7.

Flegal, K. M., D. Kruszon-Moran, M. D. Carroll, C. D. Fryar, and C. L. Ogden (2016). Trends in obesity among adults in the united states, 2005 to 2014. *Journal of the American Medical Association 315*(21), 2284–2291.

Fonarow, G. C., A. C. Keech, T. R. Pedersen, R. P. Giugliano, P. S. Sever, P. Lindgren, B. Van Hout, G. Villa, Y. Qian, R. Somaratne, and M. S. Sabatine (2017). Cost-effectiveness of evolocumab therapy for reducing cardiovascular events in patients with atherosclerotic cardiovascular disease. *JAMA Cardiology 2*(10), 1069–1078.

Friedman, J. (1991). Multivariate Adaptive Regression Splines. *The Annals of Statistics 19*(1), 1–67.

Friedman, J. H. (2002). Stochastic gradient boosting. *Computational Statistics and Data Analysis 38*(4), 367–378.

Friedman, J. H., T. J. Hastie, and R. J. Tibshirani (2010). glmnet: LASSO and Elastic-net Regularized Generalized Linear Models. *URL http://CRAN. R-project. org/package= glmnet. R package version*, 1.

Fryback, D. G., E. J. Dasbach, R. Klein, B. E. Klein, N. Dorn, K. Peterson, and P. A. Martin (1993). The beaver dam health outcomes study: initial catalog of health-state quality factors. *Medical Decision Making 13*(2), 89–102.

Galetsi, P. and K. Katsaliaki (2020). A review of the literature on big data analytics in healthcare. *Journal of the Operational Research Society 71*(10), 1511–1529.

Gentry, S., E. Chow, A. Massie, and D. Segev (2015, oct). Gerrymandering for Justice: Redistricting U.S. Liver Allocation. *Interfaces 45*(5), 462–480.

Gentry, S. E., E. K. Chow, N. Dzebisashvili, M. A. Schnitzler, K. L. Lentine, C. E. Wickliffe, E. Shteyn, J. Pyke, A. Israni, B. Kasiske, D. L. Segev, and D. A. Axelrod (2016). The Impact of Redistricting Proposals on Health Care Expenditures for Liver Transplant Candidates and Recipients. *American Journal of Transplantation 16*(2), 583–593.

Gentry, S. E., a. B. Massie, S. W. Cheek, K. L. Lentine, E. H. Chow, C. E. Wickliffe, N. Dzebashvili, P. R. Salvalaggio, M. a. Schnitzler, D. a. Axelrod, and D. L. Segev (2013). Addressing geographic disparities in liver transplantation through redistricting. *American Journal of Transplantation 13*, 2052–2058.

Ghamat, S., G. S. Zaric, and H. Pun (2017). Contracts to promote optimal use of optional diagnostic tests in cancer treatment. *Production and Operations Management 27*(12), 2184–2200.

Ginnelly, L., K. Claxton, M. J. Sculpher, and S. Golder (2005). Using value of information analysis to inform publicly funded research priorities. *Applied Health Economics and Health Policy 4*(1), 37–46.

Glover, M., E. Jones, K. Masconi, M. Sweeting, S. Thompson, J. Powell, P. Ulug, and M. Bown (2018). Discrete event simulation for decision modeling in health care: lessons from abdominal aortic aneurysm screening. *Medical Decision Making 38*(4), 439–451.

Goff, D. C., D. M. Lloyd-Jones, G. Bennett, S. Coady, R. B. D'Agostino, R. Gibbons, P. Greenland, D. T. Lackland, D. Levy, C. J. O'Donnell, J. G. Robinson, J. S. Schwartz, S. T. Shero, S. C. Smith, P. Sorlie, N. J. Stone, and P. W. Wilson (2014). 2013 ACC/AHA guideline on the assessment of cardiovascular risk: a report of the American College of Cardiology/American Heart Association Task Force on practice guidelines. *Circulation 129*(25 SUPPL. 1).

Gold, M. R., D. Stevenson, and D. G. Fryback (2002). HALYs and QALYs and DALYs, Oh My: Similarities and Differences in Summary Measures of Population Health. *Annual Review of Public Health 23*(1), 115–134.

Goldberg, D., I. C. Ditah, and K. Saeian (2017). Changes in the Prevalence of Hepatitis C Virus Infection, Non-alcoholic Steatohepatitis, and Alcoholic Liver Disease Among Patients with Cirrhosis or Liver Failure on the Waitlist for Liver Transplantation. *Gastroenterology 152*(5), 1090–1099.e1.

Goldberg, D. S., P. L. Abt, and R. K. Gilroy (2016). Reply to "increasing the Number of Organs Available to Transplant is Separate from Ensuring Equitable Distribution of Available Organs: Both Are Important Goals". *American Journal of Transplantation 16*(2), 730–731.

Goldsman, D. (1992). Simulation output analysis. In *Proceedings of the 24th conference on Winter simulation - WSC '92*, New York, New York, USA, pp. 97–103. ACM Press.

GoodRx (2017). Prescription prices, coupons & pharmacy information.

Govindan, S., L. Shapiro, K. M. Langa, and T. J. Iwashyna (2014). Death Certificates Underestimate Infections as Proximal Causes of Death in the U.S. *PLoS ONE 9*(5), 3–6.

Graham, J. A. and J. V. Guarrera (2014). "Resuscitation" of Marginal Liver Allografts for Transplantation With Machine Perfusion Technology. *Journal of Hepatology 61*(2), 418–431.

Grams, M. E., L. M. Kucirka, C. F. Hanrahan, R. A. Montgomery, A. B. Massie, and D. L. Segev (2012). Candidacy for kidney transplantation of older adults. *Journal of the American Geriatrics Society 60*(1), 1–7.

Grundy, S. M. (2004). Obesity, metabolic syndrome, and cardiovascular disease. *Journal of Clinical Endocrinology and Metabolism 89*(6), 2595–2600.

Grundy, S. M., N. J. Stone, A. L. Bailey, C. Beam, K. K. Birtcher, R. S. Blumenthal, L. T. Braun, S. De Ferranti, J. Faiella-Tommasino, D. E. Forman, R. Goldberg, P. A. Heidenreich, M. A. Hlatky, D. W. Jones, D. Lloyd-Jones, N. Lopez-Pajares, C. E. Ndumele, C. E. Orringer, C. A. Peralta, J. J. Saseen, S. C. Smith, L. Sperling, S. S. Virani, and J. Yeboah (2019). *2018 AHA/ACC/AACVPR/AAPA/ABC/ACPM/ADA/AGS/APhA/ASPC/NLA/PCNA Guideline on the Management of Blood Cholesterol: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines*, Volume 139.

Hall, E. C., B. J. Boyarsky, N. A. Deshpande, J. M. Garonzik-Wang, J. C. Berger, N. N. Dagher, and D. L. Segev (2014). Perioperative complications after live-donor hepatectomy. *JAMA Surgery 149*(3), 288–291.

Harper, P. R. and S. K. Jones (2005). Mathematical models for the early detection and treatment of colorectal cancer. *Health Care Management Science 8*(2), 101–109.

Haskell, W. B., R. Jain, and D. Kalathil (2016, may). Empirical Dynamic Programming. *Mathematics of Operations Research 41*(2), 402–429.

Hassan, M. M., R. Abdel-Wahab, A. Kaseb, A. Shalaby, A. T. Phan, H. B. El-Serag, E. Hawk, J. Morris, K. P. Singh Raghav, J. S. Lee, J. N. Vauthey, G. Bortus, H. A. Torres, C. I. Amos, R. A. Wolff, and D. Li (2015). Obesity Early in Adulthood Increases Risk but Does Not Affect Outcomes of Hepatocellular Carcinoma. *Gastroenterology 149*(1), 119–129.

Hassan, M. M., S. A. Curley, D. Li, A. Kaseb, M. Davila, E. K. Abdalla, M. Javle, D. M. Moghazy, R. D. Lozano, J. L. Abbruzzese, and J. N. Vauthey (2010). Association of diabetes duration and diabetes treatment with the risk of hepatocellular carcinoma. *Cancer 116*(8), 1938–1946.

Hassmiller Lich, K., D. A. Cornejo, M. E. Mayorga, M. Pignone, F. K. Tangka, L. C. Richardson, T.-M. Kuo, A.-M. Meyer, I. J. Hall, J. L. Smith, T. A. Durham, S. A. Chall, T. M. Crutchfield, and S. B. Wheeler (2017). Cost-effectiveness analysis of four simulated colorectal cancer screening interventions, North Carolina. *Preventing Chronic Disease 14*(1), 160158.

Hastie, T. and R. Tibshirani (1986). Generalized Additive Models. *Statistical Science 1*(3), 297–310.

Hastie, T., R. Tibshirani, and J. Friedman (2009). *The Elements of Statistical Learning* (2 ed.). Springer Series in Statistics. New York, NY: Springer New York.

Hauskrecht, M. and H. Fraser (2000). Planning treatment of ischemic heart disease with partially observable Markov decision processes. *Artificial Intelligence in Medicine 18*, 221–244.

Hayward, R. A., H. M. Krumholz, D. M. Zulman, J. W. Timbie, and S. Vijan (2010, jan). Optimizing statin treatment for primary prevention of coronary artery disease. *Annals of Internal Medicine 152*(2), 69.

Heath, A., I. Manolopoulou, and G. Baio (2017). A review of methods for analysis of the expected value of information. *Medical Decision Making 37*(7), 747–758.

Helm, J. E., M. S. Lavieri, M. P. Van Oyen, J. D. Stein, and D. C. Musch (2015). Dynamic forecasting and control algorithms of glaucoma progression for clinician decision support. *Operations Research 63*(5), 979–999.

Hernández-Lerma, O. and J. B. Lasserre (1996). *Discrete-Time Markov Control Processes*, Volume 53. New York, NY: Springer New York.

Hicklin, K., J. S. Ivy, F. C. Payton, M. Viswanathan, and E. Myerse (2018). Exploring the value of waiting during labor. *Service Science 10*(3), 334–353.

Hirose, R., S. E. Gentry, and D. C. Mulligan (2016). Increasing the Number of Organs Available to Transplant is Separate from Ensuring Equitable Distribution of Available Organs: Both Are Important Goals. *American Journal of Transplantation 16*(2), 728–729.

Hoeffding, W. (1963). Probability Inequalities for Sums of Bounded Random Variables. *Journal of the American Statistical Association 58*(301), 13–30.

Hoel, M., T. Iversen, T. Nilssen, and J. Vislie (2006). Genetic testing in competitive insurance markets with repulsion from chance: A welfare analysis. *Journal of Health Economics 25*(5), 847–860.

Hollander, M. and D. A. Wolfe (1999). *Nonparametric statistical methods*. Wiley.

Hsu, J. (1996). *Multiple comparisons: theory and methods*. CRC Press.

Hutton, D. W., D. Tan, S. K. So, and M. L. Brandeau (2007, oct). Cost-effectiveness of screening and vaccinating Asian and Pacific Islander adults for hepatitis B. *Annals of Internal Medicine 147*(7), 460.

Ibrahim, J. G., M.-H. Chen, and D. Sinha (2001). *Bayesian survival analysis*. Springer Series in Statistics. New York, NY: Springer New York.

Ioannidis, J. P. (2018). Diagnosis and treatment of hypertension in the 2017 ACC/AHA guidelines and in the real world. *JAMA - Journal of the American Medical Association 319*(2), 115–116.

James, P. A., S. Oparil, B. L. Carter, W. C. Cushman, C. Dennison-Himmelfarb, J. Handler, D. T. Lackland, M. L. LeFevre, T. D. MacKenzie, O. Ogedegbe, S. C. Smith, L. P. Svetkey, S. J. Taler, R. R. Townsend, J. T. Wright, A. S. Narva, and E. Ortiz (2014). Evidence-based guideline for the management of high blood pressure in adults. *Journal of the American Medical Association 311*(5), 507–20.

Janssen, I., P. T. Katzmarzyk, and R. Ross (2005). Body mass index is inversely related to mortality in older people after adjustment for waist circumference. *Journal of the American Geriatrics Society 53*(12), 2112–2118.

Jarmul, J., M. J. Pletcher, K. Hassmiller Lich, S. B. Wheeler, M. Weinberger, C. L. Avery, D. E. Jonas, S. Earnshaw, and M. Pignone (2018). Cardiovascular genetic risk testing for targeting statin therapy in the primary prevention of atherosclerotic cardiovascular disease. *Circulation: Cardiovascular Quality and Outcomes 11*(4), e004171.

Jenkins, D. H., P. M. Reilly, and C. W. Schwab (1999). Improving the approach to organ donation: A review. *World Journal of Surgery 23*(7), 644–649.

Jha, A. K. (2013, jan). Time to Get Serious About Pay for Performance. *JAMA 309*(4), 347.

Kapelner, A. and J. Bleich (2013). bartMachine: Machine Learning with Bayesian Additive Regression Trees. (2013).

Kathiresan, S. and D. Srivastava (2012). Genetics of human cardiovascular disease. *Cell 148*(6), 1242–1257.

Kazemian, P., J. E. Helm, M. S. Lavieri, J. D. Stein, and M. P. Van Oyen (2018, dec). Dynamic monitoring and control of irreversible chronic diseases with application to glaucoma. *Production and Operations Management 0*(0), poms.12975.

Kazi, D. S., A. E. Moran, P. G. Coxson, J. Penko, D. A. Ollendorf, S. D. Pearson, J. A. Tice, D. Guzman, and K. Bibbins-Domingo (2016). Cost-effectiveness of PCSK9 inhibitor therapy in patients with heterozygous familial hypercholesterolemia or atherosclerotic cardiovascular disease. *Journal of the American Medical Association 316*(7), 743–.

Keller, C. A. (2015). Solid organ transplantation overview and delection criteria. *The American journal of managed care 21*(1), S4–S11.

Khera, A. V., M. Chaffin, K. G. Aragam, M. E. Haas, C. Roselli, S. H. Choi, P. Natarajan, E. S. Lander, S. A. Lubitz, P. T. Ellinor, and S. Kathiresan (2018, sep). Genome-wide polygenic score to identify a monogenic risk-equivalent for coronary disease. *Nature Genetics 50*(9), 1219–1224.

Khera, A. V., C. A. Emdin, I. Drake, P. Natarajan, A. G. Bick, N. R. Cook, D. I. Chasman, U. Baber, R. Mehran, D. J. Rader, V. Fuster, E. Boerwinkle, O. Melander, M. Orho-Melander, P. M. Ridker, and S. Kathiresan (2016). Genetic risk, adherence to a healthy lifestyle, and coronary disease. *New England Journal of Medicine 375*(24), 2349–2358.

Kim, S. P., D. Gupta, A. K. Israni, and B. L. Kasiske (2015). Accept/decline decision module for the liver simulated allocation model. *Health Care Management Science 18*(1), 35–57.

Kim, W. R., J. R. Lake, J. M. Smith, M. A. Skeans, D. P. Schladt, E. B. Edwards, A. M. Harper, J. L. Wainright, J. J. Snyder, A. K. Israni, and B. L. Kasiske (2015, jan). OPTN/SRTR 2013 Annual Data Report: Liver. *American Journal of Transplantation 15*(S2), 1–28.

Kirkizlar, E., D. M. Faissol, P. M. Griffin, and J. L. Swann (2010). Timing of testing and treatment for asymptomatic diseases. *Mathematical Biosciences 226*(1), 28–37.

Kleiber, C. and A. Zeileis (2012). Applied Econometrics with R. Package ARE. *Use R*, 1–6.

Kleinbaum, D. G. and M. Klein (2005). *Survival analysis: a self-learning text* (Second ed.). New York: Springer.

Kleywegt, A. J., A. Shapiro, and T. Homem-De-Mello (2002). The sample average approximation method for stochastic discrete optimization. *SIAM Journal on Optimization 12*(2), 479–502.

Klimberg, R. and B. McCullough (2013). Fundamentals of Predictive Analytics with JMP.

Knowles, J. W. and E. A. Ashley (2018). Cardiovascular disease: The rise of the genetic risk score. *PLoS Medicine 15*(3), 1–7.

Kochanek, K. D., S. L. Murphy, J. Xu, and E. Arias (2019). Deaths: final data for 2017. *National Vital Statistics Reports 68*(9), 1–18.

Kong, N., a. J. Schaefer, B. Hunsaker, and M. S. Roberts (2010). Maximizing the Efficiency of the U.S. Liver Allocation System Through Region Design. *Management Science 56*(12), 2111–2122.

Kuczmarski, R. J. and K. M. Flegal (2000). Criteria for definition of overweight in transition: Background and recommendations for the United States. *American Journal of Clinical Nutrition 72*(5), 1074–1081.

Kuhn, M. (2008). Building Predictive Models in R Using the caret Package. *Journal Of Statistical Software 28*(5), 1–26.

Kurt, M., B. T. Denton, A. J. Schaefer, N. D. Shah, and S. a. Smith (2011). The structure of optimal statin initiation policies for patients with Type 2 diabetes. *IIE Transactions on Healthcare Systems Engineering 1*(July), 49–65.

Laber, E. B., D. J. Lizotte, and B. Ferguson (2014). Set-valued dynamic treatment regimes for competing outcomes. *Biometrics 70*(1), 53–61.

Lazaridis, N. and E. Tsochatzis (2017). Current and future treatment options in non-alcoholic steatohepatitis (NASH). *Expert Review of Gastroenterology and Hepatology 11*(4), 357–369.

Lee, C. P., G. M. Chertow, and S. A. Zenios (2008). Optimal initiation and management of dialysis therapy. *Operations Research 56*(6), 1428–1449.

Lee, E., M. S. Lavieri, and M. Volk (2018). Optimal screening for hepatocellular carcinoma: a restless bandit model. *Manufacturing & Service Operations Management* (January 2019), msom.2017.0697.

Lee, E. K., X. Wei, F. Baker-Witt, M. D. Wright, and A. Quarshie (2018, oct). Outcome-Driven Personalized Treatment Design for Managing Diabetes. *Interfaces 48*(5), 422–435.

Leshno, M., Z. Halpern, and N. Arber (2003). Cost-effectiveness of colorectal cancer screening in the average risk population. *Health Care Management Science 6*(3), 165–174.

Lewis, C. M. and E. Vassos (2020). Polygenic risk scores: From research tools to clinical instruments. *Genome Medicine 12*(1), 1–11.

Li, D., Z. Hawley, and K. Schnier (2013). Increasing organ donation via changes in the default choice or allocation rule. *Journal of health economics 32*(6), 1117–1129.

Li, H. and W. Ning (2012). Multiple comparisons with a control under heteroscedasticity. *Journal of Applied Statistics 39*(10), 2275–2283.

Lin, Y., S. Huang, G. E. Simon, and S. Liu (2018). Data-based decision rules to personalize depression follow-up. *Scientific Reports 8*(1), 4–11.

Lin, Y., S. Huang, G. E. Simon, and S. Liu (2019). Cost-effectiveness analysis of prognostic-based depression monitoring. *IISE Transactions on Healthcare Systems Engineering 9*(1), 41–54.

Liu, S., M. L. Brandeau, and J. D. Goldhaber-Fiebert (2017). Optimizing patient treatment decisions in an era of rapid technological advances: the case of hepatitis C treatment. *Health Care Management Science 20*(1), 16–32.

Lloyd-Jones, D., R. Adams, M. Carnethon, G. De Simone, T. B. Ferguson, K. Flegal, E. Ford, K. Furie, A. Go, K. Greenlund, N. Haase, S. Hailpern, M. Ho, V. Howard, B. Kissela, S. Kittner, D. Lackland, L. Lisabeth, A. Marelli, M. McDermott, J. Meigs, D. Mozaffarian, G. Nichol, C. O'Donnell, V. Roger, W. Rosamond, R. Sacco, P. Sorlie, R. Stafford, J. Steinberger, T. Thom, S. Wasserthiel-Smoller, N. Wong, J. Wylie-Rosett, and Y. Hong (2009, jan). Heart Disease and Stroke Statistics—2009 Update. *Circulation 119*(3), 119:e21–e181.

Long, E. F., E. Nohdurft, and S. Spinler (2018). Spatial resource allocation for emerging epidemics: a comparison of greedy, myopic, and dynamic policies. *Manufacturing and Service Operations Management 20*(2), 181–198.

Long, E. F., N. K. Vaidya, and M. L. Brandeau (2008). Controlling co-epidemics: analysis of HIV and tuberculosis infection dynamics. *Operations Research 56*(6), 1366–1381.

MacRae, C. A. and R. S. Vasan (2016). The future of genetics and genomics. *Circulation 133*(25), 2634–2639.

Maillart, L. M., J. S. Ivy, S. Ransom, and K. Diehl (2008, nov). Assessing dynamic breast cancer screening policies. *Operations Research 56*(6), 1411–1427.

Mak, W. K., D. P. Morton, and R. K. Wood (1999). Monte Carlo bounding techniques for determining solution quality in stochastic programs. *Operations Research Letters 24*(1), 47–56.

Malik, V. S., W. C. Willett, and F. B. Hu (2013). Global obesity: Trends, risk factors and policy implications. *Nature Reviews Endocrinology 9*(1), 13–27.

Malinoski, D. J., M. C. Daly, M. S. Patel, C. Oley-Graybill, C. E. Foster, and A. Salim (2011). Achieving donor management goals before deceased donor procurement is associated with more organs transplanted per donor. *Journal of Trauma - Injury, Infection and Critical Care 71*(4), 990–995.

Manninen, D. L. and R. W. Evans (1985). Public attitudes and behavior regarding organ donation. *Journal of the American Medical Association 253*(21), 3111–3115.

Mannor, S., D. Simester, P. Sun, and J. N. Tsitsiklis (2007). Bias and Variance Approximation in Value Function Estimates. *Management Science 53*(2), 308–322.

Mannor, S. and J. N. Tsitsiklis (2013). Algorithmic aspects of mean-variance optimization in Markov decision processes. *European Journal of Operational Research 231*(3), 645–653.

Marchesini, G., S. Petta, and R. Dalle Grave (2016). Diet, weight loss, and liver health in nonalcoholic fatty liver disease: Pathophysiology, evidence, and practice. *Hepatology 63*(6), 2032–2043.

Marks, W. H., D. Wagner, T. C. Pearson, J. P. Orlowski, P. W. Nelson, J. J. McGowan, M. K. Guidinger, and J. Burdick (2006). Organ donation and utilization, 1995-2004: Entering the collaborative era. *American Journal of Transplantation 6*(5 II), 1101–1110.

Marrero, W. J., M. S. Lavieri, S. D. Guikema, D. W. Hutton, and N. D. Parikh (2018, aug). Development of a Predictive Model for Deceased Donor Organ Yield. *Transplantation 102*(8), e364.

Marrero, W. J., M. S. Lavieri, and J. B. Sussman (2019). A simulation model to evaluate the implications of genetic testing in cholesterol treatment plans. In *Proceedings of the 2019 Winter Simulation Conference*.

Marrero, W. J., A. S. Naik, J. J. Friedewald, Y. Xu, D. W. Hutton, M. S. Lavieri, and N. D. Parikh (2016, may). Predictors of Deceased Donor Kidney Discard in the United States. *Transplantation 00*(00), 1–8.

Martikainen, J. A., A. Kivioja, T. Hallinen, and P. Vihinen (2005). Economic evaluation of temozolomide in the treatment of recurrent glioblastoma multiforme. *PharmacoEconomics 23*(8), 803–815.

Mason, J. E., B. T. Denton, N. D. Shah, and S. A. Smith (2014). Optimizing the simultaneous management of blood pressure and cholesterol for Type 2 diabetes patients. *European Journal of Operational Research 233*(3), 727–738.

Massie, A. B., B. Caffo, S. E. Gentry, E. C. Hall, D. A. Axelrod, K. L. Lentine, M. A. Schnitzler, A. Gheorghian, P. R. Salvalaggio, and D. L. Segev (2011). MELD exceptions and rates of waiting list outcomes. *American Journal of Transplantation 11*(11), 2362–2371.

Massie, A. B., E. K. Chow, C. E. Wickliffe, X. Luo, S. E. Gentry, D. C. Mulligan, and D. L. Segev (2015). Early changes in liver distribution following implementation of Share 35. *American Journal of Transplantation 15*(3), 659–667.

McLachlan, G. J. (1992, mar). *Discriminant Analysis and Statistical Pattern Recognition*. Wiley Series in Probability and Statistics. Hoboken, NJ, USA: John Wiley & Sons, Inc.

McNeil, J. J., M. R. Nelson, R. L. Woods, J. E. Lockery, R. Wolfe, C. M. Reid, B. Kirpach, R. C. Shah, D. G. Ives, E. Storey, J. Ryan, A. M. Tonkin, A. B. Newman, J. D. Williamson, K. L. Margolis, M. E. Ernst, W. P. Abhayaratna, N. Stocks, S. M. Fitzgerald, S. G. Orchard, R. E. Trevaks, L. J. Beilin, G. A. Donnan, P. Gibbs, C. I. Johnston, B. Radziszewska, R. Grimm, and A. M. Murray (2018). Effect of aspirin on all-cause mortality in the healthy elderly. *New England Journal of Medicine 379*(16), 1519–1528.

Medical Expenditure Panel Survey (2015). MEPS HC-163: 2013 full year consolidated data file.

Medved, D., P. Nugues, and J. Nilsson (2016, aug). Selection of an optimal feature set to predict heart transplantation outcomes. In *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 3290–3293. IEEE.

Mega, J. L., N. O. Stitziel, J. G. Smith, D. I. Chasman, M. J. Caulfield, J. J. Devlin, F. Nordio, C. L. Hyde, C. P. Cannon, F. M. Sacks, N. R. Poulter, P. S. Sever, P. M. Ridker, E. Braunwald, O. Melander, S. Kathiresan, and M. S. Sabatine (2015). Genetic risk, coronary

heart disease events, and the clinical benefit of statin therapy: an analysis of primary and secondary prevention trials. *The Lancet 385*(9984), 2264–2271.

Memarzadeh, M. and M. Pozzi (2016). Value of information in sequential decision making: component inspection, permanent monitoring and system-level scheduling. *Reliability Engineering and System Safety 154*, 137–151.

Messersmith, E. E., C. Arrington, C. Alexander, J. P. Orlowski, and R. Wolfe (2011). Development of donor yield models. *American Journal of Transplantation 11*(10), 2075–2084.

Miller, A. C. (1975, sep). The value of sequential information. *Management Science 22*(1), 1–11.

Miniño, A. M., M. P. Heron, S. L. Murphy, K. D. Kochanek, and Centers for Disease Control and Prevention National Center for Health Statistics National Vital Statistics System (2007, aug). Deaths: final data for 2004. *National vital statistics reports : from the Centers for Disease Control and Prevention, National Center for Health Statistics, National Vital Statistics System 55*(19), 1–119.

Moers, C., N. S. S. Kornmann, H. G. D. Leuvenink, and R. J. Ploeg (2009, aug). The influence of deceased donor age and old-for-old allocation on kidney transplant outcome. *Transplantation 88*(4), 542–52.

Musso, G., M. Cassader, C. Olivetti, F. Rosina, G. Carbone, and R. Gambino (2013). Association of obstructive sleep apnoea with the presence and severity of non-alcoholic fatty liver disease. A systematic review and meta-analysis. *Obesity Reviews 14*(5), 417–431.

Natarajan, P., R. Young, N. O. Stitziel, S. Padmanabhan, U. Baber, R. Mehran, S. Sartori, V. Fuster, D. F. Reilly, A. Butterworth, D. J. Rader, I. Ford, N. Sattar, and S. Kathiresan (2017). Polygenic risk score identifies subgroup with higher burden of atherosclerosis and greater relative benefit from statin therapy in the primary prevention setting. *Circulation 135*(22), 2091–2101.

NCHS (2017). Health, United States, 2016: with chartbook on long-term trends in health. *Center for Disease Control*, 314–317.

Negoescu, D. M., K. Bimpikis, M. L. Brandeau, and D. A. Iancu (2017). Dynamic learning of patient response types: an application to treating chronic diseases. *Management Science* (January 2019), mnsc.2017.2793.

Negoescu, D. M., K. Bimpikis, M. L. Brandeau, and D. A. Iancu (2018). Dynamic learning of patient response types: An application to treating chronic diseases. *Management Science 64*(8), 3469–3488.

Neumann, P., G. Sanders, L. Russell, and J. Siegel (2016). *Cost-effectiveness in health and medicine*. Oxford University Press.

Neumann, P. J., J. T. Cohen, and M. C. Weinstein (2014). Updating Cost-Effectiveness — The Curious Resilience of the $50,000-per-QALY Threshold. *New England Journal of Medicine 371*(9), 796–797.

Northup, P. G., N. M. Intagliata, N. L. Shah, S. J. Pelletier, C. L. Berg, and C. K. Argo (2015). Excess mortality on the liver transplant waiting list: Unintended policy consequences and model for End-Stage Liver Disease (MELD) inflation. *Hepatology 61*(1), 285–291.

Noureddin, M., A. Zhang, and R. Loomba (2016). Promising therapies for treatment of nonalcoholic steatohepatitis. *Expert Opinion on Emerging Drugs 21*(3), 343–357.

Nowok, B., G. M. Raab, and C. Dibben (2016). synthpop: Bespoke Creation of Synthetic Data in R. *Journal of Statistical Software 74*(11).

Onen, Z., S. Sayin, and b. Gurvit (2018). Optimal population screening policies for Alzheimer's disease. *IISE Transactions on Healthcare Systems Engineering 5579*, 1–36.

Organ Procurement and Transplantation Network (2020). OPTN Bylaws.

Orman, E. S., A. S. Barritt, S. B. Wheeler, and P. H. Hayashi (2013, jan). Declining liver utilization for transplantation in the United States and the impact of donation after cardiac death. *Liver Transplantation 19*(1), 59–68.

O'Sullivan, A. K., J. Rubin, J. Nyambose, A. Kuznik, D. J. Cohen, and D. Thompson (2011, aug). Cost estimation of cardiovascular disease events in the US. *PharmacoEconomics 29*(8), 693–704.

Ozcan, Y. A. (2005). *Quantitative methods in health care management: techniques and applications*. Jossey-Bass.

Oztekin, A., D. Delen, and Z. J. Kong (2009). Predicting the graft survival for heart-lung transplantation patients: An integrated data mining methodology. *International Journal of Medical Informatics 78*(12).

Pandya, A., S. Sy, S. Cho, M. C. Weinstein, and T. A. Gaziano (2015). Cost-Effectiveness of 10-Year Risk Thresholds for Initiation of Statin Therapy for Primary Prevention of Cardiovascular Disease. *Journal of the American Medical Association 314*(2), 142–150.

Parikh, N. D., D. Hutton, W. Marrero, K. Sanghani, Y. Xu, and M. Lavieri (2015, jun). Projections in donor organs available for liver transplantation in the United States: 2014-2025. *Liver transplantation : official publication of the American Association for the Study of Liver Diseases and the International Liver Transplantation Society 21*(6), 855–63.

Parikh, N. D., W. J. Marrero, C. J. Sonnenday, A. S. Lok, D. W. Hutton, and M. S. Lavieri (2017). Population-Based Analysis and Projections of Liver Supply Under Redistricting. *Transplantation 101*(9), 2048–2055.

Parikh, N. D., W. J. Marrero, J. Wang, J. Steuer, E. B. Tapper, M. Konerman, A. G. Singal, D. W. Hutton, E. Byon, and M. S. Lavieri (2017, aug). Projected increase in obesity and non-alcoholic steatohepatitis-related liver transplantation waitlist additions in the United States. *Hepatology* (5), 1–36.

Patel, N. S., I. Doycheva, M. R. Peterson, J. Hooker, T. Kisselva, B. Schnabl, E. Seki, C. B. Sirlin, and R. Loomba (2015). Effect of weight loss on magnetic resonance imaging estimation of liver fat and volume in patients with nonalcoholic steatohepatitis. *Clinical Gastroenterology and Hepatology 13*(3), 561–568.e1.

Phillips, K. A., P. A. Deverka, G. W. Hooker, and M. P. Douglas (2018). Genetic test availability and spending: Where are we now? Where are we going? *Health Affairs 37*(5), 710–716.

Pignone, M. (2007, feb). Aspirin for the primary prevention of cardiovascular disease in women. *Archives of Internal Medicine 167*(3), 290.

Pignone, M., S. Earnshaw, J. A. Tice, and M. J. Pletcher (2006). Aspirin, statins, or both drugs for the primary prevention of coronary heart disease events in men: a cost-utility analysis. *Annals of Internal Medicine 144*, 326–336.

Pletcher, M. J., M. Pignone, S. Earnshaw, C. McDade, K. A. Phillips, R. Auer, L. Zablotska, and P. Greenland (2014, mar). Using the coronary artery calcium score to guide statin therapy. *Circulation: Cardiovascular Quality and Outcomes 7*(2), 276–284.

Powell, W. B. (2011, aug). *Approximate Dynamic Programming* (2 ed.). Wiley Series in Probability and Statistics. Hoboken, NJ, USA: John Wiley & Sons, Inc.

Pozzi, M. and A. Der Kiureghian (2011, mar). Assessing the value of information for long-term structural health monitoring. In T. Kundu (Ed.), *Proc. SPIE 7984, Health Monitoring of Structural and Biological Systems 2011*, Number April 2011, pp. 79842W.

Promrat, K., D. E. Kleiner, H. M. Niemeier, E. Jackvony, M. Kearns, J. R. Wands, J. L. Fava, and R. R. Wing (2010). Randomized controlled trial testing the effects of weight loss on nonalcoholic steatohepatitis. *Hepatology 51*(1), 121–129.

Puterman, M. L. (2014). *Markov decision processes: discrete stochastic dynamic programming*. John Wiley & Sons.

R Core Team (2016). R: A Language and Environment for Statistical Computing.

Raic, M. (2019). A multivariate Berry – Esseen theorem with explicit constants. *Bernoulli 25*, 2824–2853.

Raiffa, H. and R. Schlaifer (1961, jan). *Applied statistical decision theory*. Harvard University and MIT Press.

Raji, C. G. and S. S. V. Chandra (2016). Predicting the survival of graft following liver transplantation using a nonlinear model. *Journal of Public Health (Germany) 24*(5), 443–452.

Rana, A., B. Kaplan, I. B. Riaz, M. Porubsky, S. Habib, H. Rilo, A. C. Gruessner, and R. W. Gruessner (2015). Geographic Inequities in Liver Allograft Supply and Demand. *Transplantation 99*(3), 515–520.

Rao, V., R. S. Behara, and A. Agarwal (2014). Predictive modeling for organ transplantation outcomes. *Proceedings - IEEE 14th International Conference on Bioinformatics and Bioengineering, BIBE 2014* (Figure 1), 405–408.

Rayhill, S. C., Y. M. Wu, D. A. Katz, M. D. Voigt, D. R. LaBrecque, P. A. Kirby, F. A. Mitros, R. S. Kalil, R. A. Miller, A. H. Stolpen, and W. N. Schmidt (2007). Older Donor Livers Show Early Severe Histological Activity, Fibrosis, and Graft Failure After Liver Transplantation for Hepatitis C. *Transplantation 84*(3), 331–339.

Refaeilzadeh, P., L. Tang, and H. Liu (2009). Cross-Validation. *Encyclopedia of Database Systems*, 532–538.

Ridgeway, G. (2007). Generalized Boosted Models: A guide to the gbm package. *Http://Cran.R-Project.Org/Web/Packages/Gbm/Gmb/Pdf 1*(4), 1–12.

Rithalia, A., C. McDaid, S. Suekarran, L. Myers, and A. Sowden (2009). Impact of presumed consent for organ donation on donation rates: a systematic review. *BMJ (Clinical research ed.) 338*(jan14_2), a3162.

Roberts, E. T., A. Horne, S. S. Martin, M. J. Blaha, R. Blankstein, M. J. Budoff, C. Sibley, J. F. Polak, K. D. Frick, R. S. Blumenthal, and K. Nasir (2015). Cost-effectiveness of coronary artery calcium testing for coronary heart and cardiovascular disease risk prediction to guide statin allocation: The Multi-Ethnic Study of Atherosclerosis (MESA). *PLoS ONE 10*(3), 1–20.

Robins, J., L. Orellana, and A. Rotnitzky (2008, oct). Estimation and extrapolation of optimal treatment and testing strategies. *Statistics in Medicine 27*(23), 4678–4721.

Roderick, P., R. Davies, C. Jones, T. Feest, S. Smith, and K. Farrington (2004). Simulation model of renal replacement therapy: Predicting future demand in England. *Nephrology Dialysis Transplantation 19*(3), 692–701.

Rodrigue, J. R., D. L. Cornell, and R. J. Howard (2006). Organ donation decision: Comparison of donor and nondonor families. *American Journal of Transplantation 6*(1), 190–198.

Rudd, R. A., N. Aleshire, J. E. Zibbell, and R. Matthew Gladden (2016). Increases in Drug and Opioid Overdose Deaths - United States, 2000-2014. *American Journal of Transplantation 16*(4), 1323–1327.

Sabouri, A., W. T. Huh, and S. M. Shechter (2017). Screening strategies for patients on the kidney transplant waiting list. *Operations Research 65*(5), 1131–1146.

Saini, S. D., P. Schoenfeld, K. Kaulback, and M. C. Dubinsky (2009, jun). Effect of medication dosing frequency on adherence in chronic diseases. *The American journal of managed care 15*(6), e22–33.

Salam, A., E. Atkins, J. Sundström, Y. Hirakawa, D. Ettehad, C. Emdin, B. Neal, M. Woodward, J. Chalmers, E. Berge, S. Yusuf, K. Rahimi, and A. Rodgers (2019). Effects of blood pressure lowering on cardiovascular events, in the context of regression to the mean: A systematic review of randomized trials. *Journal of Hypertension 37*(1), 16–23.

Salim, A., E. J. Ley, C. Berry, D. Schulman, S. Navarro, L. Zheng, and L. S. Chan (2014). Increasing organ donation in Hispanic Americans: The role of media and other community outreach efforts. *JAMA Surgery 149*(1), 71–76.

Saville, C. E., H. K. Smith, and K. Bijak (2018). Operational research techniques applied throughout cancer care services: a review. *Health Systems 6965*, 1–22.

Schell, G. J., W. J. Marrero, M. S. Lavieri, J. B. Sussman, and R. A. Hayward (2016, oct). Data-driven Markov decision process approximations for personalized hypertension treatment planning. *MDM Policy & Practice 1*(1).

Schlansky, B., W. E. Naugler, S. L. Orloff, and C. Kristian Enestvedt (2016). Higher Mortality and Survival Benefit in Obese Patients Awaiting Liver Transplantation. *Transplantation 100*(12), 2648–2655.

Scientific Registry of Transplant Recipients (2019). Risk Adjustment Models: OPOs.

Selck, F. W., P. Deb, and E. B. Grossman (2008). Deceased organ donor characteristics and clinical interventions associated with organ yield. *American Journal of Transplantation 8*(5), 965–974.

Serfling, R. J. (1980, nov). *Approximation Theorems of Mathematical Statistics*, Volume 145 of *Wiley Series in Probability and Statistics*. Hoboken, NJ, USA: John Wiley & Sons, Inc.

Sharma, R., M. Gupta, and G. Kapoor (2010). Some better bounds on the variance with applications. *4*(3), 355–363.

Singh, K. (1981, nov). On the Asymptotic Accuracy of Efron's Bootstrap. *The Annals of Statistics 9*(6), 1187–1195.

Skandari, M. R., S. M. Shechter, and N. Zalunardo (2015). Optimal vascular access choice for patients on hemodialysis. *Manufacturing & Service Operations Management 17*(4), 608–619.

Smolina, K., F. L. Wright, M. Rayner, and M. J. Goldacre (2012). Long-term survival and recurrence after acute myocardial infarction in England, 2004 to 2010. *Circulation: Cardiovascular Quality and Outcomes 5*(4), 532–540.

Solberg, L. I. and W. L. Miller (2018, sep). The new hypertension guideline: logical but unwise. *Family Practice 35*(5), 528–530.

Stanford R.E. (2004). A frontier analysis approach for benchmarking hospital performance in the treatment of acute myocardial infarction. *Health Care Management Science 7*, 145–154.

Starley, B. Q., C. J. Calcagno, and S. A. Harrison (2010). Nonalcoholic fatty liver disease and hepatocellular carcinoma: A weighty connection. *Hepatology 51*(5), 1820–1832.

Starzl, T. E. (1984). Implied Consent for Cadaveric Organ Donation. *JAMA: The Journal of the American Medical Association 251*(12), 1592.

Steiger, N., E. Lada, J. Wilson, C. Alexopoulos, D. Goldsman, and F. Zouaoui (2002). ASAP2: an improved batch means procedure for simulation output analysis. In *Proceedings of the Winter Simulation Conference*, Volume 1, pp. 336–344. IEEE.

Steiger, N. M., E. K. Lada, J. R. Wilson, J. A. Joines, C. Alexopoulos, and D. Goldsman (2005). ASAP3: A batch means procedure for steady-state simulation analysis. *ACM Transactions on Modeling and Computer Simulation 15*(1), 39–73.

Steiger, N. M. and J. R. Wilson (2002). An improved batch means procedure for simulation output analysis. *Management Science 48*(12), 1569–1586.

Steimle, L. N., D. L. Kaufman, and B. T. Denton (2019). Multi-model Markov Decision Processes. *Optimization Online*, 1–64.

Stekhoven, D. J. and P. Buhlmann (2012). MissForest–non-parametric missing value imputation for mixed-type data. *Bioinformatics 28*(1), 112–118.

Steuten, L., G. Van De Wetering, K. Groothuis-Oudshoorn, and V. Retèl (2013). A systematic and critical review of the evolving methods and applications of value of information in academia and practice. *PharmacoEconomics 31*(1), 25–48.

Stobierski, T. (2019). The Advantages of Data-Driven Decision-Making.

Stone, N. J., J. G. Robinson, A. H. Lichtenstein, C. N. Bairey Merz, C. B. Blum, R. H. Eckel, A. C. Goldberg, D. Gordon, D. Levy, D. M. Lloyd-Jones, P. McBride, J. S. Schwartz, S. T. Shero, S. C. Smith, K. Watson, and P. W. F. Wilson (2014, jun). 2013 ACC/AHA guideline on the treatment of blood cholesterol to reduce atherosclerotic cardiovascular risk in adults. *Circulation 129*(25 suppl 2), S1–S45.

Su, F., L. Yu, K. Berry, I. W. Liou, C. S. Landis, S. C. Rayhill, J. D. Reyes, and G. N. Ioannou (2016). Aging of Liver Transplant Registrants and Recipients: Trends and Impact on Waitlist Outcomes, Post-Transplantation Outcomes, and Transplant-Related Survival Benefit. *Gastroenterology 150*(2), 441–453e6.

Suen, S.-c., M. L. Brandeau, and J. D. Goldhaber-Fiebert (2018). Optimal timing of drug sensitivity testing for patients on first-line tuberculosis treatment. *Health Care Management Science 21*(4), 632–646.

Sun, Y., S. Goodison, J. Li, L. Liu, and W. Farmerie (2007, jan). Improved breast cancer prognosis through the combination of clinical and genetic markers. *Bioinformatics 23*(1), 30–37.

Sundström, J., H. Arima, R. Jackson, F. Turnbull, K. Rahimi, J. Chalmers, M. Woodward, and B. Neal (2015). Effects of blood pressure reduction in mild hypertension: A systematic review and meta-analysis. *Annals of Internal Medicine 162*(3), 184–191.

Sundström, J., H. Arima, M. Woodward, R. Jackson, K. Karmali, D. Lloyd-Jones, C. Baigent, J. Emberson, K. Rahimi, S. Macmahon, A. Patel, V. Perkovic, F. Turnbull, B. Neal, L. Agodoa, R. Estacio, R. Schrier, J. Lubsen, J. Chalmers, J. Cutler, B. Davis, L. Wing, N. R. Poulter, P. Sever, G. Remuzzi, P. Ruggenenti, S. Nissen, L. H. Lindholm, T. Fukui, T. Ogihara, T. Saruta, H. Black, P. Sleight, M. Lievre, H. Suzuki, K. Fox, L. Lisheng, T. Ohkubo, Y. Imai, S. Yusuf, C. J. Bulpitt, E. Lewis, M. Brown, C. Palmer, J. Wang, C. Pepine, M. Ishii, Y. Yui, K. Kuramoto, M. Pfeffer, F. W. Asselbergs, W. H. van Gilst, B. Byington, B. Pitt, B. Brenner, W. J. Remme, D. de Zeeuw, M. Rahman, G. Viberti, K. Teo, A. Zanchetti, E. Malacco, G. Mancia, J. Staessen, R. Fagard, R. Holman, L. Hansson, J. Kostis, Y. Kanno, S. Lueders, M. Matsuzaki, P. Poole-Wilson, J. Schrader, K. Rahimi, C. Anderson, J. Chalmers, N. Chapman, R. Collins, B. Neal, A. Rodgers, P. Whelton, M. Woodward, and S. Yusuf (2014). Blood pressure-lowering treatment based on cardiovascular risk: A meta-analysis of individual patient data. *The Lancet 384*(9943), 591–598.

Sussman, J., S. Vijan, and R. Hayward (2013). Using Benefit-Based Tailored Treatment to Improve the Use of Antihypertensive Medications. *Circulation 128*(21), 2309–2317.

Sussman, J. B., S. Vijan, H. Choi, and R. a. Hayward (2011). Individual and population benefits of daily aspirin therapy a proposal for personalizing national guidelines. *Circulation: Cardiovascular Quality and Outcomes 4*(3), 268–275.

Sussman, J. B., W. L. Wiitala, M. Zawistowski, T. P. Hofer, D. Bentley, and R. A. Hayward (2017, sep). The Veterans Affairs cardiac risk score: recalibrating the atherosclerotic cardiovascular disease score for applied use. *Medical Care 55*(9), 864–870.

Sutton, R. S. and A. G. Barto (2018). *Reinforcement Learning: An Introduction* (2 ed.). MIT press.

Tetri, L. H., M. Basaranoglu, E. M. Brunt, L. M. Yerian, and B. A. Neuschwander-Tetri (2008). Severe NAFLD with hepatic necroinflammatory changes in mice fed trans fats and a high-fructose corn syrup equivalent. *Am J Physiol Gastrointest Liver Physiol 295*(5), G987–95.

The CARDIoGRAMplusC4D Consortium (2013). Large-scale association analysis identifies new risk loci for coronary artery disease. *Nature genetics 45*(1), 25–33.

Thompson, K. and F. Yokota (2004). Value of information analysis in environmental health risk management decisions: Past, present, and future. *Risk Analysis 24*(3), 635–650.

Thuluvath, P. J. (2007, dec). Morbid obesity with one or more other serious comorbidities should be a contraindication for liver transplantation. *Liver Transplantation 13*(12), 1627–1629.

Tom, P. and K. S. Kumar (2016). Cadaver Kidney Demand Forecasting and Classification Modelling of Kidney Allocation–A Case Study. *Procedia Technology 25*(Raerest), 1162–1169.

Toothaker, L. (2012). *Multiple Comparison Procedures*. John Wiley & Sons.

Tu, W. and X. H. Zhou (2000). Pairwise comparisons of the means of skewed data. *Journal of Statistical Planning and Inference 88*(1), 59–74.

United States Department of Health and Human Services (2006). Coverage and Reimbursement of Genetic Tests and Services: Report of the Secretary's Advisory Committee on Genetics, Health, and Society. (February), 1–57.

US Census Bureau (2011). 2010 Census Shows 65 and Older Population Growing Faster Than Total U.S. Population - 2010 Census - Newsroom - U.S. Census Bureau.

US Census Bureau (2012). National Population Projections: Summary Tables.

US Census Bureau (2014). National Population Projections Tables.

US Census Bureau (2018). State Intercensal Datasets: 2000-2010.

US Census Bureau (2020). National Population by Characteristics: 2010-2019.

U.S. Organ Procurement and Transplantation Network (2019). Organ Procurement and Transplantation Network database from 2000 to 2018.

Van Buuren, S. and K. Groothuis-Oudshoorn (2011). Multivariate Imputation by Chained Equations. *Journal Of Statistical Software 45*(3), 1–67.

Van Hasselt, H. (2010). Double Q-learning. *Advances in Neural Information Processing Systems 23: 24th Annual Conference on Neural Information Processing Systems 2010, NIPS 2010*, 1–9.

VanWagner, L. B., M. Serper, R. Kang, J. Levitsky, S. Hohmann, M. Abecassis, A. Skaro, and D. M. Lloyd-Jones (2016). Factors Associated With Major Adverse Cardiovascular Events After Liver Transplantation Among a National Sample. *American Journal of Transplantation 16*(9), 2684–2694.

Vasan, R. S. (2006). Biomarkers of cardiovascular disease: molecular basis and practical considerations. *Circulation 113*(19), 2335–2362.

Vinogradova, Y., C. Coupland, P. Brindle, and J. Hippisley-Cox (2016). Discontinuation and restarting in patients on statin treatment: prospective open cohort study using a primary care database. *Bmj*, i3305.

Virani, S. S., A. Alonso, E. J. Benjamin, M. S. Bittencourt, C. W. Callaway, A. P. Carson, A. M. Chamberlain, A. R. Chang, S. Cheng, F. N. Delling, L. Djousse, M. S. Elkind, J. F. Ferguson, M. Fornage, S. S. Khan, B. M. Kissela, K. L. Knutson, T. W. Kwan, D. T. Lackland, T. T. Lewis, J. H. Lichtman, C. T. Longenecker, M. S. Loop, P. L. Lutsey, S. S.

Martin, K. Matsushita, A. E. Moran, M. E. Mussolino, A. M. Perak, W. D. Rosamond, G. A. Roth, U. K. Sampson, G. M. Satou, E. B. Schroeder, S. H. Shah, C. M. Shay, N. L. Spartano, A. Stokes, D. L. Tirschwell, L. B. VanWagner, C. W. Tsao, S. S. Wong, and D. G. Heard (2020). *Heart disease and stroke statistics—2020 update: A report from the American Heart Association*.

Weldon Cooper Center for Public Service (2013). National Population Projections.

Westfall, P. H. (2011, nov). On Using the Bootstrap for Multiple Comparisons. *Journal of Biopharmaceutical Statistics 21*(6), 1187–1205.

Westfall, P. H. and S. S. Young (1993). *Resampling-based multiple testing: examples and methods for P-value adjustment*. Wiley.

Whelton, P. K., R. M. Carey, W. S. Aronow, D. E. Casey, K. J. Collins, C. Dennison Himmelfarb, S. M. DePalma, S. Gidding, K. A. Jamerson, D. W. Jones, E. J. MacLaughlin, P. Muntner, B. Ovbiagele, S. C. Smith, C. C. Spencer, R. S. Stafford, S. J. Taler, R. J. Thomas, K. A. Williams, J. D. Williamson, and J. T. Wright (2018, may). 2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA Guideline for the Prevention, Detection, Evaluation, and Management of High Blood Pressure in Adults. *Journal of the American College of Cardiology 71*(19), e127–e248.

White, I. R., P. Royston, and A. M. Wood (2011). Multiple imputation using chained equations: Issues and guidance for practice. *Statistics in Medicine 30*(4), 377–399.

Williams, B., G. Mancia, W. Spiering, E. Agabiti Rosei, M. Azizi, M. Burnier, D. L. Clement, A. Coca, G. de Simone, A. Dominiczak, T. Kahan, F. Mahfoud, J. Redon, L. Ruilope, A. Zanchetti, M. Kerins, S. E. Kjeldsen, R. Kreutz, S. Laurent, G. Y. H. Lip, R. McManus, K. Narkiewicz, F. Ruschitzka, R. E. Schmieder, E. Shlyakhto, C. Tsioufis, V. Aboyans, I. Desormais, G. De Backer, A. M. Heagerty, S. Agewall, M. Bochud, C. Borghi, P. Boutouyrie, J. Brguljan, H. Bueno, E. G. Caiani, B. Carlberg, N. Chapman, R. Cífková, J. G. F. Cleland, J.-P. Collet, I. M. Coman, P. W. de Leeuw, V. Delgado, P. Dendale, H.-C. Diener, M. Dorobantu, R. Fagard, C. Farsang, M. Ferrini, I. M. Graham, G. Grassi, H. Haller, F. D. R. Hobbs, B. Jelakovic, C. Jennings, H. A. Katus, A. A. Kroon, C. Leclercq, D. Lovic, E. Lurbe, A. J. Manolis, T. A. McDonagh, F. Messerli, M. L. Muiesan, U. Nixdorff, M. H. Olsen, G. Parati, J. Perk, M. F. Piepoli, J. Polonia, P. Ponikowski, D. J. Richter, S. F. Rimoldi, M. Roffi, N. Sattar, P. M. Seferovic, I. A. Simpson, M. Sousa-Uva, A. V. Stanton, P. van de Borne, P. Vardas, M. Volpe, S. Wassmann, S. Windecker, J. L. Zamorano, S. Windecker, V. Aboyans, S. Agewall, E. Barbato, H. Bueno, A. Coca, J.-P. Collet, I. M. Coman, V. Dean, V. Delgado, D. Fitzsimons, O. Gaemperli, G. Hindricks, B. Iung, P. Jüni, H. A. Katus, J. Knuuti, P. Lancellotti, C. Leclercq, T. A. McDonagh, M. F. Piepoli, P. Ponikowski, D. J. Richter, M. Roffi, E. Shlyakhto, I. A. Simpson, M. Sousa-Uva, J. L. Zamorano, C. Tsioufis, E. Lurbe, R. Kreutz, M. Bochud, E. A. Rosei, B. Jelakovic, M. Azizi, A. Januszewics, T. Kahan, J. Polonia, P. van de Borne, B. Williams, C. Borghi, G. Mancia, G. Parati, D. L. Clement, A. Coca, A. Manolis, D. Lovic, S. Benkhedda, P. Zelveian, P. Siostrzonek, R. Najafov, O. Pavlova, M. De Pauw, L. Dizdarevic-Hudic, D. Raev, N. Karpettas, A. Linhart, M. H. Olsen, A. F. Shaker, M. Viigimaa, K. Metsärinne,

M. Vavlukis, J.-M. Halimi, Z. Pagava, H. Schunkert, C. Thomopoulos, D. Páll, K. Andersen, M. Shechter, G. Mercuro, G. Bajraktari, T. Romanova, K. Trušinskis, G. A. Saade, G. Sakalyte, S. Noppe, D. C. DeMarco, A. Caraus, J. Wittekoek, T. A. Aksnes, P. Jankowski, J. Polonia, D. Vinereanu, E. I. Baranova, M. Foscoli, A. D. Dikic, S. Filipova, Z. Fras, V. Bertomeu-Martínez, B. Carlberg, T. Burkard, W. Sdiri, S. Aydogdu, Y. Sirenko, A. Brady, T. Weber, I. Lazareva, T. D. Backer, S. Sokolovic, B. Jelakovic, J. Widimsky, M. Viigimaa, I. Pörsti, T. Denolle, B. K. Krämer, G. S. Stergiou, G. Parati, K. Trušinskis, M. Miglinas, E. Gerdts, A. Tykarski, M. de Carvalho Rodrigues, M. Dorobantu, I. Chazova, D. Lovic, S. Filipova, J. Brguljan, J. Segura, A. Gottsäter, A. Pechère-Bertschi, S. Erdine, Y. Sirenko, and A. Brady (2018, sep). 2018 ESC/ESH guidelines for the management of arterial hypertension. *European Heart Journal 39*(33), 3021–3104.

Wilt, T. J., D. Kansagara, and A. Qaseem (2018, mar). Hypertension Limbo: Balancing Benefits, Harms, and Patient Preferences Before We Lower the Bar on Blood Pressure. *Annals of Internal Medicine 168*(5), 369.

Wong, R. J., R. Cheung, and A. Ahmed (2014a). Nonalcoholic steatohepatitis is the most rapidly growing indication for liver transplantation in patients with hepatocellular carcinoma in the U.S. *Hepatology 59*(6), 2188–2195.

Wong, R. J., R. Cheung, and A. Ahmed (2014b). Nonalcoholic steatohepatitis is the most rapidly growing indication for liver transplantation in patients with hepatocellular carcinoma in the U.S. *Hepatology 59*(6), 2188–2195.

Wong, R. J., P. Devaki, L. Nguyen, R. Cheung, and M. H. Nguyen (2014, may). Ethnic disparities and liver transplantation rates in hepatocellular carcinoma patients in the recent era: Results from the surveillance, epidemiology, and end results registry. *Liver Transplantation 20*(5), 528–535.

Wood, S. (2013). Package ' mgcv '. *Available at: https://cran.r-project.org/web/packages/mgcv/index.html*.

Wood, S. N. and N. H. Augustin (2010). GAMs with integrated model selection using penalized regression splines and applications to environmental modelling. *Ecological Modelling 157*(2-3), 157–177.

Wood, S. S. (2006). Generalized Additive Models: An Introduction with R. *Chapman & Hall, UK*, 410.

Yadlowsky, S., R. A. Hayward, J. B. Sussman, R. L. McClelland, Y. I. Min, and S. Basu (2018). Clinical implications of revised pooled Cohort equations for estimating atherosclerotic cardiovascular disease risk. *Annals of Internal Medicine 169*(1), 20–29.

Yang, Y., J. D. Goldhaber-Fiebert, and L. M. Wein (2013). Analyzing screening policies for childhood obesity. *Management Science 59*(4), 782–795.

Yee, A. H., A. A. Rabinstein, P. Thapa, J. Mandrekar, and E. F. M. Wijdicks (2010). Factors influencing time to death after withdrawal of life support in neurocritical patients. *Neurology 74*(17), 1380–1385.

Yeh, H., E. Smoot, D. A. Schoenfeld, and J. F. Markmann (2011). Geographic inequity in access to livers for transplantation. *Transplantation 91*(4), 479–86.

Yokota, F. and K. M. Thompson (2004). Value of information literature analysis: a review of applications in health risk management. *Medical Decision Making 24*(3), 287–298.

Young, K., M. Aguilar, R. Gish, Z. Younossi, S. Saab, T. Bhuket, B. Liu, A. Ahmed, and R. J. Wong (2016). Lower rates of receiving model for end-stage liver disease exception and longer time to transplant among nonalcoholic steatohepatitis hepatocellular carcinoma. *Liver Transplantation 22*(10), 1356–1366.

Zargoush, M., M. Gümüş, V. Verter, and S. S. Daskalopoulou (2018). Designing risk-adjusted therapy for patients with hypertension. *Production and Operations Management 27*(12), 2291–2312.

Zaun, D. A., J. J. Snyder, E. D. Weinhandl, N. J. Salkowski, T. R. Leighton, A. K. Israni, and B. L. Kasiske (2012). Monitoring Performance of Organ Procurement Organization in the United States: Observed and Expected Donor Yield.

Zhang, J., B. T. Denton, H. Balasubramanian, N. D. Shah, and B. A. Inman (2012). Optimization of Prostate Biopsy Referral Decisions. *Manufacturing & Service Operations Management 14*(4), 529–547.