ABSTRACT

MASON, JENNIFER ELIZABETH. Markov Decision Processes and Approximate Dynamic Programming Methods for Optimal Treatment Design. (Under the direction of Brian T. Denton.)

Chronic diseases are a leading cause of death in the United States and other countries. For many chronic diseases, treatment options are available to manage the disease and reduce the risk of adverse events. We present a Markov decision process (MDP) to determine the optimal timing of medications over a patient's lifetime to control the risk of adverse events. A second MDP is used to study the optimal timing of adherence-improving interventions once a patient has begun treatment. Both models are studied in the context of treatment decisions for patients with type 2 diabetes. The first MDP considers medications for the management of blood pressure and cholesterol to reduce the risk of stroke and coronary heart disease events. The states represent the patient's blood pressure and cholesterol. We consider two objectives: maximizing rewards for life years minus costs before the patient's first event, and maximizing rewards for quality-adjusted life years minus costs of treatment over the patient's lifetime. We compare two approximate dynamic programming methods to solve the continuous-state MDP: state aggregation and basis function approximation of the value function. We use multiple basis functions, including survival functions and linear functions in terms of the state variables, and we use a linear program to solve for the basis function weights. We provide theoretical insights into the model, including providing insight into the parameters that are most important for deciding between two medications for a special case of rewards. We also provide extensive numerical results, including a comparison of optimal treatment policies with U.S. and international treatment guidelines. The second MDP considers the use of adherence-improving interventions once a patient has begun medical treatment. Adherence states are defined using enhanced electronic health records connected to pharmacy claims data to define *percentage of days covered*. We present several theoretical properties of the optimal policy. We also provide results from numerical experiments comparing the optimal policy to a simple periodic policy.

Markov Decision Processes and Approximate Dynamic Programming Methods for Optimal Treatment Design

by
Jennifer Elizabeth Mason

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Industrial Engineering

Raleigh, North Carolina

2012

APPROVED BY:

Russell E. King	David L. Roberts
Nilay D. Shah	James R. Wilson

Brian T. Denton
Chair of Advisory Committee

DEDICATION

To my parents and grandparents.

BIOGRAPHY

Jennifer Elizabeth Mason was born on January 30, 1985, in Columbia, SC. Jennifer received her Bachelor of Science in Mathematics from the University of South Carolina Honors College in 2007. She then received her Master of Science in Operations Research from North Carolina State University in 2009.

ACKNOWLEDGEMENTS

First, I would like to thank my advisor Dr. Brian Denton for his ever-present guidance, support, and patience over the last five years. I appreciate all he has done for me to help develop my research and teaching skills, and I know that this training will help me immensely as I start my career. I would also like to acknowledge the funding that supported this research. This work was supported in part by the National Science Foundation under Grant Number CMMI 0968885 (Denton). The research was also supported in part by a Doctoral Dissertation Grant from the Agency for Healthcare Research and Quality under Grant Number 1R36HS020878 (Mason).

I would like to thank my committee members Dr. Russell King, Dr. David Roberts, Dr. Nilay Shah, and Dr. James Wilson for serving on my committee and providing me with helpful suggestions and edits for this dissertation. In addition, a special thanks to Dr. Nilay Shah and Dr. Steven Smith for their help and guidance over the last four years; I owe my understanding of medical problems to both of them.

Finally, I would like to thank all of my friends for their support throughout graduate school. Thank you to Benjamin Lobo for his undying support and encouragement. Thank you to my parents and brother for believing in me.

TABLE OF CONTENTS

List of Tables . . . . . **vii**

List of Figures . . . . . **ix**

Chapter 1 Introduction **1**

Chapter 2 Literature Review **5**
2.1 Basis Function Approximation 7
2.2 Reinforcement Learning 9
2.3 Applications of ADP Methods 14
2.4 Contributions of this Dissertation 17

Chapter 3 Optimal Control of Medication Treatment Initiation **19**
3.1 Introduction 19
3.2 Diabetes Treatment Background and Literature Review 21
3.3 Model 24
3.4 Results 31
3.4.1 Data and Study Population 31
3.4.2 Model Validation 34
3.4.3 Primary Prevention Treatment Policies 37
3.4.4 Primary and Secondary Prevention Treatment Policies 40
3.4.5 Estimated Benefit of the Optimal Guidelines to the U.S. Diabetes Population 43
3.5 Conclusions 51

Chapter 4 Approximate Dynamic Programming Approaches for Optimal Treatment **54**
4.1 Continuous-State MDP Formulation 54
4.2 Finite-State MDP 58
4.3 ADP Approach 1: Policy Mapping 63
4.4 ADP Approach 2: Basis Function Approximation 63
4.4.1 Linear Programming Formulation 64
4.4.2 Basis Functions 68
4.5 Monte Carlo Simulation 71
4.6 Results 73
4.6.1 Sensitivity Analysis 75
4.7 Conclusions 80

Chapter 5 Using Electronic Health Records to Monitor and Improve Adherence to Medication **82**
5.1 Introduction 82

5.2 Background on Medication Adherence . . . 85
5.3 Literature Review . . . 87
5.3.1 Machine Maintenance Applications . . . 88
5.3.2 Medical Decision Making Applications . . . 89
5.4 Model Formulation . . . 92
5.5 Model Properties and Insights . . . 96
5.5.1 Model Assumptions . . . 96
5.5.2 Model Properties . . . 97
5.6 Case Study: Statin Adherence for Patients with Type 2 Diabetes . . . 109
5.6.1 Data and Model Parameter Estimation . . . 110
5.6.2 Numerical Results . . . 112
5.6.2.1 Active vs. Inactive Surveillance . . . 113
5.6.2.2 Sensitivity to Cost of Intervention . . . 115
5.6.2.3 Sensitivity to Individual Patient Risk Factors . . . 117
5.6.2.4 Potential Yearly Benefits of AAS to the U.S. Diabetes Population 118
5.7 Conclusions . . . 121

Chapter 6 Conclusions . . . 124

References . . . 129

LIST OF TABLES

Table 3.1 International guideline thresholds for initiation of cholesterol and blood pressure medications. Guidelines that assume diabetes patients are not considered CHD risk equivalent are represented with *. LDL is measured in mg/dL for the U.S. guidelines, and LDL, HDL, and TC are measured in mmol/L for all other guidelines. LR is unitless, and SBP is measured in mmHg. 23
Table 3.2 Ranges for TC, HDL, and SBP states based on [24]. 27
Table 3.3 Baseline characteristics for the study population (N = 663), including mean and variance. 32
Table 3.4 Percentage change in risk factors for given medications as computed from Mayo Electronic Medical Records and Diabetes Electronic Management System. 33
Table 3.5 Description of model parameters including cost inputs and utility decrements for the reward function of the MDP model. 33
Table 3.6 Costs and utility decrements for each medication used in the model. 34
Table 3.7 Male comparison of expected LYs before death, expected LYs before a stroke or CHD event, and expected LYs after an event from age 50 for our MDP model and the Framingham Heart Study (FHS). The 95% confidence intervals are provided for the FHS estimates. 36
Table 3.8 Female comparison of expected LYs before death, expected LYs before a stroke or CHD event, and expected LYs after an event from age 50 for our MDP model and the Framingham Heart Study (FHS). The 95% confidence intervals are provided for the FHS estimates. 36
Table 3.9 Yearly costs (billions) and future event-free LYs for newly diagnosed diabetes patients using no treatment, optimal guidelines (R_0 = \$100,000, R_0 = \$250,000, and R_0 = \$10 million), and U.S. I. 50
Table 3.10 Yearly costs (billions) and future QALYs for newly diagnosed diabetes patients using no treatment, optimal guidelines (R_0 = \$100,000, R_0 = \$250,000, and R_0 = \$1 billion), and U.S. I. 50

Table 4.1 Parameter values for Equations (4.37), (4.38), (4.39), and (4.40) found in Stevens et al. [97] and Kothari et al. [61]. 69
Table 4.2 Comparison among the ADP methods and no treatment of expected QALYs before a stroke, CHD event, or death from other causes for males and females. For each simulation, 120,000 patients are sampled. The 95% confidence intervals for simulated results are provided in parentheses. . . . 73
Table 4.3 Sensitivity analysis results for base case probabilities and 50% higher medication decrements to QALYs. 75
Table 4.4 Sensitivity analysis results for base case probabilities and 50% lower medication decrements to QALYs. 76

Table 4.5 Sensitivity analysis results for 25% higher probabilities and 50% higher medication decrements to QALYs. 76
Table 4.6 Sensitivity analysis results for 25% higher probabilities and base case medication decrements to QALYs. 77
Table 4.7 Sensitivity analysis results for 25% higher probabilities and 50% lower medication decrements to QALYs. 77
Table 4.8 Sensitivity analysis results for 25% lower probabilities and 50% higher medication decrements to QALYs. 78
Table 4.9 Sensitivity analysis results for 25% lower probabilities and base case medication decrements to QALYs. 78
Table 4.10 Sensitivity analysis results for 25% lower probabilities and 50% lower medication decrements to QALYs. 79
Table 4.11 Sensitivity analysis results for base case probabilities and 5 times higher medication decrements to QALYs. 79

Table 5.1 Adherence States Defined by Percentage of Days Covered (PDC) and the Corresponding Percent Change in Total Cholesterol (TC) for Patients that Initiate Statins. 88
Table 5.2 Initial hospitalization costs and follow-up events for adverse events. 111
Table 5.3 Optimal ages to begin having yearly interventions for female patients using active surveillance. Imperfect (probabilistic) interventions are assumed. Note: '–' denotes it is never optimal for the patient to have interventions. . 117
Table 5.4 Optimal ages to begin having yearly interventions for female patients using active surveillance. Perfect interventions are assumed. 118
Table 5.5 Yearly costs (billions) and future LYs for newly-diagnosed diabetes patients using no adherence interventions, yearly inactive adherence surveillance (IAS, $k = 1$), and active adherence surveillance (AAS). 120

LIST OF FIGURES

Figure 3.1 Simplified state transition diagram for the case of two medications. When medications are initiated (actions denoted by the solid lines), the risk factors are improved and the probability of the occurrence of an adverse event (denoted by the dashed lines) is reduced. 25

Figure 3.2 Comparison of optimal treatment policies for male patients to treatment by U.S. and international guidelines. 38

Figure 3.3 Comparison of optimal treatment policies for female patients to treatment by U.S. and other international guidelines. 39

Figure 3.4 Histograms to provide the difference in LYs and medication costs for males between the optimal guidelines ($R_0 = \$100{,}000$) and U.S. I. 44

Figure 3.5 Histograms to provide the difference in LYs and medication costs for females between the optimal guidelines ($R_0 = \$100{,}000$) and U.S. I. . . . 45

Figure 3.6 Comparison of optimal treatment policies for male patients to treatment by U.S. and international guidelines. 46

Figure 3.7 Comparison of optimal treatment policies for female patients to treatment by U.S. and international guidelines. 47

Figure 3.8 Histograms to provide the difference in QALYs and medication and treatment costs for males between the optimal guidelines ($R_0 = \$100{,}000$) and U.S. I. 48

Figure 3.9 Histograms to provide the difference in QALYs and medication and treatment costs for females between the optimal guidelines ($R_0 = \$100{,}000$) and U.S. I. 49

Figure 4.1 Example of the partitioned continuous state space for LR and SBP. For this particular partitioning, which is used in the numerical experiments, LR is divided into three ($q_{\text{LR}} = 3$) discrete states (low (L), medium (M), and high (H)), and SBP is divided into four ($q_{\text{SBP}} = 4$) discrete states (low (L), medium (M), high (H), and very high (V)). The dots in each cell of the partition represent the conditional mean LR and SBP values for the cell. 59

Figure 4.2 Example of the tile coding in which the bounded continuous state space for LR and SBP is partitioned using two tilings. One tiling is shown with solid lines dividing the state space, and the other tiling is shown with a dotted line, providing a single tile over the entire state space. 68

Figure 5.1 Diagram of Prescription Refills Used to Calculate the Percentage of Days Covered (PDC). 87

Figure 5.2 Comparison of expected LYs verses costs for medication, interventions, and treatment of events for active adherence surveillance (AAS) policies (with varying R values) and inactive adherence surveillance (IAS) policies (when interventions occur every k years) for female patients using imperfect interventions. Results are a weighted average of LYs and costs for the 16 possible risk states. 114
Figure 5.3 Comparison of expected LYs verses costs, as shown in Figure 5.2, for male patients. 114
Figure 5.4 Comparison of expected LYs verses costs for medication, interventions, and treatment of events for active adherence surveillance (AAS) policies (with varying R values) and yearly inactive adherence surveillance (IAS) for female patients using imperfect interventions. Results are compared for low, medium, and high risk patients. 116

Chapter 1

Introduction

Chronic diseases are the leading cause of death in the United States and other countries, accounting for seven out of ten deaths each year [62]. Fortunately, for many chronic diseases there are treatment options to manage the disease and reduce the risk of adverse events or death related to the disease. However, in many cases the cost of treatment is high, and some treatments have side effects that can reduce a patient's quality of life. In light of these facts, the optimal control of treatment for chronic diseases is very important. Improving treatment plans for chronic diseases has the potential to prolong lives, improve quality of life, and reduce costs.

Diabetes is a good example of a chronic disease that is very prevalent and has many treatment options. The American Diabetes Association estimates that 25.8 million people in the United States (over 8% of the population) have diabetes [19]. Of the affected population, more than 90% have type 2 diabetes. Diabetes patients are complex patients, often at greater risk than the general population for serious health outcomes such as cardiovascular disease. For instance, two out of three deaths of diabetes patients are caused by stroke or coronary heart disease (CHD) [3]. Blood pressure and cholesterol medications are often part of treatment plans for diabetes patients, helping them to manage their risk of stroke and CHD events. However, these medications can have side effects that reduce a patient's quality of life. Some treatment

options are expensive, and the total cost of treatment for such a large portion of the population can be high. Therefore, the optimal time and order to initiate drug treatments (if at all) over the course of a patient's lifetime is unclear.

Treatment optimization problems can pose many challenges. First, there are advantages and disadvantages to initiating medications. While treatment has the long-term benefit of reducing the probability of serious health outcomes, this must be traded off against the burden of taking medication, side effects, and the monetary cost of treatment. Second, if treatment is initiated, the decision is further complicated by choosing which medications to initiate and in which order. The large number of treatment options is coupled with a large state space that defines the possible health states of the patient based on risk factors, medication history, and the occurrence of adverse events. Uncertainties in the effects of treatment and the evolution of a patient's health state as he or she ages further complicate the decision process.

In this dissertation we present two related Markov decision process (MDP) models to study optimal treatment plans for patients with type 2 diabetes. The models are finite horizon nonstationary MDPs that incorporate the use of diabetes risk models to determine optimal decisions related to diabetes treatment control. The first model considers the optimal timing of blood pressure and cholesterol medications over the course of the patient's lifetime, with blood pressure and cholesterol represented by a finite set of states. We also extend this model to consider the continuous nature of the blood pressure and cholesterol risk factors. Due to the continuous state space, we use approximate dynamic programming (ADP) methods to solve this MDP. The ADP methods we use can be generalized to other problems where the goal is to find near-optimal solutions to MDPs with large state spaces. The second model considers the problem of imperfect adherence to medication; we use this MDP to study the optimal timing of interventions to improve patient adherence to medication once a single medication has been initiated. We present theoretical properties of the MDP, and we use the model to predict the potential impact of a new automated adherence surveillance program for patients with type 2 diabetes in the United States. Following is a detailed description of the research discussed in

Chapters 3, 4, and 5.

In Chapter 3 we describe an initial MDP model for optimal control of medication initiation decisions. This model aims to answer the following question: When and in what order should medications be initiated to reduce the risk of adverse health events? The time horizon is defined by a finite set of annual decision epochs. The states represent the patient's health status based on risk factors for cardiovascular disease and stroke. The actions represent initiating medications or deferring initiation at each epoch. Once patients initiate a medication, they are assumed to remain on that medication for the remainder of their lives. The yearly rewards include a monetary reward for the patient's quality of life minus the costs of treatment. A series of experiments are performed to evaluate the trade-off between these competing criteria. We compare expected outcomes for the optimal policy with outcomes for current blood pressure and cholesterol initiation guidelines in the United States and around the world. In general, we find that male patients should initiate treatments earlier than female patients. We report a number of findings related to the optimal sequence and time of treatment. We also present structural properties related to the initiation of one medication over another and the benefits of coordinated treatment.

In Chapter 4, we use ADP methods to approximate solutions to the true continuous state MDP underlying the approximate discrete-state MDP formulation of Chapter 3. Since the continuous-state MDP is computationally intractable, we use a basis function approximation of the true value function for the continuous-state MDP. While the discrete-state MDP formulation of Chapter 3 provides one approximation, we experiment with basis function approximation methods to find alternative policies that will achieve the best results according to specific criteria. Numerical experiments are performed using a simulation model to compare the different policies found from the ADP approach and the MDP model from Chapter 3.

Despite the long-term benefits of treatment, some patients do not take all their prescribed medication due to factors such as forgetfulness, poor understanding of how the preventative medication works, side effects, and costs. This represents a serious barrier to achieving the full

benefits of treatment. In Chapter 5 we formulate a new MDP model to answer the following question: When should adherence-improving interventions occur for patients who have already initiated medication? We use this MDP model to determine the optimal policy for interventions based on the patient's adherence state. The time horizon is defined by a finite set of annual decision epochs. The states represent the percentage of time that the patient takes his or her medication as prescribed based on pharmacy claims data. The actions are deciding to have an intervention or deferring an intervention to a later stage. This decision is revisited each year regardless of the actions chosen in the previous years. The yearly rewards for this model are a monetary reward for the patient's quality of life minus the costs of treatment and interventions. We find that the optimal policy is highly dependent on the cost and effectiveness of the intervention. In addition, we prove that the optimal policy is monotonic with respect to the patient's adherence state. We also prove theorems providing insight into how the optimal control limit changes when interventions with different effects are considered and when patients of different health statuses are considered.

The remainder of this dissertation is organized as follows. Chapter 2 presents a review of the ADP literature and applications of ADP methods to health care problems. Chapter 3 presents the first MDP model to determine the optimal sequence and timing of blood pressure and cholesterol medications over the course of a patient's lifetime. Chapter 4 outlines the ADP methods and provides numerical results comparing the different policies for the multiple medication problem. Chapter 5 presents the second MDP model to determine the optimal timing of adherence-improving interventions for patients who have already initiated treatment. Finally, Chapter 6 provides a summary of the most significant conclusions.

Chapter 2

Literature Review

The treatment of chronic diseases involves sequential decision making under uncertainty: actions (e.g., initiating a medication or waiting to treat) must be taken today without knowing what effect the actions will have on a patient's health status or what the natural progression of a patient's health will be. The need for sequential decision making under uncertainty also arises in many other settings such as machine maintenance, inventory control, and artificial intelligence.

This literature review highlights solution methods for finite-horizon, discrete-time problems for which the Markov assumption holds. In particular, the primary model to define such problems is the MDP. MDPs are Markov processes that can be controlled through actions. At each stage of the process, the optimal action is taken according to some criteria as governed by optimality equations. A finite-horizon, discrete-time MDP is defined by the following. The decision horizon is defined by a discrete set of decision epochs indexed by $t = 1, \ldots, T$. The states in the MDP, $s \in S$, capture all information needed to make decisions. The set of possible actions is defined by $a \in A$. Transition probabilities among the states define the probability of being in state s' at time $t+1$ given the state at time t is s and the action taken at time t is a:

$$p_t(s'|s,a) = \Pr\{s_{t+1} = s'|s_t = s, a_t = a\}. \tag{2.1}$$

Rewards, $r_t(s,a)$, are accrued in each time period and depend on the state at time t and the

action taken. In each decision epoch for each state, the optimal value function, $v_t(s)$, for a finite horizon MDP is defined by the optimality equations:

$$v_t(s) = \max_{a \in A} \Big\{ r_t(s,a) + \lambda \sum_{s' \in S} p_t(s'|s,a) v_{t+1}(s') \Big\}, \tag{2.2}$$

along with the following boundary condition at stage T:

$$v_T(s) = \mu_T(s), \tag{2.3}$$

where $\lambda \in (0,1]$ is the discount factor and $\mu_T(s)$ is the expected future rewards accrued after the end of the decision horizon. The optimal action, $a_t^*(s)$, at time t for state s is chosen based on the optimality equations:

$$a_t^*(s) = \operatorname*{argmax}_{a \in A} \Big\{ r_t(s,a) + \lambda \sum_{s' \in S} p_t(s'|s,a) v_{t+1}(s') \Big\}. \tag{2.4}$$

Solution methods for MDPs have been well established [81]; however, MDPs can become more difficult to solve or even intractable when the number of states and/or actions grows large or is infinite. This phenomenon is often referred to as the *curse of dimensionality* [80]. This curse can arise when considering medical decision making problems. For example, for some chronic diseases, there are many possible treatment options available (especially when considering dosage of medications and treatment of multiple risk factors), leading to a large action space. In addition, the state space can grow very large for medical decision making problems when the state includes multiple patient risk factors, the state space is continuous, or the Markov process is of higher order incorporating the dependence of state transitions on the history of the patient's health.

Even as computing power continues to improve, alternative methods are needed to address the inadequacies of solution methods for MDPs with large state and action spaces (where the problem size can be too large to store a look-up table of the optimal action for each state). This literature review highlights ADP methods to address the challenge of finding optimal policies

for MDPs that suffer from the curse of dimensionality. The ADP methods addressed include basis function approximation of the value function and sampling-based reinforcement learning (RL) techniques such as *Q-learning*. We also highlight particular examples of ADP techniques applied to different types of health care problems, including medical decision making problems.

The remainder of this chapter is organized as follows: Section 2.1 reviews the literature related to basis function approximation of the value function. Section 2.2 highlights the main methods and algorithms related to RL. Section 2.3 provides examples of health care applications of ADP methods. Finally, Section 2.4 highlights the main contributions of this dissertation.

2.1 Basis Function Approximation

Approximation methods have long been used to solve problems for which computing power is inadequate. Bellman and Dreyfus propose the use of functional approximations to solve dynamic programming problems [13]. A function is approximated by adding up other functions, referred to as *basis functions*, multiplied by appropriately chosen coefficients. Bellman and Dreyfus use the basis function approximation method to solve a recurrence relation. The basis functions used in their example are Legendre polynomials.

Basis function approximation has become a common approach for many types of problems (e.g., solving systems of partial differential equations). Powell (see Chapter 7) [80] describes the method of basis function approximation for estimating the value function of an MDP as a way of dealing with the curse of dimensionality. With this method, the number of state variables can be greatly reduced to the number of coefficients for the basis functions. The estimated value function is given by the following:

$$\tilde{v}(s) = \sum_{k \in K} w_k f_k(s), \tag{2.5}$$

where K is the set indices for the basis functions and the coefficients, w_k, that serve as the weights. In addition, the basis functions are functions of the state. Thus, the approximation

in Equation (2.5) reduces the dimensionality of the underlying problem to the selection of $|K|$ parameters.

Schweitzer and Seidmann [88] present a general framework for approximation of the value function for a stationary, infinite horizon semi-MDP, with finite state space. For problems with large state spaces, traditional solution methods for infinite-horizon MDPs (linear programs (LPs), value iteration, and policy iteration) may be too time consuming or even infeasible. For these large-scale problems it is important to reduce the size of the problem, and it may not be as important to have the optimal solution if a near-optimal solution is achievable with less effort. The value function may be approximated with K basis functions. Solving for the coefficients of these basis functions instead of using the traditional solution methods reduces the dimension of the problem from the number of the states to K. Schweitzer and Seidmann present three algorithms for estimating the values of the coefficients of the basis functions both with and without discounting: linear programming, policy iteration, and least squares. They also provide a framework for assessing the quality of the approximations.

De Farias and Van Roy [28] extend the work of Schweitzer and Seidmann by providing theoretical guarantees (error bounds) on the performance of the linear programming approach to ADP using basis functions. In addition, De Farias and Van Roy emphasize the importance of choosing appropriate coefficients for the value-function approximations in the objective function. These state-relevant weights, represented by the column vector c, affect the quality of the solution found using the approximate LP, provided here for a cost minimization problem:

$$\begin{aligned} &\max \sum_{s \in S} \left[c(s) \sum_{k \in K} w_k f_k(s) \right] \\ &\text{s.t. } r(s,a) + \lambda \sum_{s' \in S} p(s'|s,a) \sum_{k \in K} w_k f_k(s') \geq \sum_{k \in K} w_k f_k(s), \forall s \in S, a \in A. \end{aligned} \tag{2.6}$$

A suggested method for defining these coefficients $c(s)$ associated with each state s is to assign the estimated frequency with which each state will be visited. Examples of applying the approx-

imate LP solution method were provided for an uncontrolled queuing system and a controlled queuing system in which the service rate could be controlled.

De Farias and Van Roy [29] further extend the use of the approximate LP to estimate weights for basis functions. While the approximate LP may only have a small number of variables the number of constraints could be intractable, particularly when the action space is large. A constraint sampling method is presented to create a reduced linear program that still provides a near-feasible solution. A controlled queuing system is again used to demonstrate how the method could be applied.

There are several potential types of basis functions that have been shown to perform well in practice. Many sets of basis functions form a complete set of functions over $L^2(\mathbb{R})$. There are infinitely many functions in these sets, and as more of these basis functions are used to approximate the unknown function, a better approximation is achieved. As the number of basis functions used from this set approaches infinity, the approximate function approaches the true function, with appropriately chosen weights for the basis functions. Some examples of complete sets of basis functions include Legendre polynomials, radial basis functions, and Fourier series (where Fourier series form a complete set over $L^2([0, 2\pi])$).

For medical decision making problems, the topic of this dissertation, hazard functions can be used for the basis functions, as used by Lee et al. [64] (reviewed below in Section 2.3). Hazard functions can work well for medical decision making problems because often the state space includes inputs for the hazard functions. While there is no guarantee that use of this type of function will provide a good approximation of the value function being estimated (over $L^2(\mathbb{R}^+)$), these functions appear to be an intuitive choice for medical decision making problems.

2.2 Reinforcement Learning

Sutton and Barto [98] define RL as "learning what to do – how to map situations to actions – so as to maximize a numerical reward signal." In RL models, an agent interacts with the environment and takes actions which results in rewards. The main elements of RL are the model

of the environment (anything that cannot be changed by the agent), a policy that describes how the agent acts under certain conditions, rewards that indicate what is favorable in the short term, and a value function that provides information about what is favorable in the long term. The decision taken by the agent depends on the value associated with each action in the given state. RL models are closely related to MDPs, and solution methods for MDPs, such as dynamic programming, can also be considered RL techniques. RL models are divided into two categories: episodic tasks that occur for a finite length of time (e.g., finite-horizon MDPs) and continuing tasks that go on for an infinite amount of time (e.g., infinite-horizon MDPs). For the purpose of RL algorithms, $V^{\pi}(s)$ is the expected value of starting in state s and proceeding with policy π, and the *action-value function* $Q^{\pi}(s, a)$ is the expected value of starting in state s, taking action a, and proceeding with policy π thereafter. RL algorithms can be more useful than traditional dynamic programming algorithms when a perfect model of an MDP is not available. This may be the case, for example, when a large number of transition probabilities must be estimated. In addition, RL algorithms often require less computational effort.

The class of RL algorithms, including value iteration, can be described by the umbrella term generalized policy iteration (GPI). GPI involves both policy evaluation and policy improvement. Through policy evaluation the value, $V^{\pi}(s)$, of starting in a given state and proceeding with policy π is computed for each state. Policy improvement determines whether the current policy can be improved upon by comparing $V^{\pi}(s)$ and $Q^{\pi}(s, a)$.

RL algorithms can be combined with Monte Carlo methods for episodic tasks when the state of the environment is not completely known. Instead of assuming a known environment, Monte Carlo methods are based on observations from an actual environment or simulated experience from a model that is capable of generating sample transitions. GPI using Monte Carlo sampling retains and updates an approximate policy and an approximate value function at each iteration. Policy evaluation for Monte Carlo methods relies on the outcomes of many sample episodes; policy improvement takes the estimated action-value function $Q^{\pi}(s, a)$ and picks the greedy policy for each state s. Monte Carlo methods can be particularly useful when only a small

subset of states must be evaluated since the computational effort required to estimate the value of a particular state is independent of the total number of states.

Gosavi [45] presents a tutorial and recent advances in RL. Gosavi briefly describes several methods for solving discrete, stationary, infinite-horizon control problems with RL. He presents Q-learning methods that employ sampling with value iteration and policy iteration methods based on the Robbins-Monro (RM) stochastic approximation. The RM algorithm [85] was originally used to determine a unique root of a function. In Q-learning, the mean, $E[X]$, of a random variable X can be estimated using the RM algorithm, where X_m represents the sample from the m^{th} iteration, Y^m is the estimate of the mean from the m^{th} iteration, and μ_m is the m^{th} step size. The RM algorithm is provided in Algorithm 1. Convergence of this algorithm is guaranteed if $\lim_{M\to\infty}\sum_{m=1}^{M}\mu_m = \infty$ and $\lim_{M\to\infty}\sum_{m=1}^{M}\mu_m^2 < \infty$. One example step size that satisfies these conditions is $\mu_m = \frac{1}{1+m}$.

Algorithm 1 The Robbins-Monro Algorithm.

Step 1
$m := 1$
Y^0 is set to any arbitrary number
Specify $\epsilon > 0$

Step 2
$Y^m \leftarrow (1-\mu_m)Y^{m-1} + \mu_m X_m$

Step 3
if $|Y^m - Y^{m-1}| < \epsilon$ **then**
stop
else
m := m+1
return to **Step 2**
end if

Gosavi also briefly introduces TD-learning (temporal differences) methods. This set of methods is a generalization of the RM stochastic approximation. In TD-learning, feedback from the system and previous estimates of the value function are used to produce an improved

estimate of the value function. If W^k is the estimate from the k^{th} iteration, the TD-learning algorithm defines the next iterate as the following:

$$W^{k+1} \leftarrow (1-\mu)W^k + \mu[\text{feedback}_k], \tag{2.7}$$

where μ again represents the step size, and in the second term μ is multiplied by the value of feedback. The value of feedback depends on the type of TD-learning algorithm being used. However, the feedback typically is a function of the immediate reward. For the purpose of trying to estimate the value of state-action pairs, a unique W will represent each pairing. For the case of maximizing rewards, larger feedback will increase the likelihood that a particular action should be taken while negative feedback has the opposite effect. As trials proceed, the optimal policy is learned. TD(0) is the same as the RM algorithm, while TD(λ) incorporates more rewards in the future. For TD(λ), the feedback is given by the following:

$$\text{feedback}_k = r_k \sum_{i=0}^{\infty} \lambda^i r_{k+i}, \tag{2.8}$$

where $\lambda \in [0,1]$ is the discount factor and r_k is the immediate reward from iteration k.

Q-learning [111] is an algorithm that is related to the RM algorithm and TD-learning. Q-learning uses simulation in a model-free context, in which transition probabilities are not assumed, to update the action-value function, $Q(s,a)$. The Q-learning algorithm for the k^{th} iteration is provided in Algorithm 2 [112]. Note that $V_{k-1}(y) \equiv \max_b\{Q_{k-1}(y,b)\}$ in the Q-learning algorithm.

With Q-learning, there is a trade-off between *exploration* and *exploitation* that can be addressed by choosing the appropriate gain factor at each iteration. With exploration, sample paths visit many different states to widely update the value function approximations. In contrast, exploitation involves simulation of sample paths that are the same or similar, providing a very good approximation of the value function for only a small subset of states. For very large (or continuous) state spaces, a look-up table for Q-values in each particular state becomes

Algorithm 2 Q-Learning Algorithm.

Step 1

Observe the current state s_k
Select and take action a_k
Observe the next state y_k
Receive immediate reward r_k

Step 2

Update Q_{k-1} values based on the learning factor α_k:
if $s = s_k$ and $a = a_k$ **then**
$Q_k(s,a) = (1-\alpha_k)Q_{k-1}(s,a) + \alpha_k[r_k + \lambda V_{k-1}(y_k)]$
else
$Q_k(s,a) = Q_{k-1}(s,a)$
end if

impractical (or impossible). In such situations Q-values need to be approximated using regression or some other state space approximation. Gosavi also discusses further extensions, recent advances in RL, and issues of convergence.

Kaelbling et al. [58] also present a survey of RL techniques. Kaelbling et al. describe RL as "the problem faced by an agent that must learn behavior through trial-and-error interactions with a dynamic environment." Appropriate actions for the stochastic environment are chosen by searching the space of possible actions (with techniques such as genetic algorithms) to find the one that performs the best, or by using dynamic programming techniques to estimate utilities of taking certain actions. Kaelbling et al. also discuss the dilemma of exploitation versus exploration. Kaelbling et al. provide model-free methods (including TD-learning and Q-learning) in which structure of the model is not assumed, and model-based methods (including certainty equivalent methods for learning the model) in which known data is used more efficiently to learn the model and determine the best actions. In addition, Kaelbling et al. present models for partially observable environments and use of some of the models presented to applications including game play of backgammon and robotics.

In general, Q-learning and other learning techniques are best suited for model-free problems. To use Q-learning for high dimensional problems, it must be combined with dimension reduction

methods such as aggregation and basis function approximation of the value function.

2.3 Applications of ADP Methods

ADP methods have been applied to problems in many different settings including applications such as energy allocation, vehicle routing, and backgammon. More recently ADP methods have been used for health care applications, including patient scheduling and medical decision making problems. In this section we present a summary of the ADP health care applications.

Maxwell et al. [67] use an ADP approach to determine the best strategy for dynamic repositioning of ambulances in metropolitan areas in order to maximize the number of calls reached within a designated length of time. The problem is formulated as an MDP with the state space including information about the number of ambulances and the number of waiting calls in the emergency medical services system. The ambulance classification includes the ambulance's status, location, and timing of any movement. The call classification includes the call's status, location, timing, and priority level. Decisions can only be made when events occur, such as a call coming in or an ambulance transporting a patient to the hospital. Only ambulances that have just finished transporting a patient to the hospital are available for redeployment to one of the possible ambulance bases. If a call with high priority is answered in more than the designated time, a cost of 1 is incurred; otherwise, there is no cost. This simple cost function does not assign different costs according to the length of time past the threshold that the call is answered. While this may be a shortcoming, the authors note that high-priority calls are served first since calls are attended to by priority level, and the modeling framework is general enough that it could take medical outcomes into account as done by Erkut et al. [36]. The objective of the MDP is to find the policy that minimizes the total discounted number of calls that cannot be answered within the threshold time.

Approximation techniques are needed to solve the ambulance redeployment problem, as the state space is uncountable. The approach used is basis function approximation of the value function in combination with approximate policy iteration. Monte Carlo simulation is used

to evaluate expected costs. Six basis functions are used: $\phi_1(s) = 1$ to allow for the value function to be shifted, and functions to describe assigned calls that cannot be reached within the threshold, the rate of calls that cannot be met by the threshold because of ambulance distance, the rate of missed calls due to queueing issues, and the same two rates for future calls. The basis function coefficients are estimated using the cost projections found through Monte Carlo simulation. Two examples of implementation are provided for two large metropolitan areas to show the benefit of the approximately optimal policies over traditional deployment policies.

Patrick et al. [75] present a discounted, infinite-horizon MDP to schedule appointments for incoming patients of different priority levels while meeting requirements for priority-specific wait times. The model considers an N day planning horizon. The state includes information about the number of patients currently scheduled over the planning horizon and the number of patients of each priority type that are waiting to be scheduled. At each decision epoch, the person in charge of scheduling the patients must decide which appointment slots to allocate to patients waiting to be assigned. If there is more demand than time slots, then the action of diverting patients to other facilities may also be taken. Costs are associated with booking patients beyond their targeted time, diverting patients, and leaving patients unbooked. An ADP method is proposed to deal with the large state space of this MDP. Patrick et al. use a basis-function approximation of the value function and solve for the weighting parameters using an approximate LP. They also show how to derive a policy from the solution to the approximate LP. Simulation of the policy is used to estimate performance to the wait time targets in practice.

He et al. [49] present an MDP to determine the optimal treatment dosage during the controlled ovarian hyperstimulation (COH) of in vitro fertilization-embryo transfer therapy. Higher doses lead to a better chance of pregnancy, but there is also a risk of ovarian hyperstimulation syndrome (OHSS). The state space is made up of three continuous dimensions that help predict the probability of a successful pregnancy and the rate of OHSS. The decision to be made is

the daily dose of treatment the doctor should prescribe. Transition probabilities are defined among the health states, and a cost function rewards health measurements in the target treatment range. The objective is to minimize the expected cost over the COH time horizon. One solution method presented for the COH problem is to discretize the state space and solve the problem using backward induction. The authors also present an ADP method to approximate the value function as an alternative method to deal with the continuous state space problem [50]. The RM algorithm is implemented, and the value function is approximated using two different sets of piecewise linear basis functions. The basis function approximations result in costs 1% to 3% higher than the solution found by solving the MDP. However, solving the problems using basis function approximation of the value function took less than 20 seconds while the MDP took over 40 hours to solve.

Hsih et al. [53, 87] present an optimal learning approach for glycemic control for patients with type 2 diabetes. The main learning approach used is knowledge gradient. This approach uses one-step look ahead to determine the value of learning with each possible action. The state space includes the patient's fasting plasma glucose, HbA1c, body mass index, and presence of side effects from medication. Monthly time steps are used, and the decision made by the doctor is to recommend blood sugar management through diet and exercise or one of four glycemic control medications. A utility function is defined in terms of the patient's fasting plasma glucose, and the objective is to maximize the expected utility over the entire horizon. Learning techniques with one-step look ahead, including the knowledge gradient method, are used to learn the transition probabilities among the health states for each of the medication states. Numerical experiments show that this method can yield better policies than policies derived from a model which assumes estimated probabilities.

Lee et al. [64] use an ADP method to develop near-optimal strategies for initiation and control of dialysis for patients with chronic kidney failure. Cost effectiveness of current strategies is evaluated, and these strategies are compared to near-optimal policies using a previously developed and validated simulation model. Two objectives are considered with the ADP strategy:

maximization of rewards for QALYs minus costs for all medical expenditures, and minimization of costs subject to no reduction in QALYs relative to current practice. The dialysis problem is formulated as an MDP with a continuous-time model for disease progression and discrete-time control of doses used in dialysis. Lee et al. also show how the MDP can be transformed into a discrete-time stochastic shortest path problem. The MDP is solved using approximate policy iteration, incorporating the steps of policy evaluation and policy iteration. The approximate value function is expressed as the sum of basis functions multiplied by weights and the weights are estimated using the simulation model by finding weights that minimize the squared difference between the simulated rewards and the approximate value function. The basis functions are hazard rates of events dependent on the dose of dialysis, quality of life estimates for each of the dose levels, and the logarithm of each of the hazard rates and quality of life estimates. Complex dose strategies that were dependent on patient risk factors proved to be cost effective and could potentially reduce expected costs of treating patients with chronic kidney failure.

2.4 Contributions of this Dissertation

We present two novel MDP formulations of important applications: optimal control of multiple risk factors to manage the risk of stroke and CHD events, and optimal timing of adherence-improving interventions for patients who have already initiated medical treatment. We also present structural properties for each of these models that could be widely applicable to other medical decision making problems and possibly problems in other contexts. For the multiple medication model we show that a very simple comparison between QALYs and expected time to first event (e.g., stroke or CHD event) could be used to choose to initiate one medication over another. This theorem could motivate heuristics for choosing the ordering of multiple medications when many medications are being considered. For the adherence-improving interventions MDP, we prove several structural properties including the existence of an optimal control limit under certain conditions. We provide insight into the relationship of control limits for different interventions and different patients.

The ADP methods presented in this dissertation use basis-function approximation of the value function to estimate the true value function for the continuous state space multiple medication MDP based on the discrete-state MDP presented in the next chapter. Similar to Lee et al., we use hazard functions as basis functions. In Chapter 4, we compare the policies found using basis-function approximation to the policy found using state aggregation to discretize the state space of the multiple medication MDP. To our knowledge, we are the first to use constraint aggregation in an LP to solve an MDP.

Chapter 3

Optimal Control of Medication Treatment Initiation

3.1 Introduction

Currently 25.8 million people in the United States have diabetes. Approximately 1.9 million people aged 20 or older were newly diagnosed with diabetes in 2010 [19]. Treatment of the diabetes population can be costly; currently it is estimated that $113 billion per year in direct medical costs is spent for diabetes-related treatment in the United States [55]. This yearly cost is expected to triple in the next 25 years. Part of the challenge of controlling costs is reducing medication costs for diabetes patients. Another consideration for controlling costs is to prevent or delay the occurrence of stroke and CHD events for which diabetes patients are particularly at risk, thereby reducing the hospitalization and other significant follow-up costs associated with these events.

In addition to cost, important criteria from the patient perspective include life span, expected time to first event, and quality-adjusted life span. Expected time to first event (e.g., stroke, CHD event, or death) is often used as a measure of primary prevention to evaluate the benefits of treatment [25, 84, 72]. Expected quality-adjusted life years (QALYs), on the other

hand, can be used as a measure of primary and secondary prevention: QALYs trade off the benefits of treatment, including increase in event-free years, with the reduction in quality of life due to treatment or the occurrence of adverse events. QALYs are one of the most commonly used criteria in the health policy literature [42]. They define a measure of a life year on a 0 to 1 scale based on a patient's health status, with QALY decrements to represent the burden of treatment or minor illnesses and debilitating diseases or events such as stroke or CHD events. QALYs can be used to measure the trade-off between the burden of treatment and the benefits of prevention of stroke and CHD events.

Recently, several risk models have been developed to predict the probability of complications of type 2 diabetes over the course of an individual's lifetime [61, 97, 96, 34, 35]. These models serve as a guide to clinicians for establishing the importance of treatment; however, there has been little investigation of how to effectively use these risk models to design optimal treatment policies for blood pressure and cholesterol management. The research presented in this chapter seeks to bridge this gap by furthering the basic knowledge of how to optimally treat cardiovascular risk in patients with diabetes over the course of their lifetime.

We present an MDP to determine the optimal timing of medical treatment decisions for blood pressure and cholesterol control in patients with type 2 diabetes. We consider two different bi-criteria formulations of our MDP problems. First, we use our model to find the optimal treatment decision that trades off the expected time to first event and the cost of medication. Second, we use our model to find the optimal treatment decisions that trade off expected QALYs and total costs of treatment (medication costs plus one-time and follow-up treatment costs for adverse events). In both cases we combined the two criteria using a willingness-to-pay factor to balance life years (LYs) and QALYs against the costs of medication and treatment, respectively. We vary the willingness-to-pay factor to estimate the efficient frontier of treatment policies. We also evaluate the most common treatment guidelines in the United States and other countries applied to U.S. patients and compare them to the Pareto-optimal policies from our model.

Our model considers control of coexisting stochastic risk factors, which is a problem that

arises in the context of many chronic diseases. There is a significant literature on treatment optimization. However, to our knowledge ours is the first to examine simultaneous control of multiple risk factors. We highlight the benefits of coordinated treatment over the myopic nature of current guidelines by comparing costs, QALYs, and event-free LYs for the different policies. We also present structural properties for the order of treatment initiation when primary prevention is considered, and we provide discussion on the benefits of coordinated treatment for multiple risk factors.

We address several specific research questions in this chapter including the following: How much can coordinated management of coexisting risk factors improve patient outcomes (e.g., QALYs, LYs before an adverse health event) over current guidelines? What effect does treatment coordination have on costs? How should treatment plans differ for males and females? How dependent is the optimal treatment regimen on an individual patient's metabolic risk profile? To help answer these questions, we present patient-specific treatment plans based on our model. We also compare expected LYs and medication costs, and expected QALYs and total costs for optimal treatment plans and current practice guidelines.

The remainder of this chapter is organized as follows: In Section 3.2 we provide background on diabetes treatment and a review of the relevant literature. In Section 3.3 we give a detailed description of the MDP model. In Section 3.4 we present numerical results at both the individual level and the population level. Finally, in Section 3.5 we highlight main conclusions and directions for future work.

3.2 Diabetes Treatment Background and Literature Review

We focus on the prevention of stroke and CHD events since they are the leading causes of death for patients with type 2 diabetes. Most patients with diabetes use medication to manage blood pressure and cholesterol since they are the most significant controllable risk factors for stroke and CHD. Glucose control is also important, particularly for the prevention of microvascular events (such as blindness and nerve damage); however, it has not been shown that tight control

of glucose in individuals with diabetes has significant risk reduction for cardiovascular events [103, 47, 32].

There are many published recommendations in the United States and other countries for initiation of blood pressure and cholesterol medications. Table 3.1 provides a summary of U.S. and international guidelines for initiation of these medications based on well-established risk factors. For comparison, we provide both the current U.S. guideline for diabetes patients that uses the same treatment threshold for all patients (U.S. I) and the current U.S. guideline for patients without diabetes that uses risk-based treatment thresholds (U.S. II). The patients are assigned a risk level (low, medium, or high) based on risk factors such as age and gender. In the United States, initiation of blood pressure and cholesterol medications has been recommended by two independent committees [10, 22]. For diabetes patients these guidelines are "one size fits all"; all diabetes patients are treated to the same threshold, regardless of risk of events, gender, age, or any other factors. The uncoordinated treatment of these risk factors is questionable since blood pressure and cholesterol both affect the overall health of a patient and his or her risk of complications [97, 96].

U.S. and other international guidelines are typically defined by clinical thresholds for stroke and CHD risk factors (other events which are less common such as kidney failure and neuropathy also influence guidelines). The most common risk factors considered by the guidelines are cholesterol and systolic blood pressure (SBP). There are several measures associated with cholesterol including low-density lipoprotein (LDL), high-density lipoprotein (HDL), lipid ratio (LR), and total cholesterol (TC). A patient's TC is a combination of LDL, HDL, and triglycerides, a relationship estimated by the Friedewald equation [40]. A patient's LR is TC divided by HDL. If any of these risk factors are outside of the specified threshold the patient should begin an additional medication for cholesterol or blood pressure treatment, as appropriate.

Risk models that use risk factors as inputs, including the United Kingdom Prospective Diabetes Study (UKPDS) risk engine [97, 96, 61], make it possible to build models in which initiation of medications affects probabilities of complications for patients with diabetes. The

Table 3.1: International guideline thresholds for initiation of cholesterol and blood pressure medications. Guidelines that assume diabetes patients are not considered CHD risk equivalent are represented with *. LDL is measured in mg/dL for the U.S. guidelines, and LDL, HDL, and TC are measured in mmol/L for all other guidelines. LR is unitless, and SBP is measured in mmHg.

Guideline	Cholesterol	Blood Pressure
U.S. I [10, 22]	ATP III: LDL $\geq$ 100	JNC 7: SBP $>$ 130
U.S. II [10, 22]	ATP III*: High Risk: LDL $\geq$ 100, Medium Risk: LDL $\geq$ 130, Low Risk: LDL $\geq$ 190	JNC 7*: SBP $>$ 140
Australia [48]	LDL $\geq$ 2.5 or TC $\geq$ 4.0 or HDL $<$ 1.0	SBP $>$ 130
Canada [15]	LDL $\geq$ 2.5 or LR $\geq$ 4.0	SBP $>$ 130
European Union [46]	LDL $\geq$ 2.5 or TC $\geq$ 4.5	SBP $>$ 130
British [57]	LDL $\geq$ 2.0 or TC $\geq$ 4.0	SBP $>$ 130

UKPDS model is a set of risk equations based on a large cohort of diabetes patients in the United Kingdom; inputs for the risk equations include time since diagnosis of diabetes, age, SBP, LR, and gender. We use the UKPDS model to estimate probabilities of fatal and nonfatal stroke and CHD events in our MDP.

Several models related to our MDP model have been published in the medical decision making and operations research literature. In the context of type 1 diabetes, Parker et al. [74] provide an overview of control algorithms for real-time monitoring and management of blood glucose. Algorithms are presented to determine appropriate insulin delivery. Other models focus on treatment decisions for patients with type 2 diabetes. Denton et al. [30] proposed a non-stationary MDP model to study the optimal timing of statin initiation for cholesterol management, providing optimal control for a single risk factor. Kurt et al. [63] provide structural properties of the optimal statin initiation policy. Timbie et al. [102] simulated initiation of blood pressure and cholesterol medications in order to treat to the U.S. guideline targets. However, this simulation model was only used at one point in the patient's life rather than simulating the use of guidelines over a patient's lifetime. Shah et al. [89] also studied the impact of managing blood pressure and cholesterol according to guidelines for treatment (U.S. and international)

over the course of a patient's lifetime.

MDP models have also been used to determine the optimal timing of one-time medical interventions for a number of diseases other than diabetes. Alagoz et al. [7, 8] provide a discrete-time, infinite-horizon, stationary MDP model to determine the optimal timing of liver transplantation based on a patient's MELD score. They also present structural results, proving sufficient conditions for the existence of a control-limit policy for transplantations. Shechter et al. [90] present an MDP model for the optimal initiation of HIV treatment according to a patient's CD4 count with the goal of maximizing a patient's quality-adjusted lifetime. They assume a stationary, infinite-horizon model and prove that a control-limit policy exists in terms of the patient's CD4 count.

This chapter contributes to the existing literature in two main ways. First, we present a novel model formulation to determine optimal treatment policies for management of a chronic disease over the course of their lifetime. Our model involves the use of multiple medications for simultaneous control of multiple risk factors. To our knowledge, we are the first to model simultaneous control of multiple risk factors. Most related research concentrates on optimal treatment decisions for a single risk factor. Other diabetes models are more descriptive and do not provide dynamic, prescriptive policies over time as our work does. Second, we use our model to answer important policy questions regarding the benefits of coordinating treatment guidelines for cholesterol and blood pressure control. We anticipate our findings will provide insights into the ordering of treatment decisions in other contexts.

3.3 Model

Our MDP model defines the patient's health status (including SBP, TC, and HDL), current medications, number of stroke and CHD events that have occurred in the past, and risk of future stroke and CHD events based on a discrete set of health states. A discrete set of actions represent the initiation of various treatment options. The objective of our model is to find the optimal sequence and timing of medications to manage stroke and CHD risk. Figure 3.1 provides

a simplified state transition diagram of our model for the purpose of illustrating the problem. In the diagram, solid lines illustrate the actions of initiating one or both of the most common medications (statins (ST), ACE inhibitors (AI)), and dashed lines represent the occurrence of an adverse event (stroke or CHD event) or death from other causes. In each medication state, including the no medication state ($\emptyset$), patients probabilistically move between health states, here represented by L (low), M (medium), H (high), and V (very high). These health state levels represent the levels of patient risk factors (e.g., blood pressure and cholesterol). For patients on one or both medications, improvements in patient risk factors (blood pressure, cholesterol, or both) reduce the probability of adverse events.

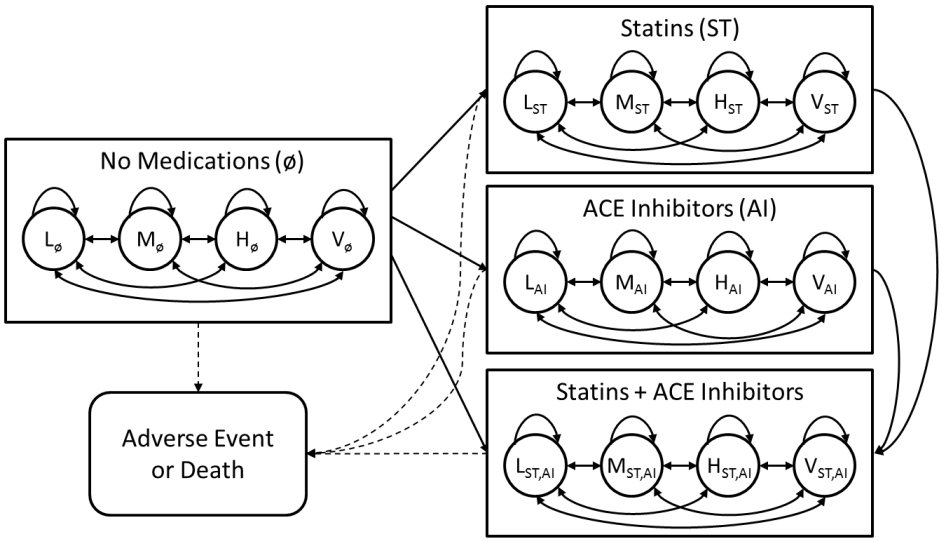

Figure 3.1: Simplified state transition diagram for the case of two medications. When medications are initiated (actions denoted by the solid lines), the risk factors are improved and the probability of the occurrence of an adverse event (denoted by the dashed lines) is reduced.

In our model treatment decisions are assumed to be irreversible. Once a patient begins a blood pressure or cholesterol medicine, it is assumed that he or she remains on the medication for the remainder of his or her life. This is consistent with clinical recommendations for these medications [94, 110, 22]. In some cases, major side effects may cause patients to discontinue

treatment (e.g., statins can cause liver problems or severe muscle pain). However, this occurs in a small proportion of patients.

The problem we explore in this chapter is a generalization of the above two-medication problem, depicted in Figure 3.1, in which the patient may elect to initiate one or more of a set of available treatments at each decision epoch. This optimal treatment problem can be viewed as a nested stopping time problem. After the first medication is initiated (the first stopping time is chosen), there is a subsequent stopping time problem for the next medication to be initiated, and so on. A brief description of the MDP model is presented below.

Actions are taken at a discrete set of decision epochs indexed by $t = 1, \ldots, T$, where epoch t represents the year $[t, t+1)$. This range constitutes the finite decision horizon. Similar to other studies [30, 63, 89], yearly decision epochs are used to represent annual visits to a clinician. Ages above T are represented by an infinite post-decision horizon, assuming no new medications are initiated, allowing for accrual of rewards for patients living past the end of the decision horizon.

States are composed of *living states* and *absorbing states*. Each living state is defined by the factors that influence a patient's cardiovascular risk: the patient's TC, HDL, and SBP levels, medication status, and history of stroke and CHD events. We denote the set of the TC states by $\mathcal{L}_{\text{TC}} = \{L, M, H, V\}$, with similar definitions for HDL, $\mathcal{L}_{\text{HDL}} = \{L, M, H, V\}$, and SBP, $\mathcal{L}_{\text{SBP}} = \{L, M, H, V\}$. The thresholds for these ranges are based on clinically-relevant cut points for treatment found in Table 3.2 [24]. The history of stroke and CHD events is defined by the current number of events the patient has had up to some maximum number, k: $\mathcal{L}_{\text{S}} = \{0, 1, \ldots, k\}$ and $\mathcal{L}_{\text{CHD}} = \{0, 1, \ldots, k\}$. Elements of these sets are indexed by ℓ_{TC}, ℓ_{HDL}, ℓ_{SBP}, ℓ_{S}, and ℓ_{CHD}, respectively. The set of health states is given by $\mathcal{L} = \mathcal{L}_{\text{TC}} \times \mathcal{L}_{\text{HDL}} \times \mathcal{L}_{\text{SBP}} \times \mathcal{L}_{\text{S}} \times \mathcal{L}_{\text{CHD}}$. Elements of $\mathcal{L}$ are indexed by ℓ.

The set of medication states is denoted by $\mathcal{M} = \{\boldsymbol{m} = (m_1, m_2, \ldots, m_n) : m_i \in \{0, 1\}, \forall i = 1, 2, \ldots, n\}$ where n denotes the number of medications. If $m_i = 0$, the patient is not currently on medication i, and if $m_i = 1$, the patient is currently on the medication. When a patient begins treatment i, the medication effects are denoted by $\omega^{\text{TC}}(i)$, representing the proportional

change in TC, $\omega^{\text{HDL}}(i)$, representing the proportional change in HDL, and $\omega^{\text{SBP}}(i)$, representing the proportional change in SBP. Note, in general cholesterol medications result in decreased TC and increased HDL, while blood pressure medications result in decreased SBP. The medications we consider are targeted specifically at either cholesterol or blood pressure and each has negligible effect on the other risk factor. For example, if medication i is a blood pressure medication, then $\omega^{\text{TC}}(i) = \omega^{\text{HDL}}(i) = 0$.

The living states in the model are denoted by $(\ell, \boldsymbol{m}) \in \mathcal{L} \times \mathcal{M}$. The absorbing states are represented by the death states: $\mathcal{D} = \{\mathcal{D}_{\text{S}}, \mathcal{D}_{\text{CHD}}, \mathcal{D}_{\text{O}}\}$. The three types of death states represent dying from a stroke, $\mathcal{D}_{\text{S}}$, a CHD event, $\mathcal{D}_{\text{CHD}}$, or other causes, $\mathcal{D}_{\text{O}}$. The absorbing states will be denoted by $d \in \mathcal{D}$. Including living and absorbing states, there are a total of $4^3 \times 2^n \times (k+1)^2 + 3$ states in our model for each time period.

At each decision epoch, it must be determined which medications to initiate (if any). The action space is dependent on the history of medications that have been initiated in previous epochs. For each medication, at each epoch, medication i can be initiated (I) or initiation can be delayed (W). These actions are defined for medication i as follows:

$$A_{(\ell,m_i)} = \begin{cases} \{I_i, W_i\} & \text{if } m_i = 0, \\ \{W_i\} & \text{if } m_i = 1, \end{cases} \tag{3.1}$$

where $\boldsymbol{A}_{(\ell,\boldsymbol{m})} = \{A_{(\ell,m_1)} \times A_{(\ell,m_2)} \times \cdots \times A_{(\ell,m_n)}\}$. Action $\boldsymbol{a} \in \boldsymbol{A}_{(\ell,\boldsymbol{m})}$ denotes the action taken in state $(\ell, \boldsymbol{m})$. If a patient is in living state $(\ell, \boldsymbol{m})$ and takes action $\boldsymbol{a}$, the medication state is then denoted by $\boldsymbol{m'}$, where m'_i is set to 1 for any medications i that are newly initiated by

Table 3.2: Ranges for TC, HDL, and SBP states based on [24].

	L	M	H	V
TC (mg/dL)	< 160	[160, 200)	[200, 240)	≥ 240
HDL (mg/dL)	< 40	[40, 50)	[50, 60)	≥ 60
SBP (mmHg)	< 120	[120, 140)	[140, 160)	≥ 160

action $\boldsymbol{a}$; $m_i' = m_i$ for all medications i which are not newly initiated. Once medication i is initiated, the patient's blood pressure and cholesterol are modified by the medication effects denoted by $\omega^{\mathrm{TC}}(i)$, $\omega^{\mathrm{HDL}}(i)$, and $\omega^{\mathrm{SBP}}(i)$, resulting in a reduction in the probability of having a stroke or CHD event.

Three types of probabilities are incorporated into the model: probabilities among health states, probability of events (both fatal and nonfatal), and probability of death from other causes. At epoch $t \in 1, \ldots, T$, death from other causes occurs with probability π_t^{O}. If the patient is in state $(\ell, \boldsymbol{m}) \in \mathcal{L} \times \mathcal{M}$, a nonfatal stroke or CHD event occurs with probability $\pi_t^{\mathrm{S}}(\ell, \boldsymbol{m})$ and $\pi_t^{\mathrm{CHD}}(\ell, \boldsymbol{m})$, respectively, which depend on the patient's age, health state, medication status, and other risk factors such as race and gender. Fatal stroke and CHD events occur with probability $\widetilde{\pi}_t^{\mathrm{S}}(\ell, \boldsymbol{m})$ and $\widetilde{\pi}_t^{\mathrm{CHD}}(\ell, \boldsymbol{m})$, respectively. Given that the patient is in state $(\ell, \boldsymbol{m})$ at epoch t, the probability of moving into one of the absorbing states $d \in \mathcal{D}$ at epoch $t+1$ is denoted by $\bar{p}_t^{\boldsymbol{m}}(d|\ell)$, where

$$\bar{p}_t^{\boldsymbol{m}}(d|\ell) = \begin{cases} \pi_t^{\mathrm{O}} & \text{if } d = \mathcal{D}_{\mathrm{O}}, \\ \widetilde{\pi}_t^{\mathrm{CHD}}(\ell, \boldsymbol{m}) & \text{if } d = \mathcal{D}_{\mathrm{CHD}}, \\ \widetilde{\pi}_t^{\mathrm{S}}(\ell, \boldsymbol{m}) & \text{if } d = \mathcal{D}_{\mathrm{S}}, \end{cases} \tag{3.2}$$

for $(\ell, \boldsymbol{m}) \in \mathcal{L} \times \mathcal{M}$, and $\bar{p}_t^{\boldsymbol{m}}(d|d) = 1$ for all $t \in 1, \ldots, T$. The probability of having a nonfatal event or dying (from an event or other causes) is denoted by $\pi_t^*(\ell, \boldsymbol{m})$, where

$$\begin{aligned} \pi_t^*(\ell, \boldsymbol{m}) = {} & (1 - \pi_t^{\mathrm{CHD}}(\ell, \boldsymbol{m}) - \widetilde{\pi}_t^{\mathrm{CHD}}(\ell, \boldsymbol{m}))\pi_t^{\mathrm{S}}(\ell, \boldsymbol{m}) \\ & + (1 - \pi_t^{\mathrm{S}}(\ell, \boldsymbol{m}) - \widetilde{\pi}_t^{\mathrm{S}}(\ell, \boldsymbol{m}))\pi_t^{\mathrm{CHD}}(\ell, \boldsymbol{m}) \\ & + \widetilde{\pi}_t^{\mathrm{S}}(\ell, \boldsymbol{m}) + \widetilde{\pi}_t^{\mathrm{CHD}}(\ell, \boldsymbol{m}) + \pi_t^{\mathrm{O}}. \end{aligned} \tag{3.3}$$

This equation defining the probability of having an event or death is based on the assumption that the probabilities of strokes and CHD events are independent. This equation could easily

be altered for events that are not assumed to be independent. Given that the patient is in health state $\ell \in \mathcal{L}$, the probability of being in health state ℓ' in the next epoch following is denoted by $q_t(\ell'|\ell)$. The transition probabilities between health states do not depend on the medication state since the transition probabilities $q_t(\ell'|\ell)$ are computed from the natural progression of blood pressure and cholesterol in the absence of medication. We define $p_t^{\boldsymbol{m}}(j|\ell)$ to be the probability of a patient being in state $j \in \mathcal{L} \cup \mathcal{D}$ at epoch $t+1$, given the patient is in living state $(\ell, \boldsymbol{m})$ at epoch t, where $\boldsymbol{m}$ incorporates the action $\boldsymbol{a}$ taken at time t. The probability $p_t^{\boldsymbol{m}}(j|\ell)$ is defined by the following:

$$p_t^{\boldsymbol{m}}(j|\ell) = \begin{cases} \big[1 - \sum_{d \in \mathcal{D}} \bar{p}_t^{\boldsymbol{m}}(d|\ell)\big] q_t(j|\ell) & \text{if } \ell, j \in \mathcal{L}, \\ \bar{p}_t^{\boldsymbol{m}}(j|\ell) & \text{if } \ell \in \mathcal{L},\ j = \mathcal{D}, \\ 1 & \text{if } \ell = j \in \mathcal{D}, \\ 0 & \text{otherwise.} \end{cases} \tag{3.4}$$

The reward $r_t(\ell, \boldsymbol{m})$ is the dollar reward for QALYs minus treatment and medication costs accrued in decision epoch t in living state $(\ell, \boldsymbol{m})$ as described in the following equation:

$$r_t(\ell, \boldsymbol{m}) = R(\ell, \boldsymbol{m}) - C^{\mathrm{O}} - \big(C^{\mathrm{S}}(\ell) + C^{\mathrm{CHD}}(\ell)\big) - \big(\mathrm{CF}^{\mathrm{S}}(\ell) + \mathrm{CF}^{\mathrm{CHD}}(\ell)\big) - C^{\mathrm{MED}}(\boldsymbol{m}), \tag{3.5}$$

for $t = 1, \ldots, T$, where $R(\ell, \boldsymbol{m}) = R_0(1 - d^{\mathrm{S}}(\ell))(1 - d^{\mathrm{CHD}}(\ell))(1 - d^{\mathrm{MED}}(\boldsymbol{m}))$ is the reward for one QALY. The quantity R_0 is the reward per QALY, which is analogous to the willingness-to-pay threshold in health economics studies. When a patient has an event or initiates statins, her quality of life is decreased. The decrement factors $d^{\mathrm{S}}(\ell)$, $d^{\mathrm{CHD}}(\ell)$, and $d^{\mathrm{MED}}(\boldsymbol{m})$ represent the decrease in quality of life from a stroke, a CHD event, or medication initiation, respectively. The costs C^{O}, $C^{\mathrm{MED}}(\boldsymbol{m})$, $C^{\mathrm{S}}(\ell)$ and $C^{\mathrm{CHD}}(\ell)$, and $\mathrm{CF}^{\mathrm{S}}(\ell)$ and $\mathrm{CF}^{\mathrm{CHD}}(\ell)$ represent cost of other

health care for diabetes patients, cost of medications, cost of initial hospitalization for stroke and CHD events, and cost of follow-up treatment for stroke and CHD events, respectively.

For a patient in living state $(\ell, \boldsymbol{m})$ in epoch t, let $v_t(\ell, \boldsymbol{m})$ denote the patient's maximum total expected discounted rewards prior to her first event or death. The following recursion defines the optimal action in each state for $t = 1, \ldots, T-1$:

$$v_t(\ell, \boldsymbol{m}) = \max_{\boldsymbol{a} \in \boldsymbol{A}_{(\ell, \boldsymbol{m})}} \left\{ r_t(\ell, \boldsymbol{m}'(\boldsymbol{a})) + \lambda \sum_{\forall j \in \mathcal{L} \cup \mathcal{D}} p_t^{\boldsymbol{m}'(\boldsymbol{a})}(j|\ell) v_{t+1}(j, \boldsymbol{m}'(\boldsymbol{a})) \right\}, \tag{3.6}$$

where j indexes states in $\mathcal{L} \cup \mathcal{D}$, $\boldsymbol{m}'(\boldsymbol{a})$ is defined as the medication state $\boldsymbol{m}$ with action $\boldsymbol{a}$ taken into account, and $\lambda \in [0, 1)$ is the discount factor per decision epoch, which is commonly set to 97% in health economic evaluations (see Chapter 7 of [41] for a discussion of this). The boundary condition is given by $v_T(\ell, \boldsymbol{m}) = r_T(\ell, \boldsymbol{m}) + E\left[\text{PDHR}|\ell, \boldsymbol{m}\right]$, where $E\left[\text{PDHR}|\ell, \boldsymbol{m}\right]$ is the expected post–decision horizon reward (PDHR). This represents expected rewards for a patient living past the decision horizon (e.g., past age 100). The PDHR depends on the state and treatment status of the patient in the last year of the decision horizon and the number of years into the post–decision horizon that the patient lives. This approximation of rewards is needed because of the limited samples in the data set for older patients.

Various criteria can be expressed using the general reward function in Equation (3.5). For example, letting $R(\ell, \boldsymbol{m}) = 1$ for all patients with $\ell_{\text{S}} = \ell_{\text{CHD}} = 0$ and all cost parameters set to zero represents the objective of primary prevention (i.e., the value function for a given state at a given decision epoch is the expected LYs to first event). Another primary prevention model criterion that can be expressed using the reward function is to let $R(\ell, \boldsymbol{m}) = 1 - d^{\text{MED}}(\boldsymbol{m})$ for patients with $\ell_{\text{S}} = \ell_{\text{CHD}} = 0$. This primary prevention model takes into account the downside of treatment by incorporating the disutility of treatment before a patient's first event. A third primary prevention model would incorporate both the disutility and cost of medications before the patient has any events. Thus, $R(\ell, \boldsymbol{m}) = R_0$ and all costs other than $C^{\text{MED}}(\boldsymbol{m})$ are set to zero. In all of these cases, patients with $\ell_{\text{S}} \neq 0$ or $\ell_{\text{CHD}} \neq 0$ have $R(\ell, \boldsymbol{m}) = 0$. As we discuss in Section 3.4, inclusion or removal of different parts of the general reward function (e.g., costs

of other medical care for diabetes patients) can greatly affect optimal policies.

3.4 Results

In this section we present numerical results illustrating optimal treatment policies for two bicriteria perspectives: (a) expected time to first event versus medication costs, and (b) expected QALYs versus medication and treatment costs. Backward induction was used to compute the optimal treatment decisions over the patient's lifetime. The model and solution method was coded in C/C++. Model instances were solved in under 40 minutes using a 2.83GHz PC with 8GB of RAM. We provide results for each perspective for a population of 40-year-old patients newly diagnosed with type 2 diabetes. The proportion of patients in each of the health states at age 40 is estimated using the Mayo cohort described in Section 3.4.1.

The remainder of this section is organized as follows: In Section 3.4.1 we define the specific parameters for our problem, the model inputs, and their sources. In Section 3.4.2 we present a comparison of outputs from our model to those found in the literature for validation purposes. In Section 3.4.3 we present a model for primary prevention with results for maximization of LYs before an event. In Section 3.4.4, we provide results from the population level for maximizing QALYs over the patient's lifetime (i.e., average results for patients with diabetes). For the results presented Sections 3.4.3 and 3.4.4, we compare the optimal treatment outcomes to the outcomes from applying U.S. and international guidelines. We also highlight the main differences in the policies for individual patients. In Section 3.4.5 we provide estimates of the yearly benefit of applying the optimal guidelines to the U.S. diabetes population over the current U.S. guidelines.

3.4.1 Data and Study Population

We used an observational data set based on medical records from the Mayo Electronic Medical Records (Mayo EMR) and Diabetes Electronic Management System (DEMS) for a large cohort of patients receiving treatment for type 2 diabetes at Mayo Clinic, Rochester, MN [44].

The DEMS dataset included 663 patients with cholesterol, HbA1c, blood pressure, and other laboratory values. Population statistics are provided in Table 3.3. The patients in this dataset are hereafter referred to as the Mayo cohort. Changes in TC, HDL, and SBP values from medications are found in Table 3.4. These values were estimated by computing the change in metabolic values before and after initiation of the given treatments using methods reported in Denton et al. [30]. These changes are assumed to be independent and additive for patients on multiple medications. It is important to note that while fibrates have been shown to improve cholesterol values, there is debate if the use of fibrates actually reduces a patient's risk of CHD events [43]. It is possibly a limitation that we use modification of surrogate markers (blood pressure and cholesterol) to reflect the benefits of medication rather than modified risk of events (stroke and CHD).

Table 3.3: Baseline characteristics for the study population (N = 663), including mean and variance.

Patient Attribute	Study Cohort
Age	52.46 (8.83)
Years with Diabetes	3.24 (5.33)
% Female	39.67
HDL	43.65 (11.58)
LDL	126.98 (37.31)
TC	216.98 (37.31)
SBP	139.11 (19.75)
HbA1c	8.01 (2.38)

The descriptions and values of utility and cost parameters can be found in Table 3.5. All costs given are for one year. In our numerical experiments we chose $R_0 = \$100,000$ as the base-case willingness-to-pay factor since this is the most commonly used value in U.S. studies [82]; however, we will use a range of willingness-to-pay values for R_0 for the population results. Costs and utility decrements for each cholesterol and blood pressure medication are found in

Table 3.4: Percentage change in risk factors for given medications as computed from Mayo Electronic Medical Records and Diabetes Electronic Management System.

Medication (i)	$\omega^{\mathrm{TC}}(i)$	$\omega^{\mathrm{HDL}}(i)$	$\omega^{\mathrm{SBP}}(i)$
Statins	-14.0	+7.3	-
Fibrates	-3.9	+4.7	-
ACE/ARBs	-	-	-3.7
Thiazides	-	-	-5.0
β Blockers	-	-	-4.6
Calcium Channel Blockers	-	-	-2.5

Table 3.6. The costs are the lower bound values based on U.S. pharmaceutical cost estimates [4], and the utility decrements are drawn from the literature [20, 77]. The costs presented in Tables 3.5 and 3.6 are in 2009 dollars. For all the numerical experiments, we consider a decision horizon from age 40 to age 100 with an infinite horizon estimate of rewards accrued after the end of the decision horizon.

Table 3.5: Description of model parameters including cost inputs and utility decrements for the reward function of the MDP model.

Parameter Type	Parameter	Value	Source
Cost Inputs	Initial hospitalization for stroke (C^{S})	\$13,204	[1]
	Initial hospitalization for CHD (C^{CHD})	\$18,590	[1]
	Follow-up for stroke (CF^{S})	\$1,664	[101]
	Follow-up for CHD ($\mathrm{CF}^{\mathrm{CHD}}$)	\$2,576	[86, 101]
	Willingness-to-pay Factor (R_0)	\$100,000	[82]
	Discount Factor (λ)	0.97	[41]
Utility Inputs	CHD decrement (d^{CHD})	0.07	[23, 106]
	Stroke decrement (d^{S})	0.21	[23, 100, 99]

The transition probabilities among health states were also computed from the DEMS

Table 3.6: Costs and utility decrements for each medication used in the model.

Medication	Cost [4]	Utility Decrement
Statins	\$212	0.003 [20]
Fibrates	\$652	0.003 [20]
ACE/ARBs	\$48	0.005 [77]
Thiazides	\$48	0.005 [77]
β Blockers	\$48	0.005 [77]
Calcium Channel Blockers	\$866	0.005 [77]

dataset. A spline fit was used to interpolate missing laboratory values for cholesterol values to obtain an estimate of yearly levels for these risk factors [30]. Each risk factor was divided into L, M, H, and V categories (as defined in Section 3.3). The transition probabilities among metabolic states were estimated from the percentages of patients that moved between each state at time t to each state at time $t+1$.

Transition probabilities to event and death states were drawn from the literature. The UKPDS risk equations [61, 97, 96] were used to compute probabilities of incurring a CHD event or stroke, both fatal and nonfatal, based on patient risk factors including age, gender, TC, SBP, and HbA1c. The Centers for Disease Control and Prevention (CDC) mortality tables [18] were used to estimate the probability of death from other causes.

3.4.2 Model Validation

In order to validate the results from our MDP, we compared outputs from our model to life expectancy estimates from the Framingham Heart Study (FHS) for diabetes patients with and without cardiovascular disease [39]. The FHS recruited 5209 study participants living in Framingham, Massachusetts between 1948 and 1951. Franco et al. [39] chose three separate 12-year follow-up periods to include in their study. The beginning years of the three periods were 1956, 1969, and 1985. A total of 9033 patient observation periods were included once patients with incomplete data or patients having cardiovascular disease were removed. The two

estimates compared from the FHS and our model were the expected LYs before a stroke or CHD event from age 50 and the expected LYs before death from age 50.

The FHS reported three sets of estimates for LYs before an event, LYs after an event, and total LYs from age 50: no diabetes, diabetes, and overall (including patients with and without diabetes). Our MDP assumes patients have diabetes, so the estimates for patients with diabetes at age 50 should provide a good lower bound for the estimates from the MDP model. Since patients without diabetes at age 50 may develop diabetes in the future, the overall estimates should provide an upper bound on the estimates for patients with diabetes at age 50. Tables 3.7 and 3.8 present the comparison of expected LYs before death, expected LYs before a stroke or CHD event, and expected LYs after an event from age 50 for our MDP model and the FHS. We present MDP results for two policies: the current U.S. guidelines (U.S. I) and no treatment. Although no data is available on the actual treatment policy for patients in the FHS, it is likely that most patients would have followed a policy that would result in expected LYs between the estimates for the two policies given. As expected, the LYs before an event from the MDP model fall between the FHS diabetes estimate and the FHS overall estimate for males and females. This is also true for female life expectancy estimates; however, the life expectancy from age 50 estimate for males using U.S. I falls outside of the corresponding FHS range. It is important to note that the additional LYs after an event for the FHS diabetes estimate is very close to the estimate for additional LYs for the MDP results. For males and females, the estimates for the additional LYs after an event from the MDP are within the 95% confidence intervals for the FHS diabetes patients.

Table 3.7: Male comparison of expected LYs before death, expected LYs before a stroke or CHD event, and expected LYs after an event from age 50 for our MDP model and the Framingham Heart Study (FHS). The 95% confidence intervals are provided for the FHS estimates.

	FHS: Diabetes Patients	FHS: Overall	MDP: U.S. I	MDP: No Treatment
Life expectancy	21.3 (19.4 to 23.1)	27.9 (27.3 to 28.6)	28.8	26.9
LYs before event	14.2 (12.3 to 16.1)	21.2 (20.5 to 22.0)	21.2	18.9
LYs after event	7.1 (6.0 to 8.3)	6.7 (6.2 to 7.1)	7.6	8.0

Table 3.8: Female comparison of expected LYs before death, expected LYs before a stroke or CHD event, and expected LYs after an event from age 50 for our MDP model and the Framingham Heart Study (FHS). The 95% confidence intervals are provided for the FHS estimates.

	FHS: Diabetes Patients	FHS: Overall	MDP: U.S. I	MDP: No Treatment
Life expectancy	26.5 (24.4 to 28.5)	33.8 (33.2 to 34.4)	32.1	29.4
LYs before event	19.6 (17.5 to 21.9)	27.3 (26.7 to 28.0)	25.3	23.1
LYs after event	6.8 (5.5 to 8.2)	6.4 (6.0 to 6.9)	6.8	6.3

Unfortunately there is no perfect way to validate medical decision making models such as ours. There are many possible reasons why our estimates for event-free years and life expectancy would differ from the estimates from the FHS. First, our model uses 2007 estimates of probabilities of death from other causes; life expectancies have increased significantly since the 1950s when the FHS study began. Second, we use the UKPDS risk equations to estimate the risk of stroke and CHD events; while these equations are widely believed to provide valid estimates of risk, they are based on observed events from a population of diabetes patients from the United Kingdom. Finally, it is impossible to know what medication policy was used by the patients in the FHS, and available medications and U.S. treatment guidelines have changed significantly since the 1950s.

3.4.3 Primary Prevention Treatment Policies

In this section we consider primary prevention of stroke and CHD events. The yearly rewards for primary prevention are defined as follows:

$$r(\ell, \boldsymbol{m}) = \begin{cases} R_0 - C^{\mathrm{O}} - C^{\mathrm{MED}}(\boldsymbol{m}) & \forall \ell : \ell_{\mathrm{S}} = \ell_{\mathrm{CHD}} = 0, \\ 0 & \text{otherwise.} \end{cases} \tag{3.7}$$

Patients receive rewards for LYs (R_0) minus costs of medication and other healthcare costs up until the occurrence of a first event including CHD or stroke (fatal or nonfatal) or death from other causes. We set all costs other than the other costs and medication costs equal to zero: $C^{\mathrm{S}}(\ell) = C^{\mathrm{CHD}}(\ell) = \mathrm{CF}^{\mathrm{S}}(\ell) = \mathrm{CF}^{\mathrm{CHD}}(\ell) = 0$. In addition, no costs are incurred in this model after a patient has an event. With this reward structure, the objective is to maximize the reward for LYs minus costs incurred prior to an event or death. In other words, the goal of this reward structure is to delay the patient's first event. This goal is in line with a physician's primary prevention goal to delay the time until a patient's first event. This goal is consistent with the U.S. guideline's goal of primary prevention [10, 22].

Figures 3.2 and 3.3 compare the primary prevention optimal treatment policy and published

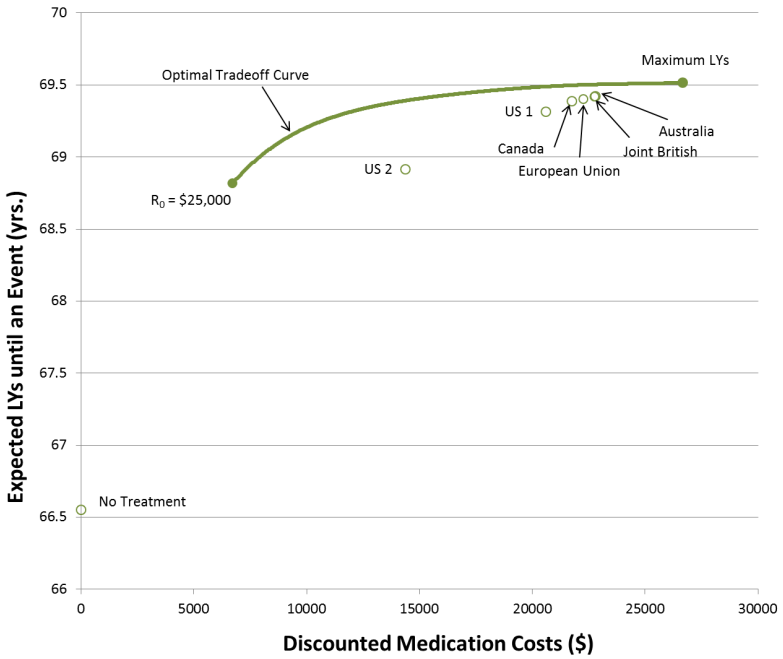

Figure 3.2: Comparison of optimal treatment policies for male patients to treatment by U.S. and international guidelines.

guideline results for males and females, respectively. These graphs present the expected LYs versus the expected discounted costs of medication before an event has occurred (the other costs have not been included in the graph). There is great similarity in costs and LYs between the U.S. and international guideline results. While the LYs achieved with the guidelines are very near the optimal policy curves for both the males and females, we see that the costs of the guidelines could be greatly reduced by implementing the optimal policies due to the flatness of the optimal policy curve as the LYs increase.

The main difference between the male and female results for LYs versus medication costs is seen on the vertical axis. According to our results, on average females can expect approximately five additional years before an event than males. Furthermore, relative to no treatment, males can increase their expected time to an event by as much as 2.97 years, while females can only increase their expected time to first event by up to 2.66 years from no treatment by following

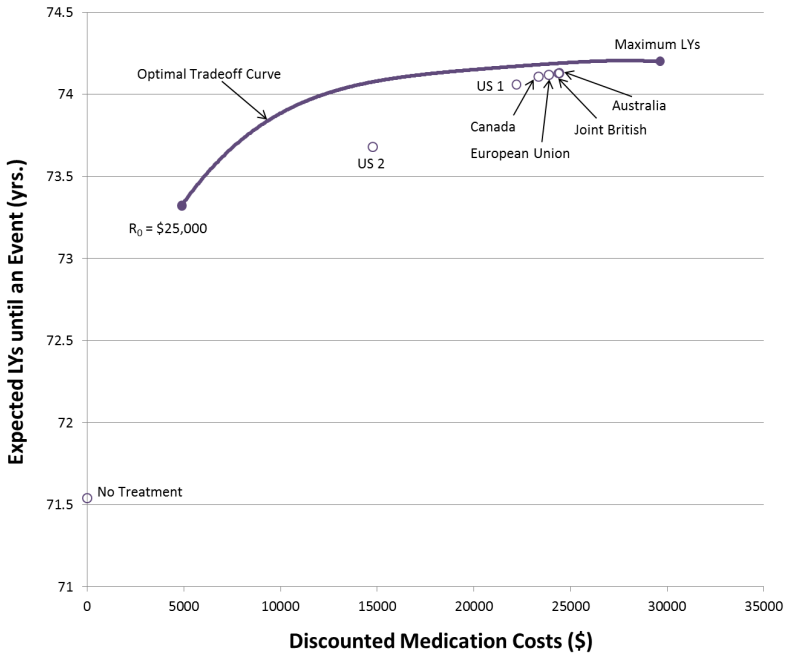

Figure 3.3: Comparison of optimal treatment policies for female patients to treatment by U.S. and other international guidelines.

the optimal guidelines. In addition, following U.S. I provides males an additional 2.76 event-free LYs over no treatment while females can only achieve an additional 2.52 event-free LYs. Thus, we see that male patients can receive a greater benefit than females from medication as seen from the primary prevention perspective. In both cases, as the value of R_0 increases, the optimal guidelines result in increased event-free LYs over U.S. I for all male and female patients.

The results presented in Figures 3.2 and 3.3 are the expected LYs and medication costs for the average male and female patient, respectively. However, not every patient receives the same benefit from treatment. Figures 3.4 and 3.5 provide histograms for the difference in event-free LYs and medication costs between the optimal guidelines, for the R_0 = \$100,000 case, and U.S. I. While the expected cost for all males is lower for the optimal guideline than U.S. I, only 55% of the patients would increase event-free LYs (by up to 0.398 LYs) using the optimal

guideline. For the other 45% of the patients, LYs could be decreased by as much as 0.084 LYs. As with male patients, for female patients the optimal guidelines result in lower expected cost. However, only 27% of females would increase event-free LYs (by up to 0.202 LYs) using optimal guidelines over U.S. I. The remaining 73% of patients would see a decrease in LYs with the use of the optimal guidelines. The reduction in LYs for these patients could be as much as 0.101 LYs.

For more insight into how the policies differ, we provide general descriptions of the policies for the optimal guidelines (R_0 = \$100,000) and U.S. I. According to the optimal guidelines (R_0 = \$100,000), all patients start statins first. The second line treatment is fibrates or thiazides depending on the health status of the patient. The third line treatment is another blood pressure medication (thiazides for those that started fibrates second and beta blockers for those that started thiazides second). The optimal time of initiation of the first four medications varies according to health status of the patient, but they are started by the time the patient is 45, with the fifth medication being started as late as age 63. In all cases, the final medication to be added for all patients is calcium channel blockers. It is started very late in life (age 70 or later) or not at all, depending on the patient's gender and health status. When applying U.S. I, the policies range from starting all six available medications by age 43 for patients in the sickest health state to starting no medications for patients in the healthiest state. For patients that do start medications, the order is to initiate statins and thiazides, then fibrates and ACE inhibitors/ARBs, followed by β blockers, and finally calcium channel blockers. These descriptions are true for both males and females.

3.4.4 Primary and Secondary Prevention Treatment Policies

In this section we consider primary and secondary prevention: QALYs are accrued before and after the first stroke and CHD event. We estimate expected QALYs and expected discounted costs over each patient's lifetime using the reward function defined in Equation (3.5). This provides another perspective on the effectiveness of the current guidelines to the optimal guide-

lines.

Figure 3.6 presents the optimal trade-off curve of QALYs versus costs of medication and hospitalization for events for all male patients by varying the reward for QALYs, R_0. Each point on the optimal trade-off curve, the no treatment point, and each of the points for the current guidelines represent the expected QALYs and expected costs for the average male. The curve represents the outcomes from optimal treatment with $R_0 = \$25{,}000$ to optimal treatment with very large values for R_0 for which QALYs are maximized ($R_0 > \$1{,}000{,}000$). We compare this curve to the U.S. and other international guidelines and no treatment. All of the guidelines besides U.S. II — the current guideline for patients without diabetes and the previous guideline for patients with diabetes — are similar with respect to expected QALYs and costs. U.S. II does not result in QALYs that are as high, but costs are lower. This suggests that treatment based on risk levels and patient risk factors could reduce expected costs with small reduction in expected QALYs.

We also see from the Figure 3.6 that the same expected QALYs can be achieved with the optimal policies as U.S. I and the international guidelines when a very high R_0 value is used. This given quality adjusted lifespan can be achieved with the optimal treatment policies at a much lower expected discounted cost of medication and hospitalization for events than the guidelines. We see a savings of at least \$4,573 (the difference in costs for maximum QALYs and U.S. I) per male patient on average with the optimal policies compared with the U.S. and international guidelines.

Figure 3.7 presents the same optimal trade-off curve for female patients by varying R_0. We again observe the similarity of U.S. I and the international guidelines. For female patients, the increase in QALYs for using U.S. I over the previous U.S. guidelines (U.S. II) costs \$60,221/QALY; in comparison to the cost of the male increase in QALYs of \$49,349/QALY. These \$/QALY estimates are found by calculating the difference in costs over the difference in QALYs between the two guidelines considered. This is likely due to the fact that U.S. II takes into account gender, risk of events, and age. Also, the risk of stroke and CHD events is

in general lower for females than for males.

We also see from Figure 3.7 that we can simultaneously improve expected QALYs and reduce costs for female patients with the optimal treatment policies over the U.S. and international guidelines. This is likely due to the fact that the guidelines treat male and female patients the same, not adjusting for differences in gender when determining medication initiation decisions. The optimal policies take into account the lower probabilities of events for females earlier in life, resulting in initiation of fewer medications, higher expected QALYs, and lower expected discounted costs of treatment. According to our model, with the optimal treatment policies, a savings of at least \$7,378 coupled with an increase of 0.072 QALYs per female patient could be realized compared to the U.S. and international guidelines.

There are a few notable differences between the QALY results for males and females. Females can expect 3.48 QALYs more than males with no treatment. Males can expect to improve their QALYs with optimal treatment by as much as 1.90 QALYs while females can improve by up to 2.43 QALYs. As a final observation, we see that optimal treatment policies would have more effect on improving QALY and cost outcomes over U.S. I for females than for males.

As with the primary prevention results of Section 3.4.3, we provide histograms for the difference in QALYs and costs for males and females. Figure 3.8 shows the difference in QALYs and costs between the optimal guidelines ($R_0 = \$100,000$) and U.S. I for males. Figure 3.9 provides the same histograms for females. All males and females simultaneously have lower QALYs and lower costs using the optimal guidelines ($R_0 = \$100{,}000$) over U.S. I. As R_0 is increased, costs for the optimal guidelines remain lower than cost for U.S. I for males and females but QALYs are higher for all females and nearly all males.

While the U.S. I polices for males and females are the same as described in Section 3.4.3 since they do not depend on the reward function, the optimal guidelines are quite different. For the base case of $R_0 = \$100{,}000$, patients begin medications later in life than with the primary prevention objective guidelines. All males and females should begin statins as the first treatment. According to the optimal policies, intensification should occur in the following

order: thiazides, β blockers, ACE inhibitors/ARBs, fibrates, and calcium channel blockers. The optimal timings of the initial treatment and intensification depend on the patient's health status and gender. For example, the time to start statins ranges from age 40 to age 66, with the start time being most affected by health status. Patients begin the last medication as late as age 98.

Multiple objective functions and their associated reward functions lead to different policies for optimal treatment. When comparing the value of these policies to the rewards associated with the current guidelines, certain policies may be more similar to the current guidelines than others. For example, we see that the current U.S. and international guidelines perform differently with respect to the primary prevention optimal trade-off curves in Figures 3.2 and 3.3 and the primary and secondary prevention optimal trade-off curves in Figures 3.6 and 3.7.

3.4.5 Estimated Benefit of the Optimal Guidelines to the U.S. Diabetes Population

We have presented the benefits of the optimal guidelines over the current U.S. guidelines for the average 40-year-old patient. In this section we present an estimate of the potential benefits of applying the optimal guidelines to all diabetes patients in the United States. We first estimated the prevalence of type 2 diabetes in the United States using population estimates from the 2010 U.S. census [107] and estimated diabetes prevalence by state and age [27]. We then estimated the number of newly-diagnosed diabetes patients each year, defined as patients that have diabetes at age 40 or patients diagnosed with diabetes later in life. At each age, the number of patients considered to be diagnosed past age 40 was the greater of zero and the number of total patients diagnosed at earlier ages minus the diagnosed population at the given age.

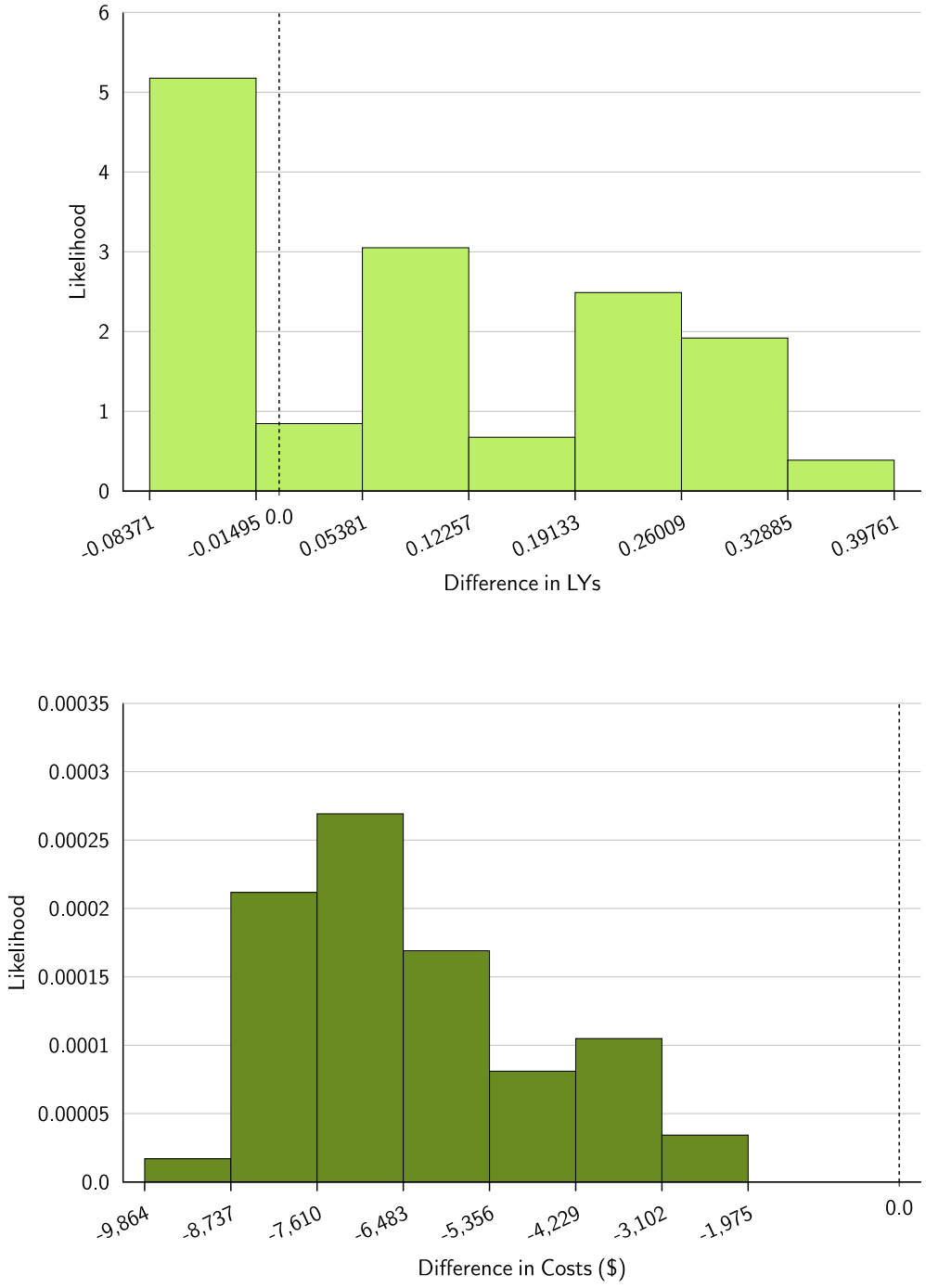

Figure 3.4: Histograms to provide the difference in LYs and medication costs for males between the optimal guidelines ($R_0 = \$100{,}000$) and U.S. I.

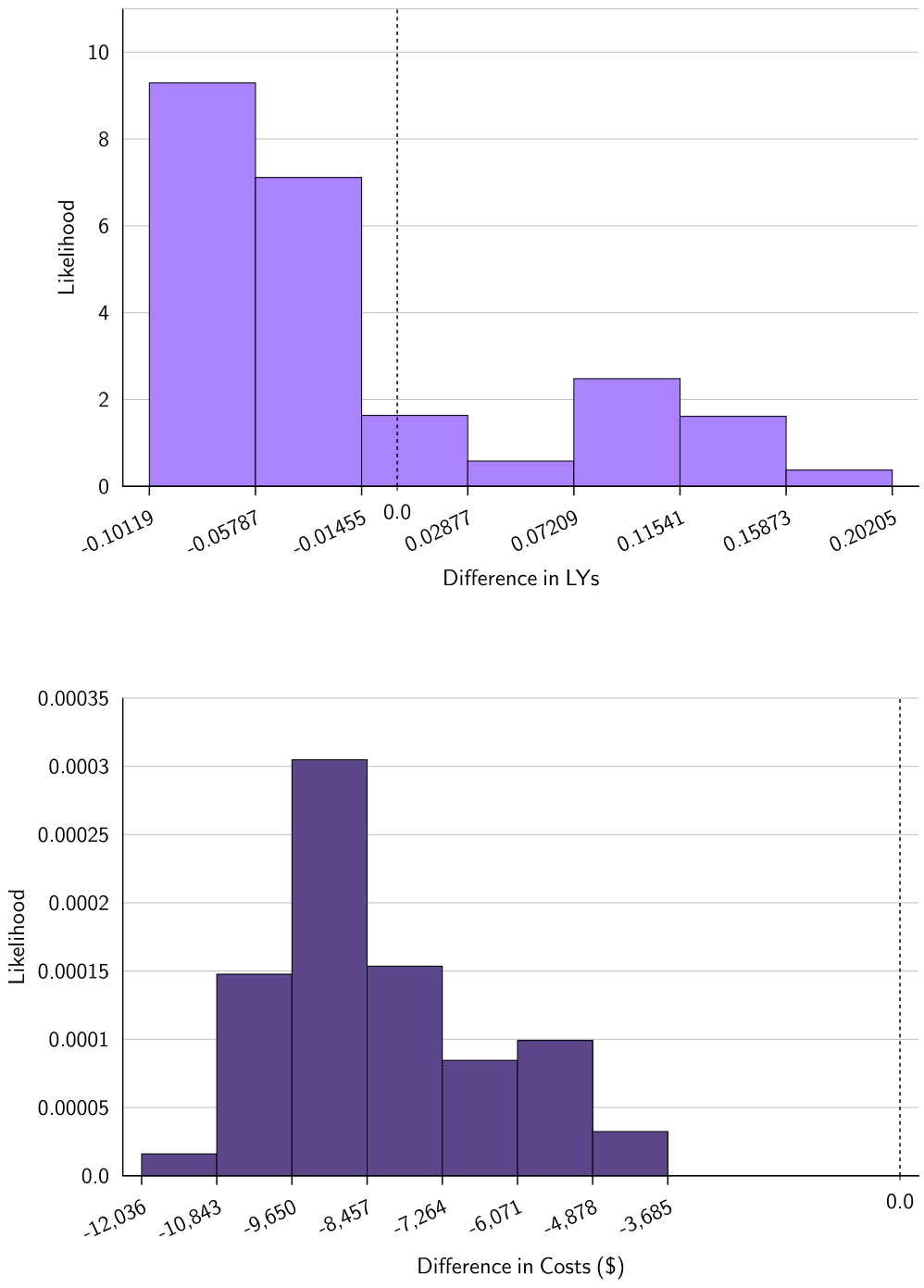

Figure 3.5: Histograms to provide the difference in LYs and medication costs for females between the optimal guidelines (R_0 = \$100,000) and U.S. I.

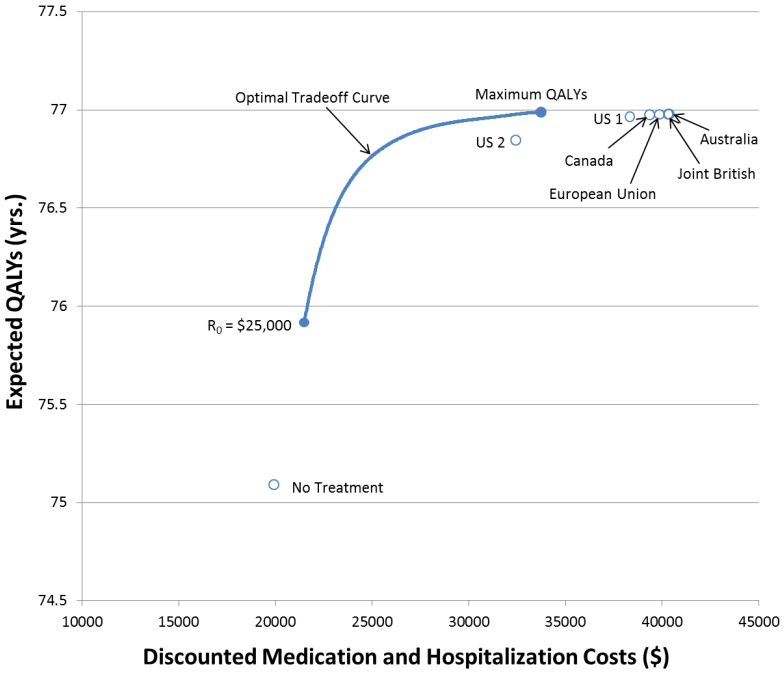

Figure 3.6: Comparison of optimal treatment policies for male patients to treatment by U.S. and international guidelines.

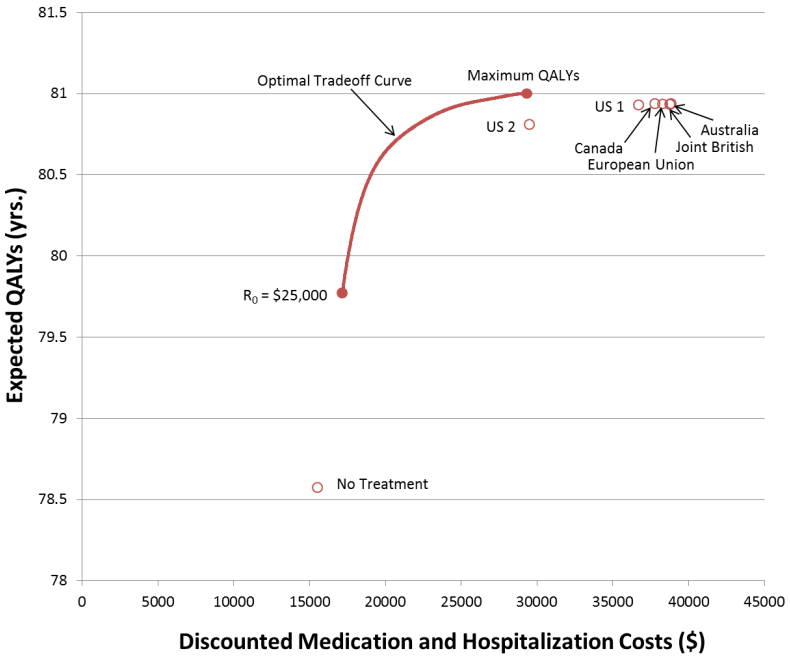

Figure 3.7: Comparison of optimal treatment policies for female patients to treatment by U.S. and international guidelines.

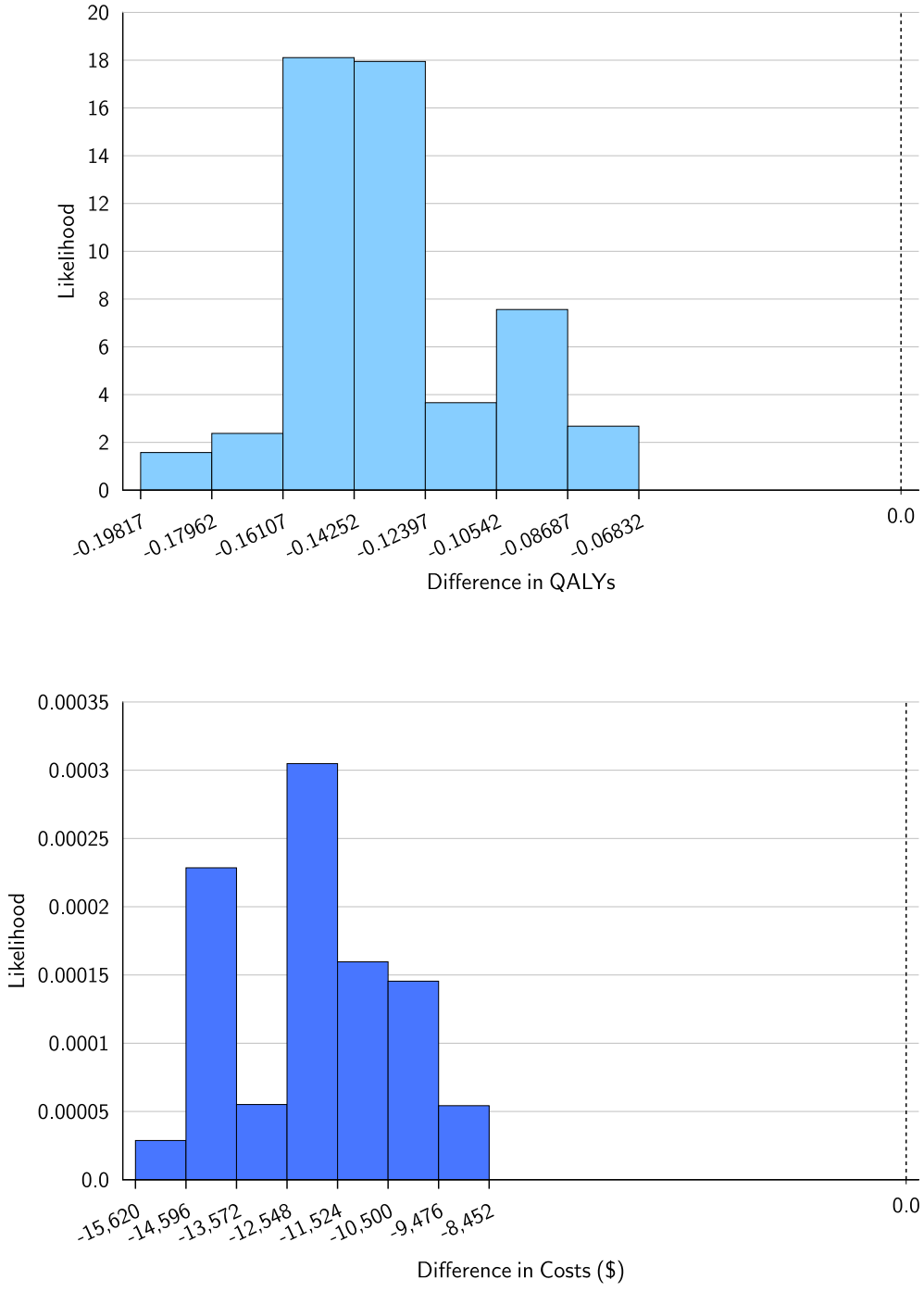

Figure 3.8: Histograms to provide the difference in QALYs and medication and treatment costs for males between the optimal guidelines (R_0 = \$100,000) and U.S. I.

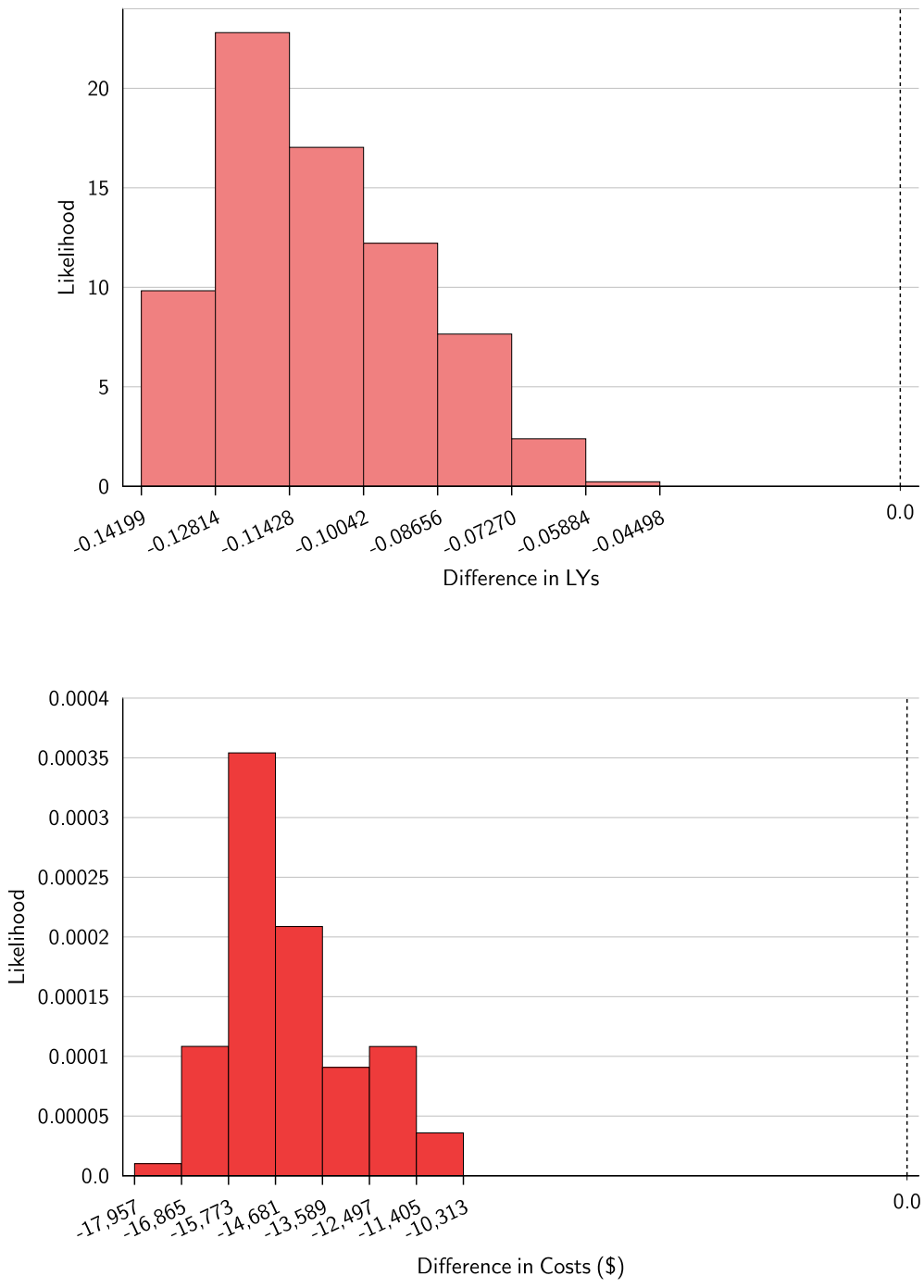

Figure 3.9: Histograms to provide the difference in QALYs and medication and treatment costs for females between the optimal guidelines ($R_0 = \$100{,}000$) and U.S. I.

Table 3.9: Yearly costs (billions) and future event-free LYs for newly diagnosed diabetes patients using no treatment, optimal guidelines (R_0 = \$100,000, R_0 = \$250,000, and R_0 = \$10 million), and U.S. I.

	Males		Females		Total Population	
	LYs	Cost (billions)	LYs	Cost (billions)	LYs	Cost (billions)
No Treatment	7,935,018	\$0	10,215,478	\$0	18,150,496	\$0
Optimal (R_0 = \$100,000)	8,904,509	\$4.71	11,168,982	\$5.22	20,073,492	\$9.93
Optimal (R_0 = \$250,000)	8,939,740	\$7.01	11,211,694	\$7.98	20,151,434	\$15.00
Optimal (R_0 = \$10 million)	8,947,470	\$8.38	11,223,534	\$10.26	20,171,004	\$18.65
U.S. I	8,873,451	\$7.20	11,161,922	\$8.49	20,035,373	\$15.70

Table 3.10: Yearly costs (billions) and future QALYs for newly diagnosed diabetes patients using no treatment, optimal guidelines (R_0 = \$100,000, R_0 = \$250,000, and R_0 = \$1 billion), and U.S. I.

	Males		Females		Total Population	
	QALYs	Cost (billions)	QALYs	Cost (billions)	QALYs	Cost (billions)
No Treatment	11,058,085	\$9.92	13,186,620	\$8.73	24,244,704	\$18.65
Optimal (R_0 = \$100,000)	11,809,410	\$13.04	14,192,325	\$12.73	26,001,736	\$25.77
Optimal (R_0 = \$250,000)	11,829,809	\$14.10	14,220,778	\$14.09	26,050,588	\$28.19
Optimal (R_0 = \$1 billion)	11,847,119	\$15.99	14,239,862	\$15.79	26,086,981	\$31.78
U.S. I	11,831,099	\$16.28	14,215,304	\$16.86	26,046,403	\$33.14

Table 3.9 provides the total expected LYs and medication costs before an event or death for newly diagnosed diabetes patients aged 40 or older with no treatment, optimal guidelines (R_0 = \$100,000, R_0 = \$250,000, and R_0 = \$10 million), and U.S. I. These results correspond to the primary prevention objective. Using the willingness-to-pay factor of R_0 = \$250,000 results in an optimal policy with the greatest LYs that still has fewer medication costs than U.S. I. Implementing the optimal guidelines (R_0 = \$250,000) would result in a total yearly savings of \$701 million in medication costs while increasing event-free LYs by over 116,000 as compared to U.S. I. The optimal guidelines policy with R_0 = \$10 million results in the maximum LYs before an event or death. However, these additional event-free LYs come at a cost of \$186,000/LY over the optimal policy with R_0 = \$250,000. In comparison, the optimal guidelines (R_0 = \$250,000) result in additional LYs over no treatment at a cost of less than \$7,500/LY.

Table 3.10 provides similar results for the primary and secondary prevention results. The potential yearly QALYs and costs of medication and treatment of events are provided for no treatment, optimal guidelines (R_0 = \$100,000, R_0 = \$250,000, and R_0 = \$1 billion), and U.S. I. For this perspective the optimal guideline policies are less costly than U.S. I for all values of R_0. Using willingness-to-pay factor R_0 = \$1 billion results in the greatest expected QALYs at \$7,126/QALY over no treatment. In comparison, U.S. I results in additional QALYs over no treatment at \$8,042/QALY.

It is important to note that the cost per event-free LY is greater for females than for males with the primary prevention results, while the cost per QALY is greater for males than for females with the primary and secondary prevention results.

3.5 Conclusions

The use of coordinated treatment of blood pressure and cholesterol can reduce the costs of treatment and hospitalization while improving length and quality of life. For primary prevention, the same event-free LYs from U.S. I could be achieved with optimal guidelines at a reduced cost. Using the optimal guidelines with large values of R_0 allow for additional LYs before an

event or death over U.S. I, though costs can be higher.

For the combined primary and secondary prevention, the model-based optimal treatment policies result in medications being initiated over a longer time period than with U.S. I, producing a reduction in costs and in some cases improvement in expected QALYs. Savings at the population level are particularly impressive when the optimal policies are applied to the female patients. Since female patients have a later onset of stroke and CHD risk, compared to males, the delayed initiation of drugs from the optimal treatment policies is sufficient to manage risk of events while improving quality of life and reducing costs. Female patients can save at least $7,378 per patient over a lifetime compared to the U.S. and international guidelines. The main impact of the optimal policies for male patients is seen on the cost side. Male patients can save at least $4,573 per patient in lifetime costs compared to the U.S. and international guidelines. Implementation of the optimal policies could produce approximately the same expected QALYs as with the implementation of U.S. I while greatly reducing the overall cost of treatment.

We have provided the potential yearly benefit of applying optimal guidelines to the U.S. diabetes population. These results show that there could be yearly savings of at least $1.4 billion at the population level in medication and hospitalization costs. Using the optimal treatment policies could also improve QALYs. The optimal policies also resulted in much of the treatment intensification occurring later in life when primary and secondary prevention is considered. Since the risk of stroke and CHD events is affected by both blood pressure and cholesterol levels, it is logical that the management of these risk factors should be done simultaneously. Our research shows that coordinated optimal treatment of cholesterol and blood pressure in patients with diabetes results in medications being prescribed later in life for each patient.

We have quantified the benefits of coordinated treatment and highlighted the need for the U.S. treatment guidelines to coordinate treatment for blood pressure and cholesterol medications. The coordination of these types of medications is particularly beneficial since both blood pressure and cholesterol medications are used to prevent the stroke and CHD events. While we have presented the benefits of coordinated treatment for the management of blood pressure

and cholesterol in diabetes patients, there are general insights that can be gained from the management of multiple risk factors for other types of patients.

Chapter 4

Approximate Dynamic Programming Approaches for Optimal Treatment

The underlying problem for the multiple medication decision problem presented in Chapter 3 is actually a continuous-state problem. Thus, the discrete-state MDP presented in Chapter 3 was an approximation of the true problem. In this chapter we explore the use of ADP methods, applied to the continuous-state version of the problem, to determine which approximation yields the best policy for maximizing expected future rewards. First we describe the exact MDP for the continuous-state problem. Next, we discuss two different ADP methods for solving the continuous-state MDP. Finally, we present a Monte Carlo simulation model to compare the policies found under several different implementations of the ADP methods.

4.1 Continuous-State MDP Formulation

We let $t = 0, \ldots, T$ denote discrete-time decision epochs at which treatment decisions are made. We define ξ_t to be a vector to represent the patient's health status at time t. This

vector is composed of elements that represent the patient's LR ($\ell_t^{\text{LR}} \in \mathcal{L}_{\text{LR}} = [0, \text{LR}^{\max}]$), SBP ($\ell_t^{\text{SBP}} \in \mathcal{L}_{\text{SBP}} = [0, \text{SBP}^{\max}]$), number of CHD events ($\ell_t^{\text{CHD}} \in \mathcal{L}_{\text{CHD}} = \{0, \ldots, c\}$), and number of stroke events ($\ell_t^{\text{S}} \in \mathcal{L}_{\text{S}} = \{0, \ldots, s\}$). Thus, $\xi_t \in \mathcal{L} = \mathcal{L}_{\text{LR}} \times \mathcal{L}_{\text{SBP}} \times \mathcal{L}_{\text{S}} \times \mathcal{L}_{\text{CHD}}$. The set of medication states is denoted by $\mathcal{M} = \{\boldsymbol{m}_t = (m_{1,t}, m_{2,t}, \ldots, m_{n,t}) : m_{i,t} \in \{0,1\}, \forall i = 1, 2, \ldots, n, \forall t = 0, \ldots, T\}$, where n denotes the number of available medications. If $m_{i,t} = 0$, the patient is not on medication i at time t, and if $m_{i,t} = 1$, the patient is on medication i at time t. As in Chapter 3, treatment decisions are assumed to be irreversible, and therefore if $m_{i,t} = 1$, then $m_{i,t'} = 1$ for all $t' > t$. The available actions are defined as follows:

$$A_{(\xi_t, m_{i,t})} = \begin{cases} \{I_i, W_i\} & \text{if } m_{i,t} = 0, \\ \{W_i\} & \text{if } m_{i,t} = 1, \end{cases} \tag{4.1}$$

where $\boldsymbol{A}_{(\xi_t, \boldsymbol{m}_t)} = \{A_{(\xi_t, m_{1,t})} \times A_{(\xi_t, m_{2,t})} \times \cdots \times A_{(\xi_t, m_{n,t})}\}$. Action $\boldsymbol{a}(\xi_t, \boldsymbol{m}_t) \in \boldsymbol{A}_{(\xi_t, \boldsymbol{m}_t)}$ denotes the action taken in state $(\xi_t, \boldsymbol{m}_t)$. The action $\boldsymbol{a}(\xi_t, \boldsymbol{m}_t)$ is also denoted by $\boldsymbol{a}$ when the health and medication states at time t are otherwise clearly defined.

A patient's LR and SBP are assumed to follow a Markov process defined as follows:

$$\ell_{t+1}^{\text{LR}} = \ell_t^{\text{LR}}\big(1 + z_t^{\text{LR}}\big)\big(1 + y_t^{\text{LR}}(\boldsymbol{a}(\xi_t, \boldsymbol{m}_t))\big), \tag{4.2}$$

$$\ell_{t+1}^{\text{SBP}} = \ell_t^{\text{SBP}}\big(1 + z_t^{\text{SBP}}\big)\big(1 + y_t^{\text{SBP}}(\boldsymbol{a}(\xi_t, \boldsymbol{m}_t))\big), \tag{4.3}$$

where z_t^{LR} and z_t^{SBP} are random variables, with probability density functions (p.d.f.s) $f_{\text{LR},t}(\cdot)$ and $f_{\text{SBP},t}(\cdot)$, defining percentage changes in LR and SBP from natural variation over time, respectively, and $y_t^{\text{LR}}(\boldsymbol{a}(\xi_t, \boldsymbol{m}_t))$ and $y_t^{\text{SBP}}(\boldsymbol{a}(\xi_t, \boldsymbol{m}_t))$ are the (deterministic) effects of medications, defining the percentage changes in LR and SBP due to medications started in year t, respectively. The number of CHD events are assumed to be 0 at $t = 0$, and the progression of ℓ_t^{CHD} is defined by the following:

$$\ell_{t+1}^{\text{CHD}} = \begin{cases} \ell_t^{\text{CHD}} + 1 & \text{if the patient has a CHD event between } t \text{ and } t+1, \\ \ell_t^{\text{CHD}} & \text{otherwise.} \end{cases} \tag{4.4}$$

The patient is also assumed to have had 0 stroke events at $t = 0$, and the progression of ℓ_t^{S} is defined by the following:

$$\ell_{t+1}^{\text{S}} = \begin{cases} \ell_t^{\text{S}} + 1 & \text{if the patient has a stroke between } t \text{ and } t+1, \\ \ell_t^{\text{S}} & \text{otherwise.} \end{cases} \tag{4.5}$$

Finally, the medication state is updated in the following way for all $i \in 1, \ldots, n$:

$$m_{i,t+1} = \begin{cases} 1 & \text{if } m_{i,t} = 0 \text{ and the action for medication } i \text{ is } I_i, \\ m_{i,t} & \text{otherwise.} \end{cases} \tag{4.6}$$

Since the state space is continuous, we assume the probabilities of transitioning from particular LR values, $\ell_t^{\text{LR}} \in \mathcal{L}_{\text{LR}}$, to a range of LR values, $L' \subset \mathcal{L}_{\text{LR}}$, specify a p.d.f. given by

$$\int_{L' \subset \mathcal{L}_{\text{LR}}} q_t^{\text{LR}}(\ell_{t+1}^{\text{LR}} | \ell_t^{\text{LR}}, \boldsymbol{m}_t) d\ell_{t+1}^{\text{LR}} = \Pr\{\ell_{t+1}^{\text{LR}} \in L' | \ell_t^{\text{LR}}, \boldsymbol{m}_t\}. \tag{4.7}$$

A similar p.d.f. is specified for SBP:

$$\int_{S' \subset \mathcal{L}_{\text{SBP}}} q_t^{\text{SBP}}(\ell_{t+1}^{\text{SBP}} | \ell_t^{\text{SBP}}, \boldsymbol{m}_t) d\ell_{t+1}^{\text{SBP}} = \Pr\{\ell_{t+1}^{\text{SBP}} \in S' | \ell_t^{\text{SBP}}, \boldsymbol{m}_t\}. \tag{4.8}$$

Fatal stroke and CHD events occur with probability $\widetilde{\pi}_t^{\text{S}}(\xi_t, \boldsymbol{m}_t)$ and $\widetilde{\pi}_t^{\text{CHD}}(\xi_t, \boldsymbol{m}_t)$, respectively. The probability of death from other causes is π_t^{O}. Given that the patient is in state ξ_t with medication status $\boldsymbol{m}_t$ at epoch t, the probability of moving into one of the absorbing states $d \in \mathcal{D}$ at epoch $t+1$ is denoted by $\bar{p}_t^{\boldsymbol{m}_t}(d|\xi_t)$, where

$$\bar{p}_t^{\boldsymbol{m}_t}(d|\xi_t) = \begin{cases} \pi_t^{\mathrm{O}} & \text{if } d = \mathcal{D}_{\mathrm{O}}, \\ \widetilde{\pi}_t^{\mathrm{CHD}}(\xi_t, \boldsymbol{m}_t) & \text{if } d = \mathcal{D}_{\mathrm{CHD}}, \\ \widetilde{\pi}_t^{\mathrm{S}}(\xi_t, \boldsymbol{m}_t) & \text{if } d = \mathcal{D}_{\mathrm{S}}, \end{cases} \tag{4.9}$$

for $\xi_t \in \mathcal{L}$ and $\boldsymbol{m}_t \in \mathcal{M}$, and $\bar{p}_t^{\boldsymbol{m}_t}(d|d) = 1$ for all $t \in 1, \ldots, T$. In addition, we assume the probabilities of transitioning from particular state values $\xi_t \in \mathcal{L}$ to a range of state values, $X' \subset \mathcal{L}$, specify a p.d.f. given by the following:

$$\int_{X' \subset \mathcal{L}} p_t^{\boldsymbol{m}_t}(\xi_{t+1}|\xi_t) d\xi_{t+1} = \left[1 - \sum_{d \in \mathcal{D}} \bar{p}_t^{\boldsymbol{m}_t}(d|\xi_t)\right] \int_{X' \subset \mathcal{L}} q_t(\xi_{t+1}|\xi_t, \boldsymbol{m}_t) d\xi_{t+1} \tag{4.10}$$

where $q_t(\xi_{t+1}|\xi_t, \boldsymbol{m}_t)$ is defined by Equations (4.7) and (4.8) for the transitions among LR and SBP states, respectively.

The following defines the optimality equations for all $\xi_t \in \mathcal{L}, \boldsymbol{m}_t \in \mathcal{M}, t = 0, \ldots, T-1$, for all patients that have not yet entered an absorbing state:

$$\begin{aligned} v_t(\xi_t, \boldsymbol{m}_t) = \max_{\boldsymbol{a} \in \boldsymbol{A}_{(\xi_t, \boldsymbol{m}_t)}} \Bigg\{ & r(\xi_t, \boldsymbol{m}_t') \\ + \lambda \int_{\ell_{t+1}^{\mathrm{LR}} \in \mathcal{L}_{\mathrm{LR}}} \int_{\ell_{t+1}^{\mathrm{SBP}} \in \mathcal{L}_{\mathrm{SBP}}} & \left(\sum_{\ell_{t+1}^{\mathrm{S}}=0}^{s} \sum_{\ell_{t+1}^{\mathrm{CHD}}=0}^{c} p_t^{\boldsymbol{m}_t'}(\xi_{t+1}|\xi_t) v_{t+1}(\xi_{t+1}, \boldsymbol{m}_t') \right) d\ell_{t+1}^{\mathrm{SBP}} d\ell_{t+1}^{\mathrm{LR}} \Bigg\}, \end{aligned} \tag{4.11}$$

where $\boldsymbol{m}_t$ represents the patient's medication status at the beginning of epoch t and $\boldsymbol{m}_t'$ represents the patient's medication status during epoch t, found by taking action $\boldsymbol{a}$ into account. For patients that have entered an absorbing state (e.g., death from other causes) $v_t(\xi_t, \boldsymbol{m}_t) = 0$. The patient's medication status at the beginning of decision epoch $t+1$ is defined as $\boldsymbol{m}_{t+1} = \boldsymbol{m}_t'$. For $t = T$, the MDP has the following boundary condition for all $\xi_T \in \mathcal{L}, \boldsymbol{m}_T \in \mathcal{M}$:

$$v_T(\xi_T, \boldsymbol{m}_T) = \mu(\xi_T, \boldsymbol{m}_T), \tag{4.12}$$

where $\mu(\xi_T, \boldsymbol{m}_T)$ represents the expected future rewards accrued after the decision horizon. The optimal action, denoted by $\boldsymbol{a}^*(\xi_t, \boldsymbol{m}_t)$, for a patient in health state ξ_t, and medication state $\boldsymbol{m}_t$, is defined by the above optimality equations and can be written as follows, for all $t = 0, \ldots, T-1$:

$$\boldsymbol{a}^*(\xi_t, \boldsymbol{m}_t) = \underset{\boldsymbol{a} \in \boldsymbol{A}_{(\xi_t, \boldsymbol{m}_t)}}{\operatorname{argmax}} \Bigg\{ r(\xi_t, \boldsymbol{m}'_t) + \lambda \int_{\ell^{\mathrm{LR}}_{t+1} \in \mathcal{L}_{\mathrm{LR}}} \int_{\ell^{\mathrm{SBP}}_{t+1} \in \mathcal{L}_{\mathrm{SBP}}} \Bigg(\sum_{\ell^{\mathrm{S}}_{t+1}=0}^{s} \sum_{\ell^{\mathrm{CHD}}_{t+1}=0}^{c} p_t^{\boldsymbol{m}'_t}(\xi_{t+1}|\xi_t) v_{t+1}(\xi_{t+1}, \boldsymbol{m}'_t) \Bigg) d\ell^{\mathrm{SBP}}_{t+1} \, d\ell^{\mathrm{LR}}_{t+1} \Bigg\}. \tag{4.13}$$

We present two ADP approaches for solving this continuous-state MDP. The remaining sections are organized as follows: In Section 4.2 we present a finite-state MDP to approximate the above continuous-state MDP. This finite-state formulation plays a role in each of the two ADP methods. In Section 4.3 we present the first ADP method which is based on solving the finite-state MDP and then mapping the finite-state policy to the true continuous-state problem. In Section 4.4 we present the second ADP method which uses a basis function approximation of the value function, where an LP is used to estimate the basis function weights. Section 4.5 provides details about specific reward functions and details about the simulation model for comparing the ADP techniques. Section 4.6 contains results of numerical experiments to evaluate the ADP methods. Finally, Section 4.7 provides conclusions.

4.2 Finite-State MDP

The basic idea of both the ADP methods we describe in this chapter is to use a finite-state version of the continuous-state MDP. One way to achieve this is to partition the LR state space $\mathcal{L}_{\mathrm{LR}} = [0, \mathrm{LR}^{\max}]$ into q_{LR} subsets ($1 \leq q_{\mathrm{LR}} < \infty$), $S^{\mathrm{LR}}_1, S^{\mathrm{LR}}_2, \ldots, S^{\mathrm{LR}}_{q_{\mathrm{LR}}}$, where $\mathcal{L}_{\mathrm{LR}} = S^{\mathrm{LR}}_1 \cup S^{\mathrm{LR}}_2 \cup \cdots \cup S^{\mathrm{LR}}_{q_{\mathrm{LR}}}$ and $S^{\mathrm{LR}}_i \cap S^{\mathrm{LR}}_j = \emptyset$, for all $i \neq j$. We also partition the SBP state space $\mathcal{L}_{\mathrm{SBP}} = [0, \mathrm{SBP}^{\max}]$ into q_{SBP} subsets ($1 \leq q_{\mathrm{SBP}} < \infty$), $S^{\mathrm{SBP}}_1, S^{\mathrm{SBP}}_2, \ldots, S^{\mathrm{SBP}}_{q_{\mathrm{SBP}}}$, where

$\mathcal{L}_{\text{SBP}} = S_1^{\text{SBP}} \cup S_2^{\text{SBP}} \cup \cdots \cup S_{q_{\text{SBP}}}^{\text{SBP}}$ and $S_i^{\text{SBP}} \cap S_j^{\text{SBP}} = \emptyset$, for all $i \neq j$. We assume $\text{LR}^{\max} \leq \infty$ and $\text{SBP}^{\max} \leq \infty$. Figure 4.1 provides an example of how a continuous, bounded state space over LR and SBP may be partitioned. The dots in each cell of the partition represent the conditional mean LR and SBP values associated with the cell. This collection of conditional means represents the finite set of states for each decision epoch t.

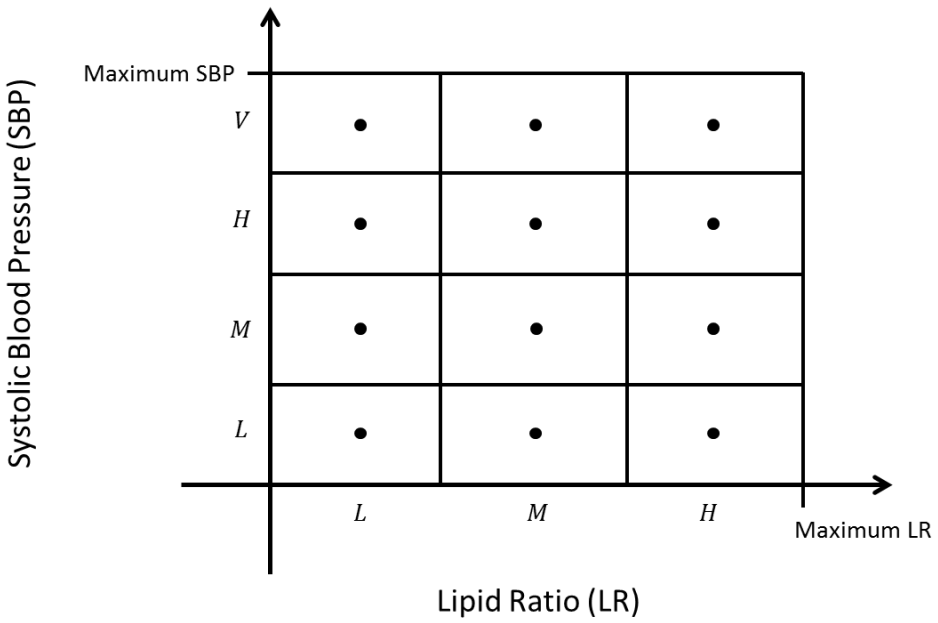

Figure 4.1: Example of the partitioned continuous state space for LR and SBP. For this particular partitioning, which is used in the numerical experiments, LR is divided into three ($q_{\text{LR}} = 3$) discrete states (low (L), medium (M), and high (H)), and SBP is divided into four ($q_{\text{SBP}} = 4$) discrete states (low (L), medium (M), high (H), and very high (V)). The dots in each cell of the partition represent the conditional mean LR and SBP values for the cell.

For $j = 1, \ldots, q_{\text{LR}}$ and $k = 1, \ldots, q_{\text{SBP}}$, we define $\text{UB}(S_j^{\text{LR}})$ and $\text{UB}(S_k^{\text{SBP}})$ to be the upper bounds of subsets S_j^{LR} and S_k^{SBP}, respectively. Note that $\text{UB}(S_{q_{\text{LR}}}^{\text{LR}}) = \text{LR}^{\max}$ and $\text{UB}(S_{q_{\text{SBP}}}^{\text{SBP}}) = \text{SBP}^{\max}$. For $j = 1, \ldots, q_{\text{LR}}$ and $k = 1, \ldots, q_{\text{SBP}}$, we define $\text{LB}(S_j^{\text{LR}})$ and $\text{LB}(S_k^{\text{SBP}})$ be the lower bounds of subsets S_j^{LR} and S_k^{SBP}, respectively. Note that $\text{LB}(S_1^{\text{LR}}) = 0$ and $\text{LB}(S_1^{\text{SBP}}) = 0$ represent theoretical lower bounds for these risk factors. For $j = 1, \ldots, q_{\text{LR}}$ and $k = 1, \ldots, q_{\text{SBP}}$,

we define $\eta(S_j^{\mathrm{LR}})$ and $\eta(S_k^{\mathrm{SBP}})$ to be the conditional mean LR and SBP values on subsets S_j^{LR} and S_k^{SBP}, respectively. A discrete LR value will be associated with each of the q_{LR} subsets for LR, and a discrete SBP value will be associated with each of the q_{SBP} subsets for SBP. The following two equations define the discrete LR and SBP values from each new subset, respectively:

$$g^{\mathrm{LR}}(\ell_t^{\mathrm{LR}}) = \begin{cases} \eta(S_1^{\mathrm{LR}}) & 0 \leq \ell_t^{\mathrm{LR}} \leq \mathrm{UB}(S_1^{\mathrm{LR}}), \\ \eta(S_2^{\mathrm{LR}}) & \mathrm{UB}(S_1^{\mathrm{LR}}) < \ell_t^{\mathrm{LR}} \leq \mathrm{UB}(S_2^{\mathrm{LR}}), \\ \vdots & \vdots \\ \eta(S_{q_{\mathrm{LR}}}^{\mathrm{LR}}) & \mathrm{UB}(S_{q_{\mathrm{LR}}-1}^{\mathrm{LR}}) < \ell_t^{\mathrm{LR}} \leq \mathrm{LR}^{\max}, \end{cases} \tag{4.14}$$

$$g^{\mathrm{SBP}}(\ell_t^{\mathrm{SBP}}) = \begin{cases} \eta(S_1^{\mathrm{SBP}}) & 0 \leq \ell_t^{\mathrm{SBP}} \leq \mathrm{UB}(S_1^{\mathrm{SBP}}), \\ \eta(S_2^{\mathrm{SBP}}) & \mathrm{UB}(S_1^{\mathrm{SBP}}) < \ell_t^{\mathrm{SBP}} \leq \mathrm{UB}(S_2^{\mathrm{SBP}}), \\ \vdots & \vdots \\ \eta(S_{q_{\mathrm{SBP}}}^{\mathrm{SBP}}) & \mathrm{UB}(S_{q_{\mathrm{SBP}}-1}^{\mathrm{SBP}}) < \ell_t^{\mathrm{SBP}} \leq \mathrm{SBP}^{\max}. \end{cases} \tag{4.15}$$

We define the following two discrete sets of LR and SBP values: $\tilde{\mathcal{L}}^{\mathrm{LR}} = \{\eta(S_1^{\mathrm{LR}}), \eta(S_2^{\mathrm{LR}}), \ldots, \eta(S_{q_{\mathrm{LR}}}^{\mathrm{LR}})\}$ and $\tilde{\mathcal{L}}^{\mathrm{SBP}} = \{\eta(S_1^{\mathrm{SBP}}), \eta(S_2^{\mathrm{SBP}}), \ldots, \eta(S_{q_{\mathrm{SBP}}}^{\mathrm{SBP}})\}$. The set of states for the finite-state MDP are defined as $\tilde{\mathcal{L}} = \tilde{\mathcal{L}}^{\mathrm{LR}} \times \tilde{\mathcal{L}}^{\mathrm{SBP}} \times \mathcal{L}^{\mathrm{CHD}} \times \mathcal{L}^{\mathrm{S}}$. The set $\tilde{\mathcal{L}}^{\mathrm{LR}}$ is indexed by $\tilde{\ell}_t^{\mathrm{LR}}$, and the set $\tilde{\mathcal{L}}^{\mathrm{SBP}}$ is indexed by $\tilde{\ell}_t^{\mathrm{SBP}}$. Summing over $\tilde{\mathcal{L}}^{\mathrm{LR}}$, $\tilde{\mathcal{L}}^{\mathrm{SBP}}$, $\mathcal{L}^{\mathrm{CHD}}$, and $\mathcal{L}^{\mathrm{S}}$ is denoted by summing over $\tilde{\mathcal{L}}$. The set $\tilde{\mathcal{L}}$ is indexed by $\tilde{\xi}_t$.

There are three types of transition probabilities used in this finite-state MDP: transition probabilities among health states, transition probabilities between health states and event states (fatal or nonfatal), and transition probabilities between health states and death from other causes. The transition probabilities among LR states are independent from the transition probabilities among SBP states. Using the p.d.f.s presented in Section 4.1, we define the transition probabilities between health states using the corresponding cumulative density functions (c.d.f.s). The c.d.f. for the LR transition probabilities is defined by $F_{z_t^{\mathrm{LR}}} = \int_{-\infty}^{x} f_{z_t^{\mathrm{LR}}}(y)dy$, and

the c.d.f. for the SBP transition probabilities is defined by $F_{z_t^{\text{SBP}}} = \int_{-\infty}^{x} f_{z_t^{\text{SBP}}}(y)dy$. Assuming the patient lives from time period t to time period $t+1$, and has medication status $\boldsymbol{m}_t$ during decision epoch t, the transition probabilities among the finite set of LR states, denoted by $\tilde{\mathcal{L}}^{\text{LR}}$, are defined by the following:

$$\begin{aligned}
&q_t^{\text{LR}}\left(\tilde{\ell}_{t+1}^{\text{LR}} = \eta(S_j^{\text{LR}}) | \tilde{\ell}_t^{\text{LR}} = \eta(S_i^{\text{LR}}), \boldsymbol{m}_t\right) \\
&= \Pr\left\{\text{LB}(S_j^{\text{LR}}) < \tilde{\ell}_{t+1}^{\text{LR}} \leq \text{UB}(S_j^{\text{LR}}) | \tilde{\ell}_t^{\text{LR}} = \eta(S_i^{\text{LR}}), \boldsymbol{m}_t\right\} \\
&= \Pr\left\{\text{LB}(S_j^{\text{LR}}) < \eta(S_i^{\text{LR}})(1+z_t^{\text{LR}})(1+y_t^{\text{LR}}(\boldsymbol{a}(\xi_t, \boldsymbol{m}_t))) \leq \text{UB}(S_j^{\text{LR}})\right\} \\
&= \Pr\left\{\frac{\text{LB}(S_j^{\text{LR}})}{\eta(S_i^{\text{LR}})(1+y_t^{\text{LR}}(\boldsymbol{a}(\xi_t, \boldsymbol{m}_t)))} - 1 < z_t^{\text{LR}} \leq \frac{\text{UB}(S_j^{\text{LR}})}{\eta(S_i^{\text{LR}})(1+y_t^{\text{LR}}(\boldsymbol{a}(\xi_t, \boldsymbol{m}_t)))} - 1\right\} \\
&= F_{z_t^{\text{LR}}}\left(\frac{\text{UB}(S_j^{\text{LR}})}{\eta(S_i^{\text{LR}})(1+y_t^{\text{LR}}(\boldsymbol{a}(\tilde{\xi}_t, \boldsymbol{m}_t)))} - 1\right) - F_{z_t^{\text{LR}}}\left(\frac{\text{LB}(S_j^{\text{LR}})}{\eta(S_i^{\text{LR}})(1+y_t^{\text{LR}}(\boldsymbol{a}(\tilde{\xi}_t, \boldsymbol{m}_t)))} - 1\right),
\end{aligned} \tag{4.16}$$

This derivation relies on the definition of the progression of a patient's LR defined in Equation (4.2). The transition probabilities among SBP states, denoted by $\tilde{\mathcal{L}}^{\text{SBP}}$, are similarly defined by the following:

$$\begin{aligned}
&q_t^{\text{SBP}}\left(\tilde{\ell}_{t+1}^{\text{SBP}} = \eta(S_j^{\text{SBP}}) | \tilde{\ell}_t^{\text{SBP}} = \eta(S_i^{\text{SBP}}), \boldsymbol{m}_t\right) \\
&= F_{z_t^{\text{SBP}}}\left(\frac{\text{UB}(S_j^{\text{LR}})}{\eta(S_i^{\text{SBP}})(1+y_t^{\text{SBP}}(\boldsymbol{a}(\tilde{\xi}_t, \boldsymbol{m}_t)))} - 1\right) - F_{z_t^{\text{SBP}}}\left(\frac{\text{LB}(S_j^{\text{SBP}})}{\eta(S_i^{\text{SBP}})(1+y_t^{\text{SBP}}(\boldsymbol{a}(\tilde{\xi}_t, \boldsymbol{m}_t)))} - 1\right).
\end{aligned} \tag{4.17}$$

The probabilities of events and death are defined by Equation (4.9). In addition, the probabilities among states are defined by the following:

$$p_t^{\boldsymbol{m}_t}(\tilde{\xi}_{t+1}|\tilde{\xi}_t) = \begin{cases} \left[1 - \sum_{d\in\mathcal{D}} \bar{p}_t^{\boldsymbol{m}_t}(d|\tilde{\xi}_t)\right] q_t(\tilde{\xi}_{t+1}|\tilde{\xi}_t, \boldsymbol{m}_t) & \text{if } \tilde{\xi}_t, \tilde{\xi}_{t+1} \in \tilde{\mathcal{L}}, \\ \sum_{d\in\mathcal{D}} \bar{p}_t^{\boldsymbol{m}_t}(d|\tilde{\xi}_t) & \text{if } \tilde{\xi}_t \in \tilde{\mathcal{L}},\ \tilde{\xi}_{t+1} \in \mathcal{D}, \\ 1 & \text{if } \tilde{\xi}_t = \tilde{\xi}_{t+1} \in \mathcal{D}, \\ 0 & \text{otherwise,} \end{cases} \tag{4.18}$$

where $q_t(\tilde{\xi}_{t+1}|\tilde{\xi}_t, \boldsymbol{m}_t)$ is defined by Equations (4.16) and (4.17) for the independent transitions among LR and SBP states, respectively.

The finite-state MDP is defined by the following optimality equations:

$$v_t(\tilde{\xi}_t, \boldsymbol{m}_t) = \max_{\boldsymbol{a}\in\boldsymbol{A}_{(\tilde{\xi}_t, \boldsymbol{m}_t')}} \left\{ r(\tilde{\xi}_t, \boldsymbol{m}_t') + \lambda \sum_{\tilde{\xi}_{t+1}\in\tilde{\mathcal{L}}} p_t^{\boldsymbol{m}_t'}(\tilde{\xi}_{t+1}|\tilde{\xi}_t) v_{t+1}(\tilde{\xi}_{t+1}, \boldsymbol{m}_t') \right\}, \tag{4.19}$$

for all $\tilde{\xi}_t \in \tilde{\mathcal{L}}, \boldsymbol{m}_t \in \mathcal{M}, t = 0, \ldots, T-1$. For $t = T$, we have the following boundary condition for all $\tilde{\xi}_T \in \tilde{\mathcal{L}}, \boldsymbol{m}_T \in \mathcal{M}$:

$$v_T(\tilde{\xi}_T, \boldsymbol{m}_T) = \mu(\tilde{\xi}_T, \boldsymbol{m}_T), \tag{4.20}$$

where $\mu(\tilde{\xi}_T, \boldsymbol{m}_T)$ is the expected future rewards after the end of the decision horizon. The optimal action for the finite-state MDP, for all $\tilde{\xi}_t \in \tilde{\mathcal{L}}, \boldsymbol{m}_t \in \mathcal{M}, t = 0, \ldots, T-1$, is given by the following:

$$\boldsymbol{a}^*(\tilde{\xi}_t, \boldsymbol{m}_t) = \operatorname*{argmax}_{\boldsymbol{a}\in\boldsymbol{A}_{(\tilde{\xi}_t, \boldsymbol{m}_t)}} \left\{ r(\tilde{\xi}_t, \boldsymbol{m}_t') + \lambda \sum_{\tilde{\xi}_{t+1}\in\tilde{\mathcal{L}}} p_t^{\boldsymbol{m}_t'}(\tilde{\xi}_{t+1}|\tilde{\xi}_t) v_{t+1}(\tilde{\xi}_{t+1}, \boldsymbol{m}_t') \right\}. \tag{4.21}$$

4.3 ADP Approach 1: Policy Mapping

For the first ADP approach we use the finite-state MDP as defined in Section 4.2. The finite-state MDP defined by Equations (4.19), (4.20), and (4.21) is solved using backward induction. The value function and optimal actions are computed for all $\tilde{\xi}_t \in \tilde{\mathcal{L}}, \boldsymbol{m}_t \in \mathcal{M}$. The approximate optimal policy found using backwards induction can be stored as a lookup table. We use the following to map the optimal actions from the finite-state MDP to the actions for the continuous states:

$$\widehat{\boldsymbol{a}}^*(\xi_t, \boldsymbol{m}_t) = \boldsymbol{a}^*(\tilde{\xi}_t, \boldsymbol{m}_t) \text{ if } \quad \text{LB}(S_i^{\text{LR}}) < \ell_t^{\text{LR}} \leq \text{UB}(S_i^{\text{LR}}), \text{ and} \tag{4.22}$$

$$\text{LB}(S_j^{\text{SBP}}) < \ell_t^{\text{SBP}} \leq \text{UB}(S_j^{\text{SBP}}). \tag{4.23}$$

The patient's LR, ℓ_t^{LR}, is mapped to $\tilde{\ell}_t^{\text{LR}} = \eta(S_i^{\text{LR}})$ for $\text{LB}(S_i^{\text{LR}}) < \ell_t^{\text{LR}} \leq \text{UB}(S_i^{\text{LR}})$. The patient's SBP, ℓ_t^{SBP}, is mapped to $\tilde{\ell}_t^{\text{SBP}} = \eta(S_j^{\text{SBP}})$ for $\text{LB}(S_j^{\text{SBP}}) < \ell_t^{\text{SBP}} \leq \text{UB}(S_j^{\text{SBP}})$. The action $\widehat{\boldsymbol{a}}^*$ is determined by the optimal action for $\tilde{\ell}_t^{\text{LR}}$ and $\tilde{\ell}_t^{\text{SBP}}$ from the finite-state MDP.

4.4 ADP Approach 2: Basis Function Approximation

In the second ADP approach, the value function for the continuous-state MDP is approximated using a basis function approximation represented as:

$$v_t(\xi_t, \boldsymbol{m}_t) \approx \tilde{v}_t(\xi_t, \boldsymbol{m}_t) = \sum_{k=1}^{K} w_{t,k}(\tilde{\xi}_t, \boldsymbol{m}_t) b_{t,k}(\xi_t, \boldsymbol{m}_t). \tag{4.24}$$

We consider K possible basis functions, $b_{t,k}(\xi_t, \boldsymbol{m}_t)$ for $k = 1, \ldots, K$. The basis functions are dependent on the patient's age (t), health status (ξ_t), and medication state ($\boldsymbol{m}_t$). Each basis function is weighted by a coefficient $w_{t,k}(\tilde{\xi}_t, \boldsymbol{m}_t)$ where $\tilde{\xi}_t$ represents the conditional mean values associated with $\tilde{\xi}_t$, defined by Equations (4.14) and (4.15). The problem reduces to finding weights $(w_{t,k}(\tilde{\xi}_t, \boldsymbol{m}_t))$ that give a good approximation of the optimal value function.

Note that the basis functions and weights are time dependent, reflecting the nonstationary nature of the MDP [5].

We chose to estimate the weights by solving an LP. The remainder of this section details the LP formulation and the specific basis functions used.

4.4.1 Linear Programming Formulation

In general, an MDP can be formulated as an LP [28, 81] where the value-to-go for a particular state and time period are the decision variables, and the constraints are defined by the optimality equations and boundary condition, such as in Equations (4.11) and (4.12). The LP to represent the continuous-state MDP is defined by the following:

$$\min z = \int_{\xi_t \in \mathcal{L}} \sum_{\boldsymbol{m}_t \in \mathcal{M}} \sum_{t=0}^{T} \alpha_t(\xi_t, \boldsymbol{m}_t) v_t(\xi_t, \boldsymbol{m}_t) \tag{4.25}$$

s.t.

$$v_t(\xi_t, \boldsymbol{m}_t) - \int_{\xi_{t+1} \in \mathcal{L}} p_t^{\boldsymbol{m}_t'}(\xi_{t+1}|\xi_t) v_{t+1}(\xi_{t+1}, \boldsymbol{m}_t') \geq r(\xi_t, \boldsymbol{m}_t'),$$

$$\text{for all } t = 0, \ldots, T-1, \boldsymbol{a} \in \boldsymbol{A}_{(\xi_t, \boldsymbol{m}_t)}, \xi_t \in \mathcal{L}, \boldsymbol{m}_t \in \mathcal{M}, \tag{4.26}$$

$$v_T(\xi_T, \boldsymbol{m}_T) = \mu(\xi_T, \boldsymbol{m}_T), \text{ for all } \xi_T \in \mathcal{L}, \boldsymbol{m}_T \in \mathcal{M}, \tag{4.27}$$

$$v_t(\xi_t, \boldsymbol{m}_t) \text{ URS, for all } \xi_t \in \mathcal{L}, \boldsymbol{m}_t \in \mathcal{M}, t = 0, \ldots, T. \tag{4.28}$$

The nonnegative objective function coefficients $\alpha_t(\xi_t, \boldsymbol{m}_t)$ can be set arbitrarily [81], though the condition $\int_{\xi_t \in \mathcal{L}} \sum_{\boldsymbol{m}_t \in \mathcal{M}} \sum_{t=0}^{T} \alpha_t(\xi_t, \boldsymbol{m}_t) = 1$ (see page 223 of Puterman [81]) provides a probability distribution over $\mathcal{L} \times \mathcal{M}$ for all time periods which may allow for clearer interpretation of the results. One way of defining the probability distribution would be to associate a probability with each state equal to the probability that the state will be visited. The choice of these objective function coefficients can influence the quality of the policies obtained [28].

The number of decision variables ($v_t(\xi_t, \boldsymbol{m}_t)$) and constraints are infinite since there are

infinitely-many states $\xi_t \in \mathcal{L}$. Therefore, we reduce the number of decision variables to a finite number by approximating $v_t(\xi_t, \boldsymbol{m}_t)$ with the basis function approximation: $v_t(\xi_t, \boldsymbol{m}_t) = \sum_{k=1}^{K} w_{t,k}(\tilde{\xi}_t, \boldsymbol{m}_t) b_{t,k}(\xi_t, \boldsymbol{m}_t)$. The approximate LP (ALP) is provided below:

$$\min z = \int_{\xi_t \in \mathcal{L}} \sum_{\boldsymbol{m}_t \in \mathcal{M}} \sum_{t=0}^{T} \sum_{k=1}^{K} \alpha_t(\xi_t, \boldsymbol{m}_t) w_{t,k}(\tilde{\xi}_t, \boldsymbol{m}_t) b_{t,k}(\xi_t, \boldsymbol{m}_t) \tag{4.29}$$

s.t.

$$\sum_{k=1}^{K} w_{t,k}(\tilde{\xi}_t, \boldsymbol{m}_t) b_{t,k}(\xi_t, \boldsymbol{m}_t) - \sum_{\xi_{t+1} \in \mathcal{L}} p_t^{\boldsymbol{m}_t'}(\xi_{t+1}|\xi_t) \sum_{k=1}^{K} w_{t+1,k}(\tilde{\xi}_{t+1}, \boldsymbol{m}_t') b_{t+1,k}(\xi_{t+1}, \boldsymbol{m}_t')$$
$$\geq r(\xi_t, \boldsymbol{m}_t'), \text{ for all } t = 0, \ldots, T-1, \boldsymbol{a} \in \boldsymbol{A}_{(\xi_t, \boldsymbol{m}_t)}, \xi_t \in \mathcal{L}, \boldsymbol{m}_t \in \mathcal{M}, \tag{4.30}$$

$$\sum_{k=1}^{K} w_{T,k}(\tilde{\xi}_T, \boldsymbol{m}_T) b_{T,k}(\xi_T, \boldsymbol{m}_T) = \mu(\xi_T, \boldsymbol{m}_T), \text{ for all } \xi_T \in \mathcal{L}, \boldsymbol{m}_T \in \mathcal{M}, \tag{4.31}$$

$$w_{t,k}(\tilde{\xi}_t, \boldsymbol{m}_t) \text{ URS, for all } t = 0, \ldots, T, k = 1, \ldots, K, \tilde{\xi}_t \in \tilde{\mathcal{L}}, \boldsymbol{m}_t \in \mathcal{M}. \tag{4.32}$$

The above ALP has the benefit over formulation (4.25) - (4.28) of having a finite number of decision variables; however, the constraints of the ALP are defined for each state-action pair, and there are infinitely-many states $\xi_t \in \mathcal{L}$. Thus, the ALP has infinitely-many constraints. Several methods have been proposed in the context of ADP to solve LPs with a large or infinite number of constraints. These approaches reduce the ALP to a reduced LP (RLP), which is a relaxation of the ALP that has a finite number of constraints. One common approach is constraint sampling [29]. Given a number of constraints to be sampled and a probability measure over the collection of state-action pairs (i.e., a probability measure over the constraints of the ALP), constraints are sampled to reduce the size of the LP and bound the probability that a non-sampled constraint will be violated. Another common approach is column generation [75]. In this approach the dual of the ALP is considered. While the dual has a finite number of constraints, there are an infinite number of decision variables, one per state-action pair. Column generation begins with a small set of feasible dual variables. One or more violated

constraints are identified in the primal, and the corresponding dual variables are added to the set of feasible dual variables. The dual is resolved, and the column generation procedure continues until there are no more violated constraints in the primal ALP or the ALP is close enough to optimality. Trick and Zin [104, 105] provide two additional approaches to generate a relaxation of the ALP. One approach uses shadow prices to develop a set of discrete grid points and determines where the grid points should be placed to increase accuracy, and the other uses cubic splines to approximate the value function.

Another approach that has been used to solve LPs with large numbers of constraints is constraint aggregation [93]. To our knowledge, this technique has not been used with LPs to solve ADPs. We implement constraint aggregation to form a RLP. We use the same state aggregation to aggregate the constraints that we used to generate the finite-state MDP. In particular, we use the finite number of aggregated constraints defined by the optimality equations and boundary condition of the finite-state MDP. With the finite state space, we can also make the weights state dependent. Thus, for the RLP, $v_t(\tilde{\xi}_t, \boldsymbol{m}_t) = \sum_{k=1}^{K} w_{t,k}(\tilde{\xi}_t, \boldsymbol{m}_t) b_{t,k}(\tilde{\xi}_t, \boldsymbol{m}_t)$. With state-dependent weights, when the LP is solved the weights found result in tight constraints associated with the optimal actions for each state, and all end-of-horizon constraints are equality constraints. If the weights are not state dependent, none of the constraints are guaranteed to be tight, and the end-of-horizon constraints would have to be inequalities. The RLP can be written as follows:

$$\min z = \sum_{\tilde{\xi}_t \in \tilde{\mathcal{L}}} \sum_{\boldsymbol{m}_t \in \mathcal{M}} \sum_{t=0}^{T} \sum_{k=1}^{K} \alpha_t(\tilde{\xi}_t, \boldsymbol{m}_t) w_{t,k}(\tilde{\xi}_t, \boldsymbol{m}_t) b_{t,k}(\tilde{\xi}_t, \boldsymbol{m}_t) \tag{4.33}$$

s.t.

$$\sum_{k=1}^{K} w_{t,k}(\tilde{\xi}_t, \boldsymbol{m}_t) b_{t,k}(\tilde{\xi}_t, \boldsymbol{m}_t) - \sum_{\tilde{\xi}_{t+1} \in \tilde{\mathcal{L}}} p_t^{\boldsymbol{m}_t'}(\tilde{\xi}_{t+1} | \tilde{\xi}_t) \sum_{k=1}^{K} w_{t+1,k}(\tilde{\xi}_{t+1}, \boldsymbol{m}_t) b_{t+1,k}(\tilde{\xi}_{t+1}, \boldsymbol{m}_t')$$

$$\geq r(\tilde{\xi}_t, \boldsymbol{m}_t'), \text{ for all } t = 0, \ldots, T-1, \boldsymbol{a} \in \boldsymbol{A}_{(\tilde{\xi}_t, \boldsymbol{m}_t)}, \tilde{\xi}_t \in \tilde{\mathcal{L}}, \boldsymbol{m}_t \in \mathcal{M}, \tag{4.34}$$

$$\sum_{k=1}^{K} w_{T,k}(\tilde{\xi}_T, \boldsymbol{m}_T) b_{T,k}(\tilde{\xi}_T, \boldsymbol{m}_T) = \mu(\tilde{\xi}_T, \boldsymbol{m}_T), \text{ for all } \tilde{\xi}_T \in \tilde{\mathcal{L}}, \boldsymbol{m}_T \in \mathcal{M}, \tag{4.35}$$

$$w_{t,k}(\tilde{\xi}_t, \boldsymbol{m}_t) \text{URS, for all } k = 1, \ldots, K, \tilde{\xi}_t \in \tilde{\mathcal{L}}, \boldsymbol{m}_t \in \mathcal{M}, t = 0, \ldots, T. \tag{4.36}$$

The ALP is a restriction of the original LP associated with the continuous-state MDP since the basis function weights are not dependent on the state (i.e., there are finitely-many decision variables). The aggregation of constraints in the RLP constitutes a relaxation of the ALP. Thus, the solution to the RLP provides a lower bound to the upper bound of the original LP. As a result, unfortunately there is no direct relationship between the solution of the original LP and the solution of the RLP.

As written in the RLP, the weights for each basis function are dependent on time, health state, and medication state. The dependency of the weights on health state could be removed, denoted by $\omega_{t,k,\boldsymbol{m}_t}$, so the basis function approximations do not depend as much on the way the states were aggregated. Using basis functions with weights $\omega_{t,k,\boldsymbol{m}_t}$ and basis functions with weights $w_{t,k}(\tilde{\xi}_t, \boldsymbol{m}_t)$ is an example of tile coding. Tile coding [92], originally referred to as *Cerebellar Model Articulatory Controllers* (CMACs), is a method in which the state space is partitioned. A particular partitioning of the state space is called a tiling, and multiple overlapping tilings are sometimes used in tile coding. The value for a particular point is estimated by summing the values for the point from each of the tilings containing the point. Figure 4.2 depicts the example of two tilings described above in which one set of weights are state depen-

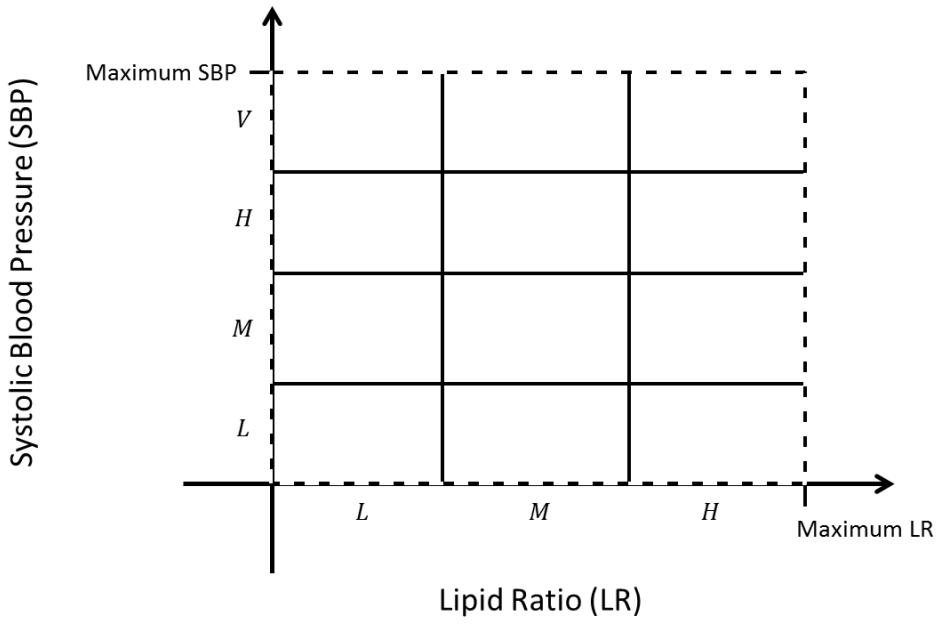

Figure 4.2: Example of the tile coding in which the bounded continuous state space for LR and SBP is partitioned using two tilings. One tiling is shown with solid lines dividing the state space, and the other tiling is shown with a dotted line, providing a single tile over the entire state space.

dent (shown with the tiling of solid lines dividing the bounded state space) and the other set of weights (shown with one tile over the entire state space represented by the dotted line) only depend on time and medication status.

4.4.2 Basis Functions

One difficulty in using the basis function approximation is finding the appropriate basis functions to use. We provide results using different types of basis functions, including survival functions. The motivation for the choice of survival functions is that the value function and optimal actions are highly influenced by the probability of nonfatal and fatal adverse events (stroke and CHD events). In particular, we use basis functions that define the patient's probability of no CHD and probability of no stroke in the next τ years, respectively, as estimated by the UKPDS risk equations [97, 61]:

Table 4.1: Parameter values for Equations (4.37), (4.38), (4.39), and (4.40) found in Stevens et al. [97] and Kothari et al. [61].

Description	Parameter	Value
Intercept	$q_{0,\text{CHD}}$	0.0112
Age	β_1	1.059
Sex	β_2	0.525
Race	β_3	0.390
Smoking Status	β_4	1.350
HbA1c	β_5	1.144
SBP	β_6	1.073
LR	β_7	3.110
Duration of Diagnosis	d_{CHD}	1.078

Description	Parameter	Value
Intercept	$q_{0,\text{S}}$	0.00186
Age	δ_1	1.092
Sex	δ_2	0.700
Smoking Status	δ_3	1.547
SBP	δ_4	1.060
LR	δ_5	1.111
Duration of Diagnosis	d_{S}	1.145

$$b_{t,1}(\xi_t, \boldsymbol{m}_t) = \exp\left\{ - q_{\text{CHD}} d_{\text{CHD}}^t \left(\frac{1 - d_{\text{CHD}}^\tau}{1 - d_{\text{CHD}}} \right) \right\}, \text{ and} \tag{4.37}$$

$$b_{t,2}(\xi_t, \boldsymbol{m}_t) = \exp\left\{ - q_{\text{S}} d_{\text{S}}^t \left(\frac{1 - d_{\text{S}}^\tau}{1 - d_{\text{S}}} \right) \right\}. \tag{4.38}$$

The quantities q_{CHD} and q_{S} are defined by the following:

$$q_{\text{CHD}} = q_{0,\text{CHD}} \beta_1^{\text{AGE}-55} \beta_2^{\text{SEX}} \beta_3^{\text{AC}} \beta_4^{\text{SMOK}} \beta_5^{H-6.72} \beta_6^{(\text{SBP}-135.7)/10} \beta_7^{\ln(\text{LR})-1.59}, \text{ and} \tag{4.39}$$

$$q_{\text{S}} = q_{0,\text{S}} \delta_1^{\text{AGE}-55} \delta_2^{\text{SEX}} \delta_3^{\text{SMOK}} \delta_4^{(\text{SBP}-135.5)/10} \delta_5^{\text{LR}-5.11}. \tag{4.40}$$

Table 4.1 provides the parameter values for the basis functions. The exponents in Equations (4.39) and (4.40) are defined as the following: AGE = t + the age of the patient at the beginning of the model, SEX = 0 for males and 1 for females, AC = 0 for Caucasians or Asian-Indians and 1 otherwise, SMOK = 0 if the patient is a nonsmoker and 1 otherwise, H = the patient's

HbA1c, SBP = ℓ_t^{SBP}, and LR = ℓ_t^{LR}. Note that ℓ_t^{SBP} and ℓ_t^{LR} incorporate the patient's current medication state $\boldsymbol{m}_t$.

Additional basis functions that are used are the following:

$$b_{t,3}(\xi_t, \boldsymbol{m}_t) = \ell_t^{\text{LR}}, \tag{4.41}$$

$$b_{t,4}(\xi_t, \boldsymbol{m}_t) = \ell_t^{\text{SBP}}, \text{ and} \tag{4.42}$$

$$b_{t,5}(\xi_t, \boldsymbol{m}_t) = 1. \tag{4.43}$$

Equations (4.41) and (4.42) are linear basis functions in terms of the decision variables. Linear basis functions are commonly used [6], however use of these functions may not always provide a good set of basis functions [79]. We also use the unit function presented in Equation (4.43). When the basis function weights for the unit function depend on t, $\tilde{\xi}_t$, and $\boldsymbol{m}_t$, the solution of the weights for the unit function is identical to the solution for $v_t(\tilde{\xi}_t, \boldsymbol{m}_t)$ for all t, $\tilde{\xi}_t$, and $\boldsymbol{m}_t$ for the finite-state MDP. Thus, the finite-state MDP is a special case of basis function approximation.

Other candidates for basis functions that have been used in other studies include Legendre polynomials, radial basis functions, and wavelets. The use of these functions will provide an exact representation of the value function as the number of basis functions used goes to infinity.

With the finite-state MDP policies, the same action $\widehat{\boldsymbol{a}}^*(\xi_t, \boldsymbol{m}_t)$ is taken for every ℓ_t^{LR} and ℓ_t^{SBP} within the finite states according to Equation (4.23). However, the optimal action found using basis function approximation methods may differ within these finite states. In this way the basis function approach allows for flexibility in the actions when the continuous state space is considered since the basis functions are continuous in terms of the patient's LR and SBP. Due to this fact, there is potential for greater rewards to be achieved with basis function approximation policies.

4.5 Monte Carlo Simulation

The policies from each ADP method are compared using simulation with continuous ranges for the patient's LR and SBP values. The patient's exact LR, SBP, and other details are plugged into the basis functions, and the basis functions are multiplied by the appropriate weights. For a given decision epoch, the action that provides the largest basis function approximation of the value function for the patient will be taken.

For the purpose of the numerical experiments we choose to consider the objective of maximizing QALYs before the patient's first stroke or CHD event or death. The rewards for this primary prevention objective are given by the following:

$$r(\xi_t, \boldsymbol{m}_t) = \begin{cases} q(\boldsymbol{m}_t) & \text{if } \ell_t^{\text{CHD}} = 0 \text{ and } \ell_t^{\text{S}} = 0, \\ 0 & \text{otherwise.} \end{cases} \tag{4.44}$$

The policies produced from the ADP approaches are compared using a Monte Carlo simulation model of the treatment of individual patients over the course of their lifetime. Patients are simulated from age 40 until death. Several patient statistics are gathered, including their value function estimate ($v_0(\xi_0, \boldsymbol{m}_0) = \sum_{t=0}^{T^{\max}} r(\xi_t, \boldsymbol{m}_t)$, where $T^{\max}$ is the patient's maximum age), estimated lifetime QALYs, number of medications initiated over their lifetime, number of CHD events, number of strokes, lifetime costs of medication and treatment, and age at the time of death. These summary statistics can be used to compare the relative performance of the ADP approaches. However, the main statistic we are interested in is the patient's QALYs before the patient's first event or death.

At age 40, the patient begins with initial LR and SBP values from one of the possible $q_{\text{LR}} \times q_{\text{SBP}}$ states for age 40. For the numerical experiments presented, the LR interval $[0, \text{LR}^{\max} = \infty)$ is divided into $q_{\text{LR}} = 3$ discrete states with $\text{UB}(S_1^{\text{LR}}) = 4$ and $\text{UB}(S_2^{\text{LR}}) = 6$. The SBP interval $[0, \text{SBP}^{\max} = \infty)$ is divided into $q_{\text{SBP}} = 4$ discrete states with $\text{UB}(S_1^{\text{SBP}}) = 120$ mmHg, $\text{UB}(S_2^{\text{SBP}}) = 140$ mmHg, and $\text{UB}(S_3^{\text{SBP}}) = 160$ mmHg. The possible initial values of LR and SBP are $\ell_0^{\text{LR}} = \eta(S_i^{\text{LR}})$ and $\ell_0^{\text{SBP}} = \eta(S_j^{\text{SBP}})$, for all $i = 1, \ldots, q_{\text{LR}}, j = 1, \ldots, q_{\text{SBP}}$. The actual

values used for $\eta(S_i^{\text{LR}})$ and $\eta(S_j^{\text{SBP}})$ were estimated from the LR and SBP values of 40-year-old patients from the Mayo cohort. From time t to $t+1$, there are seven possible patient outcomes related to events and death from other causes: no stroke and no CHD event, no stroke and a nonfatal CHD event, a nonfatal stroke and no CHD event, a fatal stroke, a fatal CHD event, and death from other causes. Each patient has probabilities of these events occurring at each epoch defined according to the UKPDS risk equations. The inputs for these risk equations are based on the patient's state at time t, including his or her LR and SBP. Of the seven possible outcomes, the outcome that occurs is determined by a random number generated from a uniform distribution on the interval $[0, 1]$. If the patient does not die, the patient's health states are updated according to Equations (4.2) and (4.3) where z_t^{LR} and z_t^{SBP} are random numbers generated from $f_{\text{LR},t}(\cdot)$ and $f_{\text{SBP},t}(\cdot)$, respectively, and the medication effects on LR and SBP, $y_t^{\text{LR}}(\boldsymbol{a}(\xi_t, \boldsymbol{m}_t))$ and $y_t^{\text{SBP}}(\boldsymbol{a}(\xi_t, \boldsymbol{m}_t))$, are determined based on the optimal actions defined by the ADP approach being considered.

The distributions for the yearly percentage change in LR and the yearly percentage change in SBP were estimated from patients in the Mayo cohort defined in Section 3.4.1. The LR estimate was found using the yearly percentage changes in LR for 415 patients on no cholesterol medications, and the SBP estimate was found using the yearly percentage changes in SBP for 323 patients on no blood pressure medications. Multiple estimates were used for patients that had more than two years of readings while they were not on the specified medications. The percentage change in LR was found by calculating the percentage change in the average of all spline-fit LR readings for year t and the average of all spline-fit LR readings for year $t+1$. The description of the spline fitting method used can be found in Denton et al. [30]. The yearly percentage change values for SBP were found in an analogous way. We fit distributions to the percentage change values using Arena Input Analyzer [59]. The percentage change in LR was estimated to be normally distributed with mean -0.0106 and standard deviation 0.113. The percentage change in SBP was estimated to be normally distributed with mean 0.000169 and standard deviation 0.068. Since it is not reasonable for the percentage change in a patient's

Table 4.2: Comparison among the ADP methods and no treatment of expected QALYs before a stroke, CHD event, or death from other causes for males and females. For each simulation, 120,000 patients are sampled. The 95% confidence intervals for simulated results are provided in parentheses.

	Males	Females
Finite-State MDP (ADP 1)		
MDP Results	69.25	75.11
Simulation of MDP Policy	69.57 (69.34, 69.80)	75.52 (75.27, 75.78)
Basis Function Approximation (ADP 2)		
Unit	69.57 (69.34, 69.80)	75.52 (75.27, 75.78)
LR	69.59 (69.36, 69.83)	75.52 (75.27, 75.78)
SBP	69.57 (69.34, 69.80)	75.55 (75.30, 75.80)
LR and SBP	69.59 (69.36, 69.83)	75.52 (75.26, 75.77)
CHD	69.56 (69.33, 69.79)	75.53 (75.28, 75.78)
Stroke	69.57 (69.34, 69.80)	75.52 (75.27, 75.78)
CHD and Stroke	69.57 (69.33, 69.80)	75.53 (75.28, 75.78)
Patient Baseline		
No Treatment	67.86 (67.63, 68.09)	73.85 (73.60, 74.10)
U.S. I	69.46 (69.23, 69.69)	75.43 (75.18, 75.68)

LR or SBP to be unbounded, in the simulation the values are truncated at a maximum $\pm 50\%$ change. For LR, the probability of the percentage change falling outside of $[-50\%, 50\%]$ is less than 1.5×10^{-5}. For SBP, the probability of the percentage change falling outside of $[-50\%, 50\%]$ is less than 2×10^{-13}.

Patients that are alive at the end of the decision horizon are assumed to remain on the medication used at time T for the rest of their lives. The expected future rewards past the decision horizon are estimated by Equation (4.20).

4.6 Results

We present results for the case of maximizing QALYs before an event for the two ADP approaches (ADP 1 and ADP 2). The expected QALYs before an event or death for males and females using the simulation model with base case parameters are presented in Table 4.2. For

each simulation we use 120,000 sample paths (patients), and we provide 95% confidence intervals for each result. For ADP 1, we present expected event-free QALYs from solving the finite-state MDP as well as the result of simulating the mapping of the finite-state policy to the continuous state space as defined in Equation (4.23). For ADP 2, we present results for using the following basis functions:

- Unit: $b_{t,5}(\xi_t, \boldsymbol{m}_t)$ (Equation (4.43))
- LR: $b_{t,3}(\xi_t, \boldsymbol{m}_t)$ (Equation (4.41))
- SBP: $b_{t,4}(\xi_t, \boldsymbol{m}_t)$ (Equation (4.42))
- LR and SBP: $b_{t,3}(\xi_t, \boldsymbol{m}_t)$ and $b_{t,4}(\xi_t, \boldsymbol{m}_t)$ (Equations (4.41) and (4.42))
- CHD: $b_{t,1}(\xi_t, \boldsymbol{m}_t)$ (Equation (4.37))
- Stroke: $b_{t,2}(\xi_t, \boldsymbol{m}_t)$ (Equation (4.38))
- CHD and Stroke: $b_{t,1}(\xi_t, \boldsymbol{m}_t)$ and $b_{t,2}(\xi_t, \boldsymbol{m}_t)$ (Equations (4.37) and (4.38))

We find that the expected QALYs found through simulation of the finite-state MDP policy are significantly greater than the value function estimate found through solving the finite-state MDP using backward induction. This difference is slightly more pronounced for females. The basis function approximations do not provide significantly different results using the base case parameters, as shown in Table 4.2. All of the basis functions tested perform approximately the same. While the relative results are very similar for males and females, the difference between the results for males and females for each policy is approximately 6 QALYs.

We also simulated the the U.S. guidelines. Since U.S. I (see Table 3.1) depends on a patient's LDL and SBP, we use age-dependent estimates of TC and triglycerides (TG) estimated from the Mayo cohort in order to calculate a patient's LDL from their LR using Friedewald's equation [40]. According to this equation, TC is estimated by the following: TC = HDL + LDL + $k\times$ TG, where k is 0.20 when the cholesterol estimates are measured in mg/dL. In the simulation,

Table 4.3: Sensitivity analysis results for base case probabilities and 50% higher medication decrements to QALYs.

	Males	Females
Finite-State MDP (ADP 1)		
Simulation of MDP Policy	69.31 (69.08, 69.54)	75.25 (75.00, 75.50)
Basis Function Approximation (ADP 2)		
LR	69.31 (69.08, 69.54)	75.25 (75.00, 75.50)
SBP	69.29 (69.06, 69.52)	75.25 (75.00, 75.50)
LR and SBP	69.31 (69.08, 69.54)	75.24 (74.99, 75.49)
CHD	69.32 (69.08, 69.55)	75.25 (75.00, 75.50)
Stroke	69.31 (69.08, 69.54)	75.25 (75.00, 75.50)
CHD and Stroke	69.32 (69.09, 69.55)	75.26 (75.01, 75.51)
Patient Baseline		
No Treatment	67.86 (67.63, 68.09)	73.85 (73.60, 74.10)
U.S. I	69.20 (68.97, 69.42)	75.10 (74.86, 75.35)

we estimate the patient's LDL at time t by $\text{TC} - \frac{\text{TC}}{\ell_t^{\text{LR}}} - 0.20 \times \text{TG}$. From Table 4.2 we see that the mean expected QALYs from using U.S. I is slightly less than the mean expected QALYs found from using the policies generated from the basis function approximations. However, none of the basis function approximation policies result in significantly higher QALYs. This result is not very surprising since the expected given the results in Chapter 3. Although Figures 3.6 and 3.7 consider QALYs over a patient's entire lifetime, maximum QALYs using optimal treatment only results in slightly higher QALYs than U.S. I.

4.6.1 Sensitivity Analysis

Tables 4.3 through 4.11 present the results of sensitivity analysis on the probability of events and the decrements to QALYs due to medication costs. Note that the results for the unit basis function approximation are not included in these tables since these results are identical to the results from simulating the finite-state MDP policy. In Tables 4.3 through 4.10, probabilities of events range from 25% lower than the base case probabilities to 25% higher than the base case probabilities. The medication decrements to QALYs ($d^{\text{MED}}(\boldsymbol{m}_t)$ where $q(\boldsymbol{m}_t) = 1 - d^{\text{MED}}(\boldsymbol{m}_t)$)

Table 4.4: Sensitivity analysis results for base case probabilities and 50% lower medication decrements to QALYs.

	Males	Females
Finite-State MDP (ADP 1)		
Simulation of MDP Policy	69.90 (69.67, 70.13)	75.85 (75.60, 76.11)
Basis Function Approximation (ADP 2)		
LR	69.90 (69.67, 70.13)	75.85 (75.60, 76.11)
SBP	69.93 (69.70, 70.16)	75.92 (75.67, 76.17)
LR and SBP	69.90 (69.67, 70.14)	75.85 (75.60, 76.11)
CHD	69.90 (69.67, 70.14)	75.85 (75.60, 76.10)
Stroke	69.90 (69.67, 70.13)	75.85 (75.60, 76.10)
CHD and Stroke	69.91 (69.67, 70.14)	75.85 (75.60, 76.11)
Patient Baseline		
No Treatment	67.86 (67.63, 68.09)	73.85 (73.60, 74.10)
U.S. I	69.72 (69.49, 69.95)	75.75 (75.50, 76.00)

Table 4.5: Sensitivity analysis results for 25% higher probabilities and 50% higher medication decrements to QALYs.

	Males	Females
Finite-State MDP (ADP 1)		
Simulation of MDP Policy	67.54 (67.32, 67.76)	73.23 (72.99, 73.47)
Basis Function Approximation (ADP 2)		
LR	67.54 (67.32, 67.76)	73.23 (72.99, 73.47)
SBP	67.52 (67.30, 67.74)	73.23 (72.99, 73.47)
LR and SBP	67.54 (67.32, 67.77)	73.23 (72.99, 73.47)
CHD	67.54 (67.32, 67.76)	73.23 (72.99, 73.47)
Stroke	67.54 (67.32, 67.76)	73.23 (72.99, 73.47)
CHD and Stroke	67.54 (67.31, 67.76)	73.24 (73.00, 73.48)
Patient Baseline		
No Treatment	66.03 (65.81, 66.26)	71.79 (71.55, 72.03)
U.S. I	67.43 (67.21, 67.65)	73.09 (72.85, 73.33)

Table 4.6: Sensitivity analysis results for 25% higher probabilities and base case medication decrements to QALYs.

	Males	Females
Finite-State MDP (ADP 1)		
Simulation of MDP Policy	67.80 (67.58, 68.02)	73.49 (73.25, 73.73)
Basis Function Approximation (ADP 2)		
LR	67.80 (67.58, 68.02)	73.49 (73.25, 73.73)
SBP	67.81 (67.58, 68.03)	73.51 (73.27, 73.75)
LR and SBP	67.80 (67.58, 68.03)	73.49 (73.25, 73.73)
CHD	67.80 (67.58, 68.02)	73.49 (73.25, 73.73)
Stroke	67.80 (67.58, 68.02)	73.49 (73.25, 73.73)
CHD and Stroke	67.80 (67.57, 68.02)	73.49 (73.24, 73.73)
Patient Baseline		
No Treatment	66.03 (65.81, 66.26)	71.79 (71.55, 72.03)
U.S. I	67.67 (67.45, 67.90)	73.39 (73.16, 73.63)

Table 4.7: Sensitivity analysis results for 25% higher probabilities and 50% lower medication decrements to QALYs.

	Males	Females
Finite-State MDP (ADP 1)		
Simulation of MDP Policy	68.15 (67.92, 68.37)	73.82 (73.58, 74.07)
Basis Function Approximation (ADP 2)		
LR	68.15 (67.92, 68.37)	73.82 (73.58, 74.07)
SBP	68.17 (67.94, 68.39)	73.89 (73.65, 74.14)
LR and SBP	68.15 (67.92, 68.37)	73.82 (73.58, 74.06)
CHD	68.14 (67.92, 68.37)	73.82 (73.58, 74.07)
Stroke	68.15 (67.92, 68.37)	73.82 (73.58, 74.06)
CHD and Stroke	68.14 (67.92, 68.36)	73.82 (73.58, 74.06)
Patient Baseline		
No Treatment	66.03 (65.81, 66.26)	71.79 (71.55, 72.03)
U.S. I	67.92 (67.70, 68.14)	73.70 (73.46, 73.94)

Table 4.8: Sensitivity analysis results for 25% lower probabilities and 50% higher medication decrements to QALYs.

	Males	Females
Finite-State MDP (ADP 1)		
Simulation of MDP Policy	71.58 (71.34, 71.82)	77.90 (77.64, 78.17)
Basis Function Approximation (ADP 2)		
LR	71.58 (71.34, 71.82)	77.90 (77.64, 78.17)
SBP	71.57 (71.33, 71.81)	77.91 (77.65, 78.17)
LR and SBP	71.59 (71.35, 71.83)	77.90 (77.64, 78.16)
CHD	71.59 (71.35, 71.83)	77.91 (77.65, 78.17)
Stroke	71.58 (71.34, 71.82)	77.91 (77.65, 78.17)
CHD and Stroke	71.59 (71.35, 71.83)	77.91 (77.65, 78.17)
Patient Baseline		
No Treatment	70.23 (69.99, 70.47)	76.57 (76.31, 76.83)
U.S. I	71.44 (71.21, 71.68)	77.74 (77.48, 78.00)

Table 4.9: Sensitivity analysis results for 25% lower probabilities and base case medication decrements to QALYs.

	Males	Females
Finite-State MDP (ADP 1)		
Simulation of MDP Policy	71.82 (71.58, 72.06)	78.15 (77.88, 78.41)
Basis Function Approximation (ADP 2)		
LR	71.82 (71.58, 72.06)	78.15 (77.88, 78.41)
SBP	71.83 (71.59, 72.07)	78.18 (77.92, 78.45)
LR and SBP	71.82 (71.58, 72.07)	78.14 (77.88, 78.41)
CHD	71.82 (71.58, 72.06)	78.15 (77.89, 78.42)
Stroke	71.82 (71.58, 72.06)	78.15 (77.88, 78.41)
CHD and Stroke	71.82 (71.58, 72.06)	78.16 (77.89, 78.42)
Patient Baseline		
No Treatment	70.23 (69.99, 70.47)	76.57 (76.31, 76.83)
U.S. I	71.73 (71.49, 71.97)	78.09 (77.83, 78.35)

Table 4.10: Sensitivity analysis results for 25% lower probabilities and 50% lower medication decrements to QALYs.

	Males	Females
Finite-State MDP (ADP 1)		
Simulation of MDP Policy	72.16 (71.91, 72.40)	78.45 (78.19, 78.72)
Basis Function Approximation (ADP 2)		
LR	72.16 (71.91, 72.40)	78.45 (78.19, 78.72)
SBP	72.20 (71.96, 72.44)	78.56 (78.30, 78.83)
LR and SBP	72.16 (71.91, 72.40)	78.44 (78.18, 78.71)
CHD	72.15 (71.91, 72.40)	78.46 (78.20, 78.73)
Stroke	72.15 (71.91, 72.40)	78.46 (78.19, 78.72)
CHD and Stroke	72.17 (71.93, 72.41)	78.48 (78.22, 78.75)
Patient Baseline		
No Treatment	70.23 (69.99, 70.47)	76.57 (76.31, 76.83)
U.S. I	72.02 (71.77, 72.26)	78.44 (78.18, 78.71)

Table 4.11: Sensitivity analysis results for base case probabilities and 5 times higher medication decrements to QALYs.

	Males	Females
Finite-State MDP (ADP 1)		
Simulation of MDP Policy	68.43 (68.20, 68.66)	74.26 (74.02, 74.50)
Basis Function Approximation (ADP 2)		
LR	68.35 (68.12, 68.57)	74.19 (73.94, 74.43)
SBP	68.42 (68.19, 68.64)	74.25 (74.01, 74.50)
LR and SBP	68.33 (68.11, 68.56)	74.14 (73.90, 74.38)
CHD	68.45 (68.23, 68.68)	74.29 (74.05, 74.54)
Stroke	68.44 (68.22, 68.67)	74.29 (74.04, 74.53)
CHD and Stroke	68.45 (68.22, 68.67)	74.30 (74.05, 74.54)
Patient Baseline		
No Treatment	67.86 (67.63, 68.09)	73.85 (73.60, 74.10)
U.S. I	67.35 (67.14, 67.57)	72.84 (72.61, 73.07)

range from 50% lower than the base case medication decrements to 50% higher than the base case case medication decrements.

Overall we see that the expected QALYs are sensitive to event probabilities and QALY decrements, but the relative values of the expected QALYs from using ADP 1, the basis function approximations, and U.S. I are very similar. The difference between QALYs under no treatment and QALYs from the other policies is dependent on the probability and QALY decrement values, though the policies always result in significantly higher QALYs over no treatment.

From the results in Tables 4.3 through 4.10 there are not large differences in the mean values among the ADP methods. However, the mean for the SBP basis function policy is slightly higher than the other means when lower medication decrements to QALYs are considered.

We also consider the case of medication QALYs being five times the base case values in Table 4.11. When medication decrements to QALYs are assumed to be this large, the LR basis function policy and the LR and SBP basis function policy means are not as high as the means for the other BF policies. It may be that LR is not as important of a feature as SBP. In addition the MDP policy and all basis function policies perform significantly better than U.S. I.

4.7 Conclusions

We have presented the continuous-state MDP that represents the true underlying problem for the discrete-state MDP presented in Chapter 3. The approximation of this continuous-state MDP that we presented in Chapter 3 may not be the best approximation, so we explored other methods for solving this continuous-state problem. In particular we focused on the use of basis function approximation of the value function.

Based on the numerical experiments provided, the expected event-free QALYs do not appear to be very sensitive to the ADP method used. In addition, there was no significant difference in expected QALYs before an event between each of the basis function results and the results for U.S. I. This observation holds true for the base case results as well as the results under sensitivity analysis on the probability of events and the QALY decrements for medication use except when

very large medication QALY decrements are considered. In summary, for the objective function and approximations considered, it appears it is not possible to achieve significantly greater QALYs than the current U.S. guidelines. It is possible that we find these results because of the assumptions we made about a patient's TC and TG in the simulation model in order to implement U.S. I. It is also possible that we could see more significant differences between the results of the ADP methods and U.S. I if additional medications were considered. Use of other methods for solving the ALP, such as constraint sampling and column generation, may provide more insight into these results.

Chapter 5

Using Electronic Health Records to Monitor and Improve Adherence to Medication

5.1 Introduction

Poor medication adherence has been estimated to cost approximately $100 billion per year in preventable hospitalizations in the United States alone [73]. Recent studies show that while improving adherence results in an increase in medication costs, there are significant overall cost savings, particularly among patients with chronic diseases [95, 51]. Improved adherence can also reduce the risk of adverse events and improve the quality and length of life for many patients. In spite of the benefits of high adherence, poor adherence is recognized as a major challenge in the medical community [26]. In 2007 the National Institutes of Health (NIH) implemented the *Adherence Research Network* to promote research on adherence [71]. The initiative supports 14 institutes and centers across NIH, highlights NIH funding for adherence research, synthesizes current scientific findings on adherence, and provides leadership on future research directions.

While it is difficult to directly measure the medication taken by patients, there are widely

accepted proxy measures of adherence, including patient self reporting, electronic medication monitors on pill canisters, and rates of prescription refills calculated from electronic health records (EHRs). Based on prescription refill estimates of adherence, studies suggest that only 25% of patients remain highly adherent to common treatments such as cholesterol-lowering medication [14, 66]. Adherence-improving interventions, such as collaborative decision making and the use of decision aids to choose medications, have been shown to improve adherence [113]. However, barriers to such interventions include the perception that they take time and effort and are often not reimbursed by third-party payers. Furthermore, information about an individual patient's adherence to their prescribed medications is normally unavailable to physicians at the time of encounter with a patient.

Recently, considerable attention has been given to the use of EHRs to improve efficiency and effectiveness of health care delivery. EHRs are systematic collections of patient health information that can aid physicians in making medical decisions. In the United States, the Centers for Medicare and Medicaid Services (CMS) have recently introduced a *Meaningful Use* initiative [108]. The goals of the initiative are to improve safety and efficiency of health care delivery through the use of EHRs, and there is over $20 billion available from the Health Information Technology for Economic and Clinical Health Act (HITECH Act) to promote the adoption of information technology for health care and train skilled workers in this field. Due to incentives created by this program, health care managers are under pressure to meet the objectives of the Meaningful Use initiative and to submit clinical quality measures (CQMs) using certified EHR technology.

EHRs have the potential to enable monitoring of adherence and to identify patients that would benefit most from an adherence-improving intervention. By actively monitoring patients' adherence to medications, which we refer to as *active adherence surveillance* (AAS), such decisions could be made in real time at the point of care. However, implementation of a surveillance system comes at a cost. Therefore in this chapter we aim to answer the following research question: What are the potential benefits of using EHRs to improve adherence to medication? To

answer this question, we use pharmacy claims data for a large population to estimate patient adherence levels to the most commonly-used medication for cholesterol control. We present an MDP model to determine the optimal timing of adherence-improving interventions based on AAS of individual patients' adherence using EHRs. Our model considers both the perspective of the patient, who stands to benefit from the prevention of adverse health events related to poor adherence, and the perspective of the third-party payer (health insurer) that bears the burden of the cost of interventions, medication, hospitalizations, and follow-up care for adverse events related to poor adherence. We present structural properties of our model, including conditions under which a control limit policy exists, and how the control limit policy changes based on a patient's health status and the effectiveness of an intervention.

There are many prescription medications for which poor adherence is recognized as a challenge in preventing the onset or progression of disease (e.g., blood pressure control medications, asthma medications). In this chapter we provide a specific example based on adherence to *statins*, the most common cholesterol-lowering medication. We evaluate the costs and benefits associated with AAS by using our MDP model to determine the following: (a) the expected QALYs before a stroke, a CHD event (such as a heart attack), or death; and (b) medication and intervention costs and costs associated with the occurrence of strokes and CHD events (the most significant outcomes associated with cholesterol control). To estimate the marginal benefits of implementing the EHR-based system, we compare AAS to a much simpler, and easier to implement, schedule of interventions at regularly-spaced intervals (e.g., yearly interventions) which we refer to as *inactive adherence surveillance* (IAS). We also compare our results to outcomes for patients who receive no interventions. In addition, we estimate the potential yearly benefits of applying AAS to the U.S. population.

Our findings have the potential to influence several different stakeholders. First, our findings will help inform CMS about the potential benefits of AAS, and whether such implementations should be added to the list of objectives for their Meaningful Use or other future initiatives. Understanding and improving medication adherence is a natural extension to the current Mean-

ingful Use requirement of *medical reconciliation*, which requires an accurate list of medications the patient is currently taking. Second, our results will help inform third-party health insurers about the potential benefits of reimbursing health care providers for adherence-improving interventions. Third, physicians will benefit from an improved understanding of the relative benefits of addressing adherence to medications for chronic conditions. Finally, patients could directly benefit from improved quality of life and the lower costs that can be achieved by improved adherence.

The remainder of this chapter is organized as follows. In Section 5.2 we provide some background on adherence interventions and methods for estimating adherence from EHRs. We also provide a specific example that illustrates measurement of adherence to statins and its relationship to health outcomes. In Section 5.3 we present an overview of related literature in the areas of machine maintenance and medical decision making. In Section 5.4 we present our MDP model for determining the optimal time for interventions, and in Section 5.5 we explore some general insights that can be drawn from our model. In Section 5.6 we present a case study of cholesterol-lowering treatment to prevent cardiovascular disease in patients with type 2 diabetes. Finally, in Section 5.7 we provide concluding remarks and discuss future research opportunities.

5.2 Background on Medication Adherence

Motivation for understanding adherence to medication is summed up in a quote by C. Everett Koop, M.D.: "Drugs don't work in patients who don't take them." Osterberg and Blaschke [73] cite patient forgetfulness and lack of understanding as possible causes of poor adherence. The authors describe several types of interventions for improving medical adherence including patient education, increased access to medical care, and improved communication between patients and physicians. For example, performing screening tests and reviewing a patient's risk of an adverse health event (e.g., 10-year risk of a stroke or CHD event), or educating a patient about the risk reduction associated with a particular medication, has been shown to improve

patient adherence [113].

A common method for measuring patient adherence is to observe the percentage of days covered (PDC) by prescription refills over time. Prescription refills can be observed from pharmacy claims data, a portion of administrative claims data generated as a result of a patient's encounter with the health system. Claims data is an important part of the extended EHR that is collected by third-party payers for payment purposes. If Meaningful Use program objectives are met, more than 80% of patients will have pharmacy refills recorded as structured data by the end of 2012.

The standard formula for PDC is as follows [17]:

$$\text{PDC} = 100 \times \left(\frac{\text{days with an available supply of medication in the time period}}{\text{days in time period}} \right)\%. \quad (5.1)$$

Figure 5.1 provides an example of a patient's pharmacy claims for which PDC is estimated over a one-year period. In this example, the patient begins taking the medication with a 30-day supply. The patient makes four refills, each with 90-day supply, during the year. Gaps between the end of the days' supply for one prescription fill and the beginning of the next fill are interpreted as gaps in the patient's adherence to the medication. As shown in Figure 5.1, refills that have supply exceeding the amount of time to the end of the year (time period) are carried over to the calculation of the PDC for the next time period. Note that this method for computing PDC is not restricted by the days' supply of refills or the refill method (by mail or local pharmacy).

Combining pharmacy claims data with laboratory data (e.g., cholesterol, blood sugar, blood pressure) and other sources of data in the EHR is often necessary to measure the effects of adherence. For example, the PDC can be linked with the patient's percentage change in metabolic values over the same time period. We illustrate this with a specific example. Consider the case of patients initiating statins to lower their cholesterol and therefore lower their risk of stroke and CHD events. States for the PDC over the course of a year after initiation are defined by

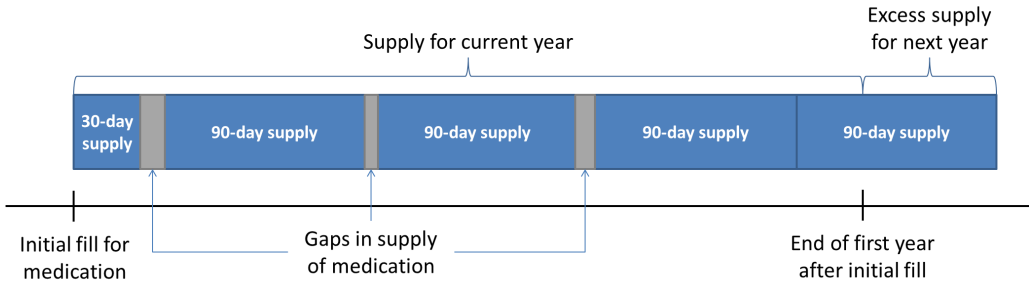

Figure 5.1: Diagram of Prescription Refills Used to Calculate the Percentage of Days Covered (PDC).

the four categories given in Table 5.1. The adherence states are defined as follows: NON ($0\% \leq$ PDC $\leq 10\%$), LOW ($10\% <$ PDC $\leq 40\%$), MED ($40\% <$ PDC $\leq 80\%$), and HIGH ($80\% <$ PDC $\leq 100\%$). These specific choices of adherence states are based on those commonly used in the health services research literature (for example see [83]). By using laboratory data, these adherence states are linked with changes in TC from initiation to one year after initiation. Large data sets that combine pharmacy claims data with laboratory data for a large sample of patients can thus be used to estimate the expected change in TC for each PDC level.

The results in Table 5.1 are based on a study reported by Mason et al. [66]. Table 5.1 establishes the link between a patient's percentage change in TC and her adherence to medication. Since the risk of cardiovascular events is affected by TC, the patient's risk is also correlated with their adherence to the medication [61, 97]. For this reason interventions that improve adherence have the potential to reduce cardiovascular risk over time. A method to estimate a stochastic model for changes in PDC over time is revisited in Section 5.6.

5.3 Literature Review

The problem of finding the optimal time to perform an intervention to improve a patient's adherence to a medical treatment is analogous to problems studied in the machine maintenance literature. This has been an active area of research for over fifty years. While we do not attempt to fully review this literature, we highlight related articles. In addition to the literature

Table 5.1: Adherence States Defined by Percentage of Days Covered (PDC) and the Corresponding Percent Change in Total Cholesterol (TC) for Patients that Initiate Statins.

Adherence State	PDC	Change in TC
NON	0 – 10%	−5.22%
LOW	10 – 40%	−8.21%
MED	40 – 80%	−18.08%
HIGH	80 – 100%	−25.25%

on machine maintenance, we also review articles on the use of the operations research (OR) models and methods for medical decision making.

5.3.1 Machine Maintenance Applications

Pioneering work on maintenance systems was done by Klein [60], who considers a stochastically deteriorating system that could be replaced or kept after inspection by a manager. If the system was kept, the decision would then be to repair the system or decide the time of the next inspection. The author assumes inspection gives the manager enough information to determine the state of the system. The objective of the model is to minimize the long run average cost of the policy. The model uses a Markov chain to represent probabilistic deterioration of the system. The main difference between Klein's model and previous machine maintenance models is that the time between successive transitions (inspections) is under the control of the decision maker (manager). This is similar to the model we present in which a decision maker must choose the optimal time for an adherence-improving intervention. The assumption that the manager gains knowledge of the machine's state through inspection is analogous to a physician gaining knowledge about the patient's adherence behavior through an office visit.

In [68], [78], [91], [56], [109], and [76], the authors survey maintenance policies for stochastically failing machines and imperfect repair. The latter in particular is consistent with adherence interventions that have an uncertain outcome, which we consider in this chapter. In [16] the authors consider a *hazardous inspection* model in which, similar to adherence intervention, there

is potential harm from inspection (e.g., for adherence intervention this could include a monetary cost of intervention or a loss of utility on behalf of the patient). Nakagawa [69] extends imperfect repair models to a system degrading over time as it ages, similar to the increasing probability of death from other causes in the model we present. In [11] the authors extend deterioration models to the case of preemptive maintenance. This is analogous to the goal of improving a patient's adherence to reduce the likelihood of future adverse events. All these studies have similarities to adherence control (albeit in a much different context), but none combine all the characteristics of our model.

Other notable references include the work of Anderson [9], which presents three continuous-time MDPs, each with an infinite horizon, continuous state space, and actions for maintenance or replacement of the machine. The models differ in terms of the rate of deterioration. The models are transformed into discrete-time finite-horizon MDPs to prove structural properties that provide insights on their continuous-time counterparts. Anderson provides conditions for each model for which a control limit structure exits and the preventative maintenance level is nonincreasing. Hopp and Wu [52] extend the work of Anderson on a machine maintenance model with preventative maintenance using an infinite-horizon MDP with a finite state space. They prove a control limit policy and monotonicity. In addition, Hopp and Wu consider the effects of alternate assumptions on the structural properties they prove for their model. For example, they show the structural properties still hold when the system must go down for an entire period when maintenance is performed.

5.3.2 Medical Decision Making Applications

There has been a recent increase in collaboration between the OR community and nonprofit organizations on the treatment and prevention of HIV/AIDS, including studies to improve adherence to treatment. The Population Council, an international nongovernmental organization, published a handbook on designing HIV/AIDS prevention studies using OR methods [38]. The handbook focuses on descriptive models, applying statistical tests and running cost-effectiveness

analysis. The Doris Duke Charitable Foundation has awarded grants recently for the use of OR methods on AIDS Care and Treatment in Africa (ORACTA) [31]. These grants include funding for studying effectiveness of interventions (e.g., HIV education, text message medication reminders, and home visits by peer educators). This work helps motivate the potential for prescriptive OR models, such as we discuss in this chapter, to improve adherence to medication for chronic diseases.

MDPs have been used in a number of medical applications for determining when a particular treatment should start or procedure should take place. For example, Alagoz et al. [7] consider the optimal timing of liver transplantation using a live donor in order to maximize the patient's total reward. The authors use an infinite-horizon MDP model to determine the optimal timing of this one-time decision. Structural properties are derived, including showing the existence of a control-limit policy under certain assumptions. Shechter et al. [90] also present an infinite-horizon MDP model to determine the optimal timing of HIV therapy. The states in the model represent the patient's CD4 count, and the objective is to maximize life years or QALYs over the patient's lifetime. Results suggest earlier treatment is optimal, contrary to treatment trends at the time of publication.

Maillart et al. [65] present a partially observable Markov chain model to evaluate various breast cancer screening policies considering implications of patient adherence to screening guidelines and differences in breast cancer incidence and aggression as women age. Evaluation, rather than optimization of policies, is used to selectively compare easy-to-implement policies. Efficient policies are identified based on the trade-off between lifetime breast mortality risk and the expected number of mammograms over a woman's lifetime. Chhatwal et al. [21] present a finite-horizon discrete-time MDP to determine the optimal timing of breast biopsy given the outcome of a mammogram and the patient's demographic features. The decision epochs are years after age 40, the states represent the patient's risk score, determined after a mammogram, and the actions are to have a biopsy or to have another mammogram the following year. Once the action of biopsy is taken, the patient leaves the decision process. Rewards are defined by

QALYs accrued by patients. Chhatwal et al. prove structural properties for their model, including the existence of a control-limit type policy. Results suggest that the decision to biopsy should depend on the patient's age.

Denton et al. [30] propose an MDP model to find the optimal time to initiate statins in patients with type 2 diabetes for the prevention of cardiovascular events. The states represent the patient's metabolic risk factors. The rewards are monetary rewards for QALYs minus costs of medication and treatment for cardiovascular events, and the action to initiate or defer initiation of treatment is revisited each year. The authors consider the effects of using different cardiovascular risk models to estimate the probability of adverse events, concluding that the risk model chosen can dramatically affect the optimal start times. Their model assumes perfect adherence to treatment. Mason et al. [66] propose a related MDP model to find the optimal time to initiate statins given the possibility of imperfect adherence. The authors incorporate a Markov model for adherence after the patient begins statins. The authors conclude that timing of initiation does not have as great of an effect on patient outcomes as improving adherence; however, they note that adherence-improving interventions can be costly. This study provides motivation for the study of the optimal time of adherence-improving interventions once treatment has begun.

To our knowledge, the problem of finding the optimal time to perform an intervention to improve medication adherence has not been studied before. This problem is analogous to problems studied in the machine maintenance literature; however, there are differences. First, in our model there is no available action to replace the system; only preventative maintenance may be performed. Second, we consider a system that is deteriorating in a nonstationary fashion over a finite horizon. Our model also differs in several ways from the literature on MDP models for medical decision making described above. First, the decision to initiate an adherence-improving intervention is a recurring decision and not a one-time decision as considered by [30, 7, 21], and [90]. Second, unlike the majority of the above models, which are infinite-horizon models, our model is finite horizon and nonstationary, to reflect nonstationarity in the risk of

adverse events with respect to a patient's age. Finally, and most notably, our study is unique in its specific application, in the methods and data used to estimate the model parameters, and in the research question we answer.

As a result of the differences between our model and those already in the literature, we make several contributions. We present new structural properties that provide insight into optimal adherence intervention policies, and motivate easy-to-implement rules of thumb for when to initiate an intervention. We use a large data set that combines pharmacy claims data with the laboratory data necessary to construct and solve the MDP model, and we present results based on this model for a specific example in the context of statin treatment for a population of patients at high risk of stroke and CHD events. To our knowledge, these results are the first estimates of the potential benefits that may be derived from active surveillance of patient adherence to medication.

5.4 Model Formulation

In each of a set of discrete decision epochs, a patient on a particular medication is observed to be in a specific health state. The health states are divided into *adherence states* and an *absorbing state*. The adherence states represent the patient's level of adherence to the medication (e.g., statins), and the absorbing state represents the occurrence of events that the treatment aims to prevent (e.g., a stroke or CHD event) or death from other causes. In each decision epoch, the decision maker (e.g., the physician) must decide whether or not to implement an intervention with the patient. Thus, one of two possible actions is taken: *implement an intervention* or *defer the decision until the next epoch*. This decision is faced at each decision epoch, provided the patient does not enter the absorbing state. The following is a detailed description of the MDP model.

Time Horizon: The decision to initiate an adherence-improving intervention is revisited periodically over a finite horizon with T yearly decision epochs. The decision epochs are indexed

by $t = 0, 1, 2, \ldots, T-1$ where time epoch t represents the time interval $[t, t+1)$. Time $t = 0$ represents the initial epoch when the patient begins surveillance (the patient begins taking the medication), and T is chosen as a reasonable upper bound on a typical patient's age (e.g., 100 years). For patients who have not entered the absorbing state at time T, a reward is obtained that estimates the benefits and costs associated with their future survival, based on an estimate of the patient's future remaining life years.

States: The states of the patient are represented by the set $S \equiv \{0, 1, 2, \ldots, M\}$; for each time $t = 0, \ldots, T$, we let $s_t \in S \setminus \{0\}$ denote the patient's adherence level for time epoch t, while $s_t = 0$ indicates that the patient had an adverse health event (fatal or nonfatal) or that the patient died from other causes. For $s_t \in S \setminus \{0\}$, a larger value of s_t corresponds to an increased (improved) level of adherence for the patient at time t.

Actions: An intervention may be initiated or deferred at any epoch, $t = 1, \ldots, T-1$, and in any state, $s_t \in S \setminus \{0\}$. The possible set of actions is defined as the following:

$$A_t(s_t) = \begin{cases} \{W, I\} & \text{for } s_t \in S \setminus \{0\} \text{ and } t = 1, \ldots, T-1, \\ \{W\} & \text{for } s_t = 0 \text{ or } t = T, \end{cases}$$

so that $a_t(s_t) \in A_t(s_t)$ denotes the action taken at time t when the patient is in state s_t, where the action $a_t = I$ denotes an intervention and the action $a_t = W$ denotes the action of waiting, or deferring the decision until the next epoch. The total action space is defined by $A = \{W, I\}$.

Transition Probabilities: There are two types of transition probabilities: (a) transitions between adherence states; and (b) transitions from adherence states to the absorbing state. Given avoidance of state 0, the conditional transition probabilities between the adherence states are represented by the matrix $\tilde{P}_t(a_t) \in \mathbb{R}^{M \times M}$ so that $[\tilde{P}_t(a_t)]_{i,j}$, the (i, j) element of $\tilde{P}_t(a_t)$, is equal to the conditional probability $\Pr\{s_{t+1} = j | s_t = i \text{ and } s_{t+1} \neq 0\}$ for $1 \leq i, j \leq M$. Transitions from adherence states to the absorbing state are represented by the vector $\bar{p}_t \in \mathbb{R}^M$ so that $[\bar{p}_t]_i$, the ith element of the $M \times 1$ (column) vector $\bar{p}_t$, is equal to the conditional probability

$\Pr\{s_{t+1} = 0 | s_t = i\}$ under action a_t for $1 \leq i \leq M$. The complete state transition probability matrix is

$$P_t(a_t) = \begin{bmatrix} 1 & \mathbf{0}_M^T \\ \bar{p}_t & \text{diag}[\mathbf{1}_M - \bar{p}_t]\tilde{P}_t(a_t) \end{bmatrix}, \tag{5.2}$$

where $\mathbf{0}_M$ is the $M \times 1$ (column) vector of zeros and $\mathbf{1}_M$ is the $M \times 1$ (column) vector of ones.

Rewards: There are many possible reward structures for our model depending on the decision maker's perspective. In this chapter we define rewards to be composed of four parts: (i) a reward for quality-adjusted time gained in the most recent period (e.g., a QALY for an annual decision epoch); (ii) a cost associated with an adherence intervention; (iii) a state-dependent cost of medication; and (iv) a penalty cost for entering the absorbing state. We define $r_t(s_t, a_t)$ to be the reward accrued in state s_t given action a_t is taken. For $t = 1, \ldots, T-1$, the reward function is defined as

$$r_t(s_t, a_t) = \begin{cases} R \times Q(s_t) - C^{\text{MED}}(s_t) & \text{for } a_t = W \text{ and } s_t = 1, \ldots, M, \\ R \times Q(s_t) - C^{\text{MED}}(s_t) - C^{\text{INT}} & \text{for } a_t = I \text{ and } s_t = 1, \ldots, M, \\ -C_t^{\text{F}} & \text{for } s_t = 0, \end{cases} \tag{5.3}$$

where R is the *willingness-to-pay* factor defining a monetary value per QALY and $Q(s_t)$ represents the QALYs for a patient in state s_t during time epoch t. The quantity $C^{\text{MED}}(s_t)$ denotes the cost of medication for time epoch t; this cost depends on the patient's adherence state since patients do not pay for medication they do not have in their possession. The quantity C^{INT} denotes the cost of an adherence-improving intervention.

The quantity C_t^{F} represents a one-time lump sum for the expected future costs of a patient entering the absorbing state 0. This cost penalty reflects a loss associated with failure to avoid an adverse health event. This loss could include costs associated with hospitalization and/or

future treatment. This one-time negative reward occurs upon entering the absorbing state, and all future rewards for patients in the absorbing state are zero. Thus, the expected future rewards, or value-to-go, for patients that have just entered the absorbing state is zero. The state space could be amended to add an additional state $s_t = 0'$, representing the transient state patients enter when an event or death occurs. Patients in state $s_t = 0'$ would then transition to state $s_{t+1} = 0$ at time $t + 1$. For this amended state space the immediate reward for being in state $s_t = 0'$ would be $-C_t^{\mathrm{F}}$ and the immediate reward for being in state $s_t = 0$ would be 0.

The reward structure presented above represents a combination of the patient objective of maximizing quality-adjusted time to first event (which is frequently the clinical intent of preventive treatment [24]) and the objective of minimizing costs of treatment, considering both costs before the patient enters the absorbing state and expected costs after the patient enters the absorbing state. Additional assumptions about the reward structure are provided in Section 5.5, and specific values for rewards are provided in Section 5.6 in the context of cardiovascular disease prevention.

For a patient in state $s_t \in S$ in epoch t, the optimality equations can be written as

$$v_t(s_t) = \max_{a_t \in A_t(s_t)} \left\{ r_t(s_t, a_t) + \lambda \sum_{s_{t+1} \in S} p_t(s_{t+1}|s_t, a_t) v_{t+1}(s_{t+1}) \right\}, \text{ for every } t = 1, \ldots, T-1, \tag{5.4}$$

where $p_t(s_{t+1}|s_t, a_t)$ is the (s_t, s_{t+1}) element of $P_t(a_t)$, $v_t(s_t)$ is the optimal value function, and $\lambda \in [0, 1)$ is the discount factor which calculates the time t value of rewards received at time $t + 1$. For every time $t \in \{0, 1, \ldots, T\}$, we define $\mu_t(s_t)$ to be the expected difference between the rewards for quality-adjusted survival benefits and the associated costs, assuming no future interventions, for every $s_t \in S$. We take $\mu_t(0) = -C_t^{\mathrm{F}}$ provided that $s_{t-1} \neq 0$, and $\mu_t(0) = 0$ otherwise. The end-of-horizon boundary condition is

$$v_T(s_T) = \mu_T(s_T), \text{ for every } s_T \in S. \tag{5.5}$$

The last decision epoch, T, is selected to represent a reasonable upper bound on the age at which adherence-improving interventions would no longer be advisable due to high competing risks of death from other causes. This end-of-horizon assumption has been made in a number of other medical decision making studies [30, 21, 63].

5.5 Model Properties and Insights

This section provides insights into the structure of our model. First, we discuss some of the assumptions of our model. Next, we present some properties of our model that can reduce the computational effort to solve the MDP, and that provide some insight into the optimal policy for interventions defined by our model. We prove the existence of an optimal control limit policy. Next, we present a theorem relating the effectiveness of interventions to the optimal control limits for the interventions. Finally, we present a theorem comparing the optimal control limits for two patients where one patient is at a greater risk for adverse health events than the other.

5.5.1 Model Assumptions

There are many possible choices for the reward function to use in our MDP model. We chose to blend two criteria for our reward function: the patient reward for quality-adjusted time to first event and the payer cost of treatment, intervention, and care associated with an adverse health event. We make the following assumptions about our model:

(i) $\tilde{P}_t(a_t)$ has the increasing failure rate (IFR) property for every $a_t \in A$, and for every $t = 1, \ldots, T-1$;

(ii) $\mu_t(s_t)$ is nondecreasing in s_t for $t = 1, \ldots, T$;

(iii) $[\bar{p}_t]_i \equiv \Pr\{s_{t+1} = 0 | s_t = i\}$ is nonincreasing in s_t for $t = 1, \ldots, T-1$; and

(iv) $R \times Q(s_t) - C^{\mathrm{MED}}(s_t)$ is a nondecreasing function of s_t for $t = 1, \ldots, T-1$.

Assumption (i) states that the Markov chain defining a patient's adherence exhibits the IFR property (see Barlow and Proschan [12] for a definition of this property). This can be interpreted to mean that the better a patient's adherence level the better it is likely to be in the future. Our study using observational data (see Section 5.6) suggests that this is a reasonable assumption. This property has also been observed for a number of other health characteristics ([7, 63, 21]). Assumption (ii) states that a patient's expected future rewards for QALYs minus costs, assuming no future interventions, does not decrease as their adherence improves. This assumption is reasonable since improved adherence causes treatment to be more effective at preventing adverse events. Assumption (iii) states that the probability of moving to the absorbing state is nonincreasing in the adherence state. Finally, assumption (iv) states that the difference between $R \times Q(s_t)$, the reward for living through a decision epoch, and $C^{\text{MED}}(s_t)$, the cost of medication, is a nondecreasing function of the adherence state s_t. In addition to the above assumptions, we assume that $R, Q(s_t), C^{\text{INT}}, C_t^{\text{F}}$, and $C^{\text{MED}}(s_t)$ are nonnegative for every value of s_t.

5.5.2 Model Properties

We now discuss some properties associated with the optimal adherence intervention policy and draw comparisons between different types of patients and interventions. We begin by presenting two lemmas that are used to prove our main results.

Assumption (i) stated that transitions among adherence states, defined by $\tilde{P}_t(a_t)$, are IFR. Lemma 5.5.1 extends this by stating that the complete transition probability matrix, $P_t(a_t)$, is IFR if the conditional transition matrix among the adherence states given avoidance of state 0 is IFR. We show that if the rows of $\tilde{P}_t(a_t)$ are in increasing stochastic order, then the rows of $P_t(a_t)$ must also be in increasing stochastic order.

Lemma 5.5.1. *If $\tilde{P}_t(a_t)$ is IFR and assumption (iii) holds, then $P_t(a_t)$ is IFR for $t = 1, \ldots, T-1$.*

Proof. Since $\tilde{P}_t(a_t)$ is IFR by assumption (i), with (i,j) element $\tilde{p}_t(j|i,a_t) \equiv [\tilde{P}_t(a_t)]_{i,j}$, it follows that for each $k \in \{1,\ldots,M\}$, the quantity

$$\tilde{q}_t(k|i,a_t) = \sum_{j=k}^{M} \tilde{p}_t(j|i,a_t) \tag{5.6}$$

is nondecreasing in i for $i = 1,\ldots,M$. The matrix multiplication $\text{diag}[\mathbf{1}_M - \bar{p}_t]\tilde{P}_t(a_t)$ involves multiplying the ith row of $\tilde{P}_t(a_t)$ through by $1 - [\bar{p}_t]_i$ for $i \in \{1,\ldots,M\}$. Therefore, since $1 - [\bar{p}_t]_i$ is nondecreasing in i by assumption (iii), it follows that the $(M+1) \times M$ matrix

$$Z \equiv \begin{bmatrix} \mathbf{0}_M^T \\ \text{diag}[\mathbf{1}_M - \bar{p}_t]\tilde{P}_t(a_t) \end{bmatrix} \tag{5.7}$$

with (u,v) element $Z_{u,v}$ for $u \in \{0,1,\ldots,M\}$ and $v \in \{1,\ldots,M\}$ satisfies the following IFR-like property: for each fixed $k \in \{1,\ldots,M\}$, the function

$$z(u) \equiv \sum_{v=k}^{M} Z_{u,v} \tag{5.8}$$

is nondecreasing in u for $u \in \{0,1,\ldots,M\}$. Finally, we see that $P_t(a_t)$ is IFR since $\sum_{j=0}^{M} p_t(j|i,a_t) = 1$, for every $i = 0,\ldots,M$. □

The next lemma states that the value function does not decrease as the patient's adherence improves.

Lemma 5.5.2. *The value function $v_t(s_t)$ is nondecreasing in s_t, for $t = 1,\ldots,T$.*

Proof. This follows from Proposition 4.7.3 of [81] and the fact that the following conditions hold:

1. The quantity $r_t(s_t,a_t)$ is nondecreasing in s_t, for every $a_t \in A_t(s_t)$ and for $t = 1,\ldots,T-1$, by assumption (iv) and Equation (5.3).

2. For each $k \in \{0, 1, \ldots, M\}$, the analogue of Equation (5.6) for the matrix $P_t(a_t)$,

$$q_t(k|s_t, a_t) \equiv \sum_{j=k}^{M} p_t(j|s_t, a_t) \tag{5.9}$$

is nondecreasing in s_t, for every $a_t \in A_t(s_t)$, by Lemma 5.5.1 for $t = 1, \ldots, T-1$.

3. The function $\mu_T(s_T)$ is nondecreasing in s_T by assumption (ii).

□

Lemma 5.5.2 shows that the patient's expected future rewards do not decrease as adherence to treatment improves. This fact is used to prove Theorem 5.5.1, which states that the optimal intervention policy has a simple control-limit structure for the adherence states $s_t = 1, \ldots, M$.

Theorem 5.5.1. *If the effect of an intervention is independent of the patient's current adherence state, then there exists an optimal control limit $s_t^* \in S \setminus \{0\}$, for every $t \in \{1, \ldots, T-1\}$, such that the optimal action $a_t^*(s_t)$ is given by*

$$a_t^*(s_t) = \left\{ \begin{array}{ll} I, & \text{if } s_t \leq s_t^*, \text{ for every } s_t \in S \setminus \{0\}, \\ W, & \text{otherwise,} \end{array} \right\} \text{for } t = 1, \ldots, T-1. \tag{5.10}$$

Proof. We prove the existence of the control-limit policy for the entire set S, where it is understood that the action for $s_t = 0$ is defined as W. If we associate, for example, the numerical value 0 with the action W and the numerical value 1 with the action I, then from Theorem 4.7.4 of Puterman (1994) it is sufficient to prove the following: (a) $r_t(s_t, a_t)$ is nondecreasing in s_t for all $a_t \in A_t(s_t)$; (b) $q_t(k|s_t, a_t)$ is nondecreasing in s_t for all $k \in S$ and $a_t \in A_t(s_t)$; (c) $r_t(s_t, a_t)$ is a subadditive function on $S \times A_t(s_t)$; (d) $q_t(k|s_t, a_t)$ is a subadditive function on $S \times A_t(s_t)$ for all $k \in S$; and (e) $\mu_T(s_T)$ is nondecreasing in s_T. Conditions (a), (b), and (e) follow by assumption (iv), Lemma 5.5.1, and assumption (ii), respectively. Condition (c) follows because $\big(r_t(s_t + 1, 1) - r_t(s_t + 1, 0)\big) - \big(r_t(s_t, 1) - r_t(s_t, 0)\big) \leq 0$, for every $s_t \in S$. To

prove condition (d) we must show

$$q_t(k|i+1,1) - q_t(k|i+1,0) \leq q_t(k|i,1) - q_t(k|i,0)$$

for every $k \in S$ and for every $s_t \in S$. In order for the above inequality to be true, the following must also hold

$$\sum_{j=k}^{M} p_t(j|i+1,1) - \sum_{j=k}^{M} p_t(j|i+1,0) \leq \sum_{j=k}^{M} p_t(j|i,1) - \sum_{j=k}^{M} p_t(j|i,0),$$

which can be expressed as

$$\sum_{j=k}^{M} p_t(j|i+1,1) + \sum_{j=k}^{M} p_t(j|i,0) \leq \sum_{j=k}^{M} p_t(j|i,1) + \sum_{j=k}^{M} p_t(j|i+1,0). \tag{5.11}$$

This inequality holds for the first terms on each side of Equation (5.11) by the assumption that the effect of an intervention is independent of the patient's current adherence state. The inequality holds for the second terms on each side of Equation (5.11) by Lemma 5.5.1. □

Theorem 5.5.1 provides sufficient conditions under which the optimal intervention policy has a simple structure, which is important for clinical applications in practice. This structure can also be exploited to achieve computational advantages in computing the optimal policy. This could be particularly relevant for applications involving real-time clinical intervention decisions.

The remainder of this section presents theorems based on the comparison of optimal policies for different types of interventions and different types of patients. We begin with a definition of stochastic dominance relevant to the two theorems.

Definition 5.5.1. *Given $t \in \{1, \ldots, T-1\}$, $s_t \in S$, and $a_t \in A_t(s_t)$, the one-step transition probability matrix $P_t^{(1)}(a_t)$ is said to stochastically dominate $P_t^{(2)}(a_t)$, denoted by $P_t^{(1)}(a_t) \succcurlyeq P_t^{(2)}(a_t)$, if*

$$\sum_{j=k}^{M} P_{(1)}t^{(}j|i,a_t) \geq \sum_{j=k}^{M} P_t^{(2)}(j|i,a_t), \text{ for every } i,k \in S.$$

In order to differentiate the control limits for two interventions, we introduce the following notation: $s_t^*(I)$ represents the optimal control limit for intervention I. In addition, we use a superscript to differentiate probabilities and value functions from the two MDPs in the following theorem.

Theorem 5.5.2. *Given an MDP with intervention I_1 and a second MDP with intervention I_2, if $v_t^{(1)}(s_t) - v_t^{(2)}(s_t)$ is nondecreasing in s_t for $t \in \{1, \ldots, T-1\}$ and the two MDPs are identical except $P_t^{(1)}(I_1) \succcurlyeq P_t^{(2)}(I_2)$ and $P_t^{(2)}(I_2) \succcurlyeq P_t^{(2)}(W)$ for $t \in \{1, \ldots, T-1\}$, then $s_t^*(I_1) \geq s_t^*(I_2)$ for $t \in \{1, \ldots, T-1\}$.*

Proof. The proof is by contradiction. If the desired conclusion of the theorem is false, then there is a time $u \in \{1, \ldots, T-1\}$ for which $s_u^*(I_1) < s_u^*(I_2)$; therefore we can find a state $s_u \in S$ such that $s_u^*(I_1) < s_u \leq s_u^*(I_2)$ and

$$\begin{aligned} &R \times Q(s_u) - C^{\text{MED}}(s_u) - C^{\text{INT}} + \sum_{s_{u+1}=0}^{M} p_u^{(1)}(s_{u+1}|s_u, I_1) v_{u+1}^{(1)}(s_{u+1}) \\ &< R \times Q(s_u) - C^{\text{MED}}(s_u) + \sum_{s_{u+1}=0}^{M} p_u^{(1)}(s_{u+1}|s_u, W) v_{u+1}^{(1)}(s_{u+1}). \end{aligned}$$

The above inequality simplifies to

$$\sum_{s_{u+1}=0}^{M} p_u^{(1)}(s_{u+1}|s_u, I_1) v_{u+1}^{(1)}(s_{u+1}) < \sum_{s_{u+1}=0}^{M} p_u^{(1)}(s_{u+1}|s_u, W) v_{u+1}^{(1)}(s_{u+1}) + C^{\text{INT}}.$$

This implies the following two conditions hold simultaneously

$$\sum_{s_{u+1}=0}^{M} \big(p_u^{(1)}(s_{u+1}|s_u, I_1) - p_u^{(1)}(s_{u+1}|s_u, W)\big) v_{u+1}^{(1)}(s_{u+1}) < C^{\text{INT}} \text{ (because } s_u^*(I_1) < s_u\text{), and} \tag{5.12}$$

$$\sum_{s_{u+1}=0}^{M} \big(p_u^{(2)}(s_{u+1}|s_u, I_2) - p_u^{(2)}(s_{u+1}|s_u, W)\big) v_{u+1}^{(2)}(s_{u+1}) \geq C^{\text{INT}} \text{ (because } s_u \leq s_u^*(I_2)). \quad (5.13)$$

Therefore it follows that

$$\begin{aligned} &\sum_{s_{u+1}=0}^{M} \big(p_u^{(1)}(s_{u+1}|s_u, I_1) - p_u^{(1)}(s_{u+1}|s_u, W)\big) v_{u+1}^{(1)}(s_{u+1}) \\ &< \sum_{s_{u+1}=0}^{M} \big(p_u^{(2)}(s_{u+1}|s_u, I_2) - p_u^{(2)}(s_{u+1}|s_u, W)\big) v_{u+1}^{(2)}(s_{u+1}). \end{aligned} \quad (5.14)$$

We will show this is a contradiction (using Lemma 4.7.2 of Puterman) if the following two conditions hold for all $t \in \{1, \ldots, T-1\}$:

$$\begin{aligned} &\sum_{s_{t+1}=k}^{M} \big(p_t^{(1)}(s_{t+1}|s_t, I_1) - p_t^{(1)}(s_{t+1}|s_t, W)\big) \\ &\geq \sum_{s_{t+1}=k}^{M} \big(p_t^{(2)}(s_{t+1}|s_t, I_2) - p_t^{(2)}(s_{t+1}|s_t, W)\big), \text{ for every } k \in S, \text{ and} \end{aligned} \quad (5.15)$$

$$v_{t+1}^{(1)}(s_{t+1}) \geq v_{t+1}^{(2)}(s_{t+1}). \quad (5.16)$$

Condition (5.15) follows from the assumptions that $P_t^{(1)}(I_1) \succcurlyeq P_t^{(2)}(I_2)$ and $P_t^{(1)}(W) \equiv P_t^{(2)}(W)$ for $t \in \{1, \ldots, T-1\}$. The proof of condition (5.16) is by induction. For the base case $t = T$, we have the following

$$v_T^{(1)}(s_T) = \mu_T(s_T) = v_T^{(2)}(s_T), \text{for } s_T \in S.$$

Thus, $v_T^{(1)}(s_T) \geq v_T^{(2)}(s_T)$ for $s_T \in S$. For the inductive step we assume $v_{t+1}^{(1)}(s_{t+1}) \geq v_{t+1}^{(2)}(s_{t+1})$ for $s_{t+1} \in S$ and $t \in \{1, \ldots, T-1\}$. Now we must show $v_t^{(1)}(s_t) \geq v_t^{(2)}(s_t)$ for $s_t \in S$ and $t \in \{1, \ldots, T-1\}$. Let $a_t^{(2)*}(s_t)$ be the optimal action for MDP 2 at time t for a patient in

state $s_t \in S$ and $t \in \{1, \ldots, T-1\}$. It follows that

$$v_t^{(1)}(s_t) \geq r_t(s_t, a_t^{(2)*}(s_t)) + \lambda \sum_{s_{t+1}=0}^{M} p_t^{(1)}(s_{t+1}|s_t, a_t^{(2)*}(s_t)) v_{t+1}^{(1)}(s_{t+1}) \tag{5.17}$$

$$\geq r_t(s_t, a_t^{(2)*}(s_t)) + \lambda \sum_{s_{t+1}=0}^{M} p_t^{(2)}(s_{t+1}|s_t, a_t^{(2)*}(s_t)) v_{t+1}^{(1)}(s_{t+1}) \tag{5.18}$$

$$\geq r_t(s_t, a_t^{(2)*}(s_t)) + \lambda \sum_{s_{t+1}=0}^{M} p_t^{(2)}(s_{t+1}|s_t, a_t^{(2)*}(s_t)) v_{t+1}^{(2)}(s_{t+1}) \tag{5.19}$$

$$= v_t^{(2)}(s_t), \text{ for } s_t \in S \text{ and } t \in \{1, \ldots, T-1\}.$$

Inequality (5.17) follows from the fact that $v_t^{(1)}(s_t)$, the optimal value function, is lower bounded by the value function for any other policy (in this case the optimal policy for MDP 2). Inequality (5.18) holds by the assumptions that $P_t^{(1)}(I_1) \succcurlyeq P_t^{(2)}(I_2)$ and $P_t^{(1)}(W) \equiv P_t^{(2)}(W)$, and Inequality (5.19) holds by the inductive hypothesis. Thus, $v_t^{(1)}(s_t) \geq v_t^{(2)}(s_t)$ for all t and for all $s_t \in S$, and the proof of condition (5.16) is complete.

Now we use conditions (5.15) and (5.16) to complete the proof. By Lemma 4.7.2 of Puterman the following inequality holds for $t \in \{1, \ldots, T-1\}$:

$$\sum_{s_{t+1}=0}^{M} \big(p_t^{(1)}(s_{t+1}|s_t, I_1) - p_t^{(1)}(s_{t+1}|s_t, W)\big) v_{t+1}^{(1)}(s_{t+1})$$

$$\geq \sum_{s_{t+1}=0}^{M} \big(p_t^{(2)}(s_{t+1}|s_t, I_2) - p_t^{(2)}(s_{t+1}|s_t, W)\big) v_{t+1}^{(1)}(s_{t+1}), \tag{5.20}$$

since condition (5.15) holds and $v_{t+1}^{(1)}(s_{t+1}+1) \geq v_{t+1}^{(1)}(s_{t+1})$ by Lemma 5.5.2. Finally, the following inequality holds by applying the assumptions that $v_t^{(1)}(s_t) - v_t^{(2)}(s_t)$ is nondecreasing in s_t for $t \in \{1, \ldots, T-1\}$ and $P_t^{(2)}(I_2) \succcurlyeq P_t^{(2)}(W)$ for $t \in \{1, \ldots, T-1\}$ together with condition (5.16):

$$\sum_{s_{t+1}=0}^{M} \big(p_t^{(2)}(s_{t+1}|s_t, I_2) - p_t^{(2)}(s_{t+1}|s_t, W)\big) v_{t+1}^{(1)}(s_{t+1})$$
$$\geq \sum_{s_{t+1}=0}^{M} \big(p_t^{(2)}(s_{t+1}|s_t, I_2) - p_t^{(2)}(s_{t+1}|s_t, W)\big) v_{t+1}^{(2)}(s_{t+1}) \text{ for } t \in \{1, \ldots, T-1\}. \quad (5.21)$$

Therefore from (5.20) and (5.21) we have

$$\sum_{s_{t+1}=0}^{M} \big(p_t^{(1)}(s_{t+1}|s_t, I_1) - p_t^{(1)}(s_{t+1}|s_t, W)\big) v_{t+1}^{(1)}(s_{t+1})$$
$$\geq \sum_{s_{t+1}=0}^{M} \big(p_t^{(2)}(s_{t+1}|s_t, I_2) - p_t^{(2)}(s_{t+1}|s_t, W)\big) v_{t+1}^{(2)}(s_{t+1}) \text{ for } t \in \{1, \ldots, T-1\}. \quad (5.22)$$

Thus, we have our contradiction since Inequality (5.22) violates Inequality (5.14). □

Theorem 5.5.2 can be interpreted as follows. If intervention I_1 is more effective than intervention I_2, then the optimal control limit for I_1 in MDP 1 should be greater than or equal to the optimal control limit for I_2 in MDP 2. In other words, under the optimal policy, intervention I_1 would be implemented for a wider range of adherence states. I_1 may be used for patients in better adherence states than I_2.

In the next, and final, theorem we use a superscript to index patients in order to compare the optimal intervention thresholds for two types of patients. The superscript for patient 1 is $(1')$ and the superscript for patient 2 is $(2')$.

Theorem 5.5.3. *If $\tilde{P}_t(I) \succcurlyeq \tilde{P}_t(W)$ for $t \in \{1, \ldots, T-1\}$ and $v_t^{(2')}(s_t) - v_t^{(1')}(s_t)$ is nondecreasing in s_t for $t \in \{1, \ldots, T-1\}$, then for two patients that are identical except*

$$[\bar{p}_t^{(1')}]_i \geq [\bar{p}_t^{(2')}]_i, \text{ for every } i = 1, \ldots M \text{ and for } t \in \{1, \ldots, T-1\}, \quad (5.23)$$

and

$$\mu_T^{(1')}(s_T) \leq \mu_T^{(2')}(s_T), \text{ for every } s_T \in S \tag{5.24}$$

then

$$s_t^{*(1')} \leq s_t^{*(2')} \text{ for } t \in \{1, \ldots, T-1\}. \tag{5.25}$$

Proof. Since $\tilde{P}_t(a_t)$ is the same for both patients, it follows that $P_t^{(2')}(W) \succcurlyeq P_t^{(1')}(W)$ and $P_t^{(2')}(I) \succcurlyeq P_t^{(1')}(I)$ for $t \in \{1, \ldots, T-1\}$. The proof for this theorem is similar to the proof for Theorem 5.5.2.

The proof is by contradiction. If the desired conclusion (5.25) of the theorem is false, then there is a time $u \in \{1, \ldots, T-1\}$ for which $s_u^{*(1')} > s_u^{*(2')}$; therefore we can find a state $s_u \in S$ such that $s_u^{*2'} < s_u \leq s_u^{*(1')}$ and

$$\begin{aligned} &R \times Q(s_u) - C^{\text{MED}}(s_u) - C^{\text{INT}} + \sum_{s_{u+1}=0}^{M} p_u^{(2')}(s_{u+1}|s_u, I) v_{u+1}^{(2')}(s_{u+1}) \\ &< R \times Q(s_u) - C^{\text{MED}}(s_u) + \sum_{s_{u+1}=0}^{M} p_u^{(2')}(s_{u+1}|s_u, W) v_{u+1}^{(2')}(s_{u+1}). \end{aligned}$$

This simplifies to

$$\sum_{s_{u+1}=0}^{M} p_u^{(2')}(s_{u+1}|s_u, I) v_{u+1}^{(2')}(s_{u+1}) < \sum_{s_{u+1}=0}^{M} p_u^{(2')}(s_{u+1}|s_u, W) v_{u+1}^{(2')}(s_{u+1}) + C^{\text{INT}}.$$

This implies the following two conditions hold simultaneously

$$\sum_{s_{u+1}=0}^{M} \big(p_u^{(2')}(s_{u+1}|s_u, I) - p_u^{(2')}(s_{u+1}|s_u, W)\big) v_{u+1}^{(2')}(s_{u+1}) < C^{\text{INT}} \text{ (because } s_u^{*(2')}\text{), and} \tag{5.26}$$

$$\sum_{s_{u+1}=0}^{M} \big(p_u^{(1')}(s_{u+1}|s_u, I) - p_u^{(1')}(s_{u+1}|s_u, W)\big) v_{u+1}^{(1')}(s_{u+1}) \geq C^{\text{INT}} \text{ (because } s_u \leq s_u^{*(1')}\text{).} \tag{5.27}$$

Therefore it follows that

$$\begin{aligned}
&\sum_{s_{u+1}=0}^{M} \big(p_u^{(2')}(s_{u+1}|s_u, I) - p_u^{(2')}(s_{u+1}|s_u, W)\big) v_{u+1}^{(2')}(s_{u+1}) \\
&< \sum_{s_{u+1}=0}^{M} \big(p_u^{(1')}(s_{u+1}|s_u, I) - p_u^{(1')}(s_{u+1}|s_u, W)\big) v_{u+1}^{(1')}(s_{u+1}). \qquad (5.28)
\end{aligned}$$

We prove this is a contradiction if the following two conditions hold for all $t \in \{1, \ldots, T-1\}$:

$$\begin{aligned}
&\sum_{s_{t+1}=k}^{M} \big(p_t^{(2')}(s_{t+1}|s_t, I) - p_t^{(2')}(s_{t+1}|s_t, W)\big) \\
&\geq \sum_{s_{t+1}=k}^{M} \big(p_t^{(1')}(s_{t+1}|s_t, I) - p_t^{(1')}(s_{t+1}|s_t, W)\big), \text{ for every } k \in S, \text{ and} \qquad (5.29)
\end{aligned}$$

$$v_{t+1}^{(2')}(s_{t+1}) \geq v_{t+1}^{(1')}(s_{t+1}). \qquad (5.30)$$

The proof of condition (5.29) is as follows. For $k = 0$, the inequality holds trivially. For $k = 1, \ldots, M$, we have the following

$$
\begin{aligned}
&\sum_{s_{t+1}=k}^{M} \big(p_t^{(2')}(s_{t+1}|s_t, I) - p_t^{(2')}(s_{t+1}|s_t, W)\big) \\
=&\sum_{s_{t+1}=k}^{M} \big((1 - [\bar{p}_t^{(2')}]_{s_t})\tilde{p}_t(s_{t+1}|s_t, I) - (1 - [\bar{p}_t^{(2')}]_{s_t})\tilde{p}_t(s_{t+1}|s_t, W)\big) \\
=&(1 - [\bar{p}_t^{(2')}]_{s_t}) \sum_{s_{t+1}=k}^{M} \big(\tilde{p}_t(s_{t+1}|s_t, I) - \tilde{p}_t(s_{t+1}|s_t, W)\big) \\
\geq&(1 - [\bar{p}_t^{(1')}]_{s_t}) \sum_{s_{t+1}=k}^{M} \big(\tilde{p}_t(s_{t+1}|s_t, I) - \tilde{p}_t(s_{t+1}|s_t, W)\big) \qquad (5.31) \\
=&\sum_{s_{t+1}=k}^{M} \big((1 - [\bar{p}_t^{(1')}]_{s_t})\tilde{p}_t(s_{t+1}|s_t, I) - (1 - [\bar{p}_t^{(1')}]_{s_t})\tilde{p}_t(s_{t+1}|s_t, W)\big) \\
=&\sum_{s_{t+1}=k}^{M} \big(p_t^{(1')}(s_{t+1}|s_t, I) - p_t^{(1')}(s_{t+1}|s_t, W)\big),
\end{aligned}
$$

where $\tilde{p}_t(s_{t+1}|s_t, a_t)$ is the (s_t, s_{t+1}) element of $\tilde{P}_t(a_t)$. Inequality (5.31) follows by the assumptions that $[\bar{p}_t^{(1')}]_i \geq [\bar{p}_t^{(2')}]_i$ and $\tilde{P}_t(I) \succcurlyeq \tilde{P}_t(W)$ for $t \in \{1, \ldots, T-1\}$.

The proof of condition (5.30) is by induction. For the base case $t = T$, we have the following

$$
v_T^{(2')}(s_T) = \mu_T^{(2')}(s_T) \geq \mu_T^{(1')}(s_T) = v_T^{(1')}(s_T).
$$

Thus, $v_T^{(2')}(s_T) \geq v_T^{(1')}(s_T)$. For the inductive step we assume $v_{t+1}^{(2')}(s_{t+1}) \geq v_{t+1}^{(1')}(s_{t+1})$. Now we must show $v_t^{(2')}(s_t) \geq v_t^{(1')}(s_t)$ for $s_t \in S$ and $t \in \{1, \ldots, T-1\}$. Let $a_t^{(1')*}(s_t)$ be the optimal

action at time t for patient 1 in state s_t for $s_t \in S$ and $t \in \{1, \ldots, T-1\}$. It follows that

$$
\begin{aligned}
v_t^{(2')}(s_t) &\geq r_t(s_t, a_t^{(1')*}(s_t)) + \lambda \sum_{s_{t+1}=0}^{M} p_t^{(2')}(s_{t+1}|s_t, a_t^{(1')*}(s_t)) v_{t+1}^{(2')}(s_{t+1}) && (5.32)\\
&\geq r_t(s_t, a_t^{(1')*}(s_t)) + \lambda \sum_{s_{t+1}=0}^{M} p_t^{(1')}(s_{t+1}|s_t, a_t^{(1')*}(s_t)) v_{t+1}^{(2')}(s_{t+1}) && (5.33)\\
&\geq r_t(s_t, a_t^{(1')*}(s_t)) + \lambda \sum_{s_{t+1}=0}^{M} p_t^{(1')}(s_{t+1}|s_t, a_t^{(1')*}(s_t)) v_{t+1}^{(1')}(s_{t+1}) && (5.34)\\
&= v_t^{(1')}(s_t).
\end{aligned}
$$

Inequality (5.33) holds since $P_t^{(2')}(W) \succcurlyeq P_t^{(1')}(W)$ and $P_t^{(2')}(I) \succcurlyeq P_t^{(1')}(I)$. Thus, $v_t^{(2')}(s_t) \geq v_t^{(1')}(s_t)$ for all t and for all $s_t \in S$, and the proof of condition (5.30) is complete.

Now we use conditions (5.29) and (5.30) to complete the proof. By Lemma 4.7.2 of Puterman, the following inequality holds for $t \in \{1, \ldots, T-1\}$.

$$
\begin{aligned}
&\sum_{s_{t+1}=0}^{M} \big(p_t^{(2')}(s_{t+1}|s_t, I) - p_t^{(2')}(s_{t+1}|s_t, W)\big) v_{t+1}^{(2')}(s_{t+1})\\
&\geq \sum_{s_{t+1}=0}^{M} \big(p_t^{(1')}(s_{t+1}|s_t, I) - p_t^{(1')}(s_{t+1}|s_t, W)\big) v_{t+1}^{(2')}(s_{t+1}), \qquad (5.35)
\end{aligned}
$$

since condition (5.29) holds and $v_{t+1}^{(2')}(s_{t+1}+1) \geq v_{t+1}^{(2')}(s_{t+1})$ by Lemma 5.5.2. Finally, the following condition holds by applying the assumptions $P_t^{(1')}(I) \succcurlyeq P_t^{(1')}(W)$ for $t \in \{1, \ldots, T-1\}$ and $v_t^{(2')}(s_t) - v_t^{(1')}(s_t)$ is nondecreasing in s_t for $t \in \{1, \ldots, T-1\}$ together with condition (5.30) and Lemma 4.7.2 of Puterman (1994):

$$
\begin{aligned}
&\sum_{s_{t+1}=0}^{M} \big(p_t^{(1')}(s_{t+1}|s_t, I) - p_t^{(1')}(s_{t+1}|s_t, W)\big) v_{t+1}^{(2')}(s_{t+1})\\
&\geq \sum_{s_{t+1}=0}^{M} \big(p_t^{(1')}(s_{t+1}|s_t, I) - p_t^{(1')}(s_{t+1}|s_t, W)\big) v_{t+1}^{(1')}(s_{t+1}). \qquad (5.36)
\end{aligned}
$$

Therefore from (5.35) and (5.36)

$$\sum_{s_{t+1}=0}^{M} \big(p_t^{(2')}(s_{t+1}|s_t, I) - p_t^{(2')}(s_{t+1}|s_t, W)\big) v_{t+1}^{(2')}(s_{t+1})$$
$$\geq \sum_{s_{t+1}=0}^{M} \big(p_t^{(1')}(s_{t+1}|s_t, I) - p_t^{(1')}(s_{t+1}|s_t, W)\big) v_{t+1}^{(1')}(s_{t+1}). \tag{5.37}$$

Thus, we have our contradiction since Inequality (5.37) violates Inequality (5.28). □

Theorem 5.5.3 states that if patient 1 has a higher probability of moving to an absorbing state than patient 2, then patient 2 should have interventions in the same or better adherence states than patient 1. Since $P_t^{(2')}(I) \succcurlyeq P_t^{(1')}(I)$, patient 1, the sicker patient, receives less benefit from interventions than patient 2. Interventions that are optimal for patient 2 with better adherence may not be optimal for patient 1. This theorem provides a simple criterion for sorting patients on the basis of importance of an intervention, which could be useful for resource constrained settings.

5.6 Case Study: Statin Adherence for Patients with Type 2 Diabetes

In this section we present a case study to illustrate the application of our model to evaluate a hypothetical EHR-based AAS system in the context of preventive treatment for cardiovascular disease. Specifically, we investigate adherence interventions for statin treatment among patients with type 2 diabetes. Statins are particularly important for patients with diabetes, since these patients are at two to four times higher risk for stroke and CHD events over patients without diabetes [19]. Furthermore, long-term adherence to statins is known to be poor [14, 66].

In Section 5.6.1 we provide our data sources and model parameters. In Section 5.6.2 we compare active and inactive surveillance policies using the MDP model described in Section

5.4. We present the optimal policies and expected LYs and costs associated with these policies. We also explore the effects of gender, the patient's health risk, the cost of an intervention, the willingness-to-pay factor, and the type of intervention on the optimal policy. We conclude this section with an estimate of total benefits of AAS to the U.S. diabetes population.

5.6.1 Data and Model Parameter Estimation

The transition probabilities among adherence states were computed from the administrative medical and pharmacy claims data from a large health insurance company that insures patients across the United States. A cohort of 54,036 diabetes patients from this dataset were identified using Healthcare Effectiveness Data and Information Set (HEDIS) criteria for diagnosis of diabetes [70]. Patients included in the set were required to have five years of continuous enrollment with first encounter dates ranging from January 1995 to June 2004. The PDC by pharmacy fills, described in Section 5.2, was used as a proxy for patient adherence rates. Once the PDC was computed for each patient, the transition probabilities were computed by counting the number of patients in each adherence state that transitioned to each adherence state in the next year. The associated effect of statins on the patient's TC level for each adherence level was derived from this observational data set as well (see [66] for a detailed description).

The transition probabilities for stroke and CHD events were derived from the UKPDS risk models [61, 97], and the probabilities for death from other causes were calculated from the CDC mortality tables [3]. The state of the patient's health (other than their adherence level), which we used to estimate stroke and CHD event probabilities with the UKPDS model, was based on observations from a large cohort of 663 patients receiving treatment for type 2 diabetes at Mayo Clinic, Rochester, MN. Approximately $15,000$ measurements of HbA1c (a patient's average blood sugar over two to three months), blood pressure, and cholesterol were collected between 1997 and 2006 through the Mayo Clinic Diabetes Electronic Management System (DEMS) [44].

For all of our experiments we assumed a maximum age of $T = 100$ as the age at which

Table 5.2: Initial hospitalization costs and follow-up events for adverse events.

Parameter	Cost	Citation
Initial Hospitalization for Stroke	\$13,204	[1]
Initial Hospitalization for CHD	\$18,590	[1]
Yearly Follow-up for Stroke	\$1664	[101]
Yearly Follow-up for CHD	\$2576	[101]

interventions would be discontinued and a discount factor of $\lambda = 0.97$ which corresponds to a 3% yearly discount rate [41]. For the base case, we assumed a willingness to pay of $R = \$100,000$ [37] and a cost of statins of $C^{\text{MED}}(s_t) = \$212 \times \delta(s_t)$, where $\delta(s_t)$ represents the mean PDC of a patient in adherence state s_t [4]. The cost of an intervention was estimated to be $C^{\text{INT}} = \$123$ from a cost-benefit analysis of interventions by [33] that was inflated to 2009 dollars using the consumer price index method [2]. This intervention cost includes clarification of the doctor's message, family member reinforcement, and group meetings to improve patient adherence. The initial and follow-up costs of stroke and CHD events were drawn from sources in the health services research literature provided in Table 5.2. The one-time penalty of entering the absorbing state, C_t^{F}, is computed with a Markov chain using these costs and probabilities governing patient survival.

The adherence states used in the numerical experiments are NON ($0\% \leq \text{PDC} \leq 10\%$), LOW ($10\% < \text{PDC} \leq 40\%$), MED ($40\% < \text{PDC} \leq 80\%$), and HIGH ($80\% < \text{PDC} \leq 100\%$) ([66]). The transition probability matrices, $\tilde{P}_t(a_t)$, were estimated to be

$$\tilde{P}_t(W) = \begin{array}{c} \\ \text{NON} \\ \text{LOW} \\ \text{MED} \\ \text{HIGH} \end{array} \begin{array}{c} \begin{array}{cccc} \text{NON} & \text{LOW} & \text{MED} & \text{HIGH} \end{array} \\ \begin{pmatrix} 0.787 & 0.106 & 0.082 & 0.025 \\ 0.498 & 0.205 & 0.213 & 0.084 \\ 0.199 & 0.154 & 0.390 & 0.257 \\ 0.028 & 0.046 & 0.189 & 0.737 \end{pmatrix} \end{array},$$

and

$$\tilde{P}_t(I) = \begin{array}{c} \\ \text{NON} \\ \text{LOW} \\ \text{MED} \\ \text{HIGH} \end{array} \begin{array}{c} \begin{array}{cccc} \text{NON} & \text{LOW} & \text{MED} & \text{HIGH} \end{array} \\ \begin{pmatrix} 0.091 & 0.165 & 0.257 & 0.487 \\ 0.091 & 0.165 & 0.257 & 0.487 \\ 0.091 & 0.165 & 0.257 & 0.487 \\ 0.091 & 0.165 & 0.257 & 0.487 \end{pmatrix} \end{array}.$$

The matrix $\tilde{P}_t(I)$ was estimated based on the proportion of patients occupying each of the adherence states in their first year of treatment. This assumption was made since an intervention may act to "reset" a patient's adherence level to the level it was when the patient initially began treatment. In addition, we considered the more optimistic case that a patient moves to state HIGH with probability 1. Use of this intervention provides a conservative estimate of the improvement achievable through interventions. The use of actual data to estimate the probabilities among the adherence states inherently includes the effects of diet, exercise, and other behavioral changes.

5.6.2 Numerical Results

Numerical experiments were conducted to find the optimal policy for adherence-improving interventions based on the above model parameters. The model was solved using backwards recursion, implemented in C/C++. Each experiment took less than ten seconds to run using a 2.83GHz PC with 8GB of RAM. Experiments were run for males and females, starting at age 40, assuming a variety of different risk states and different intervention cost estimates. The perfect and imperfect interventions described in Section 5.6.1 were both evaluated. We represent different risk states by the patient's TC and high-density lipoprotein (HDL), also known as "good" cholesterol, each given as one of low (L), medium (M), high (H), and very high (V). These are the most significant metabolic factors influencing a patient's risk of stroke or CHD

events according to the UKPDS model. While there are a total of 16 patient risk states defined by clinically-relevant thresholds [24], for brevity we provide policies and numerical results for representative patients with low risk (low TC and very high HDL), medium risk (medium TC and medium HDL), and high risk (very high TC and low HDL).

5.6.2.1 Active vs. Inactive Surveillance

To estimate the potential benefits of using EHRs to improve adherence to medication at the population level, we compared the expected LYs from age 40 prior to an event or death and the expected discounted total costs comprising the costs of intervention, statin treatment, and hospitalizations and follow-up care for CHD events and stroke found using the optimal AAS policy and the IAS policy. IAS involves periodic interventions that do not rely on a patient's adherence level. We considered interventions that occur every k years ($k = 1, 2, 3, 4$, or 5) after a patient begins taking medication, regardless of the patient's adherence state. The IAS policy is useful for comparison since it requires no pharmacy or laboratory data and is therefore much easier to implement in practice. We also considered the use of no interventions.

Figures 5.2 and 5.3 show the expected LYs vs. costs for AAS, IAS, and no treatment, for females and males. Imperfect interventions were used for these results. We evaluated different AAS policies by varying the willingness-to-pay factor from $R = \$0$ to $R = \$1{,}000{,}000$. When the willingness-to-pay factor is varied, different weights are placed on LYs and costs. As this factor increases, a larger weight is placed on maximizing the patient's LYs rather than minimizing costs. We observe that AAS outperforms IAS, for females and males, yielding greater expected LYs before an event or death and lower expected costs when $R \geq \$25{,}000$. When $R = \$100{,}000$, the base case value for our experiments, the average female patient using AAS receives an expected 0.17 additional LYs with a \$1727 reduction in costs over IAS ($k = 1$), and the average male patient using AAS receives an expected 0.19 additional LYs with a \$1808 reduction in costs over IAS ($k = 1$). AAS resulted in no interventions for patients with HIGH adherence. The higher expected costs incurred by IAS are presumably due in

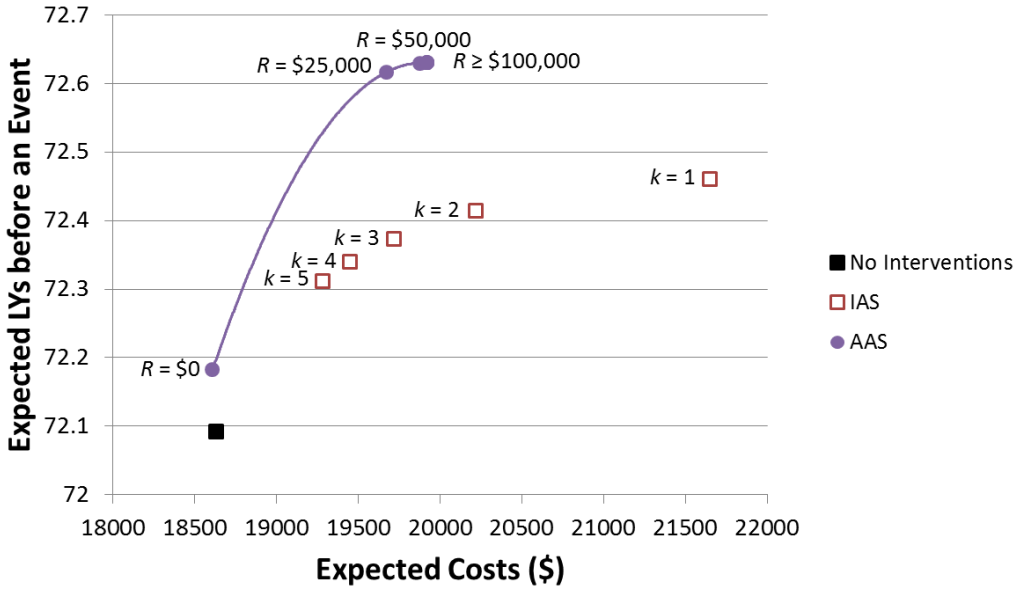

Figure 5.2: Comparison of expected LYs verses costs for medication, interventions, and treatment of events for active adherence surveillance (AAS) policies (with varying R values) and inactive adherence surveillance (IAS) policies (when interventions occur every k years) for female patients using imperfect interventions. Results are a weighted average of LYs and costs for the 16 possible risk states.

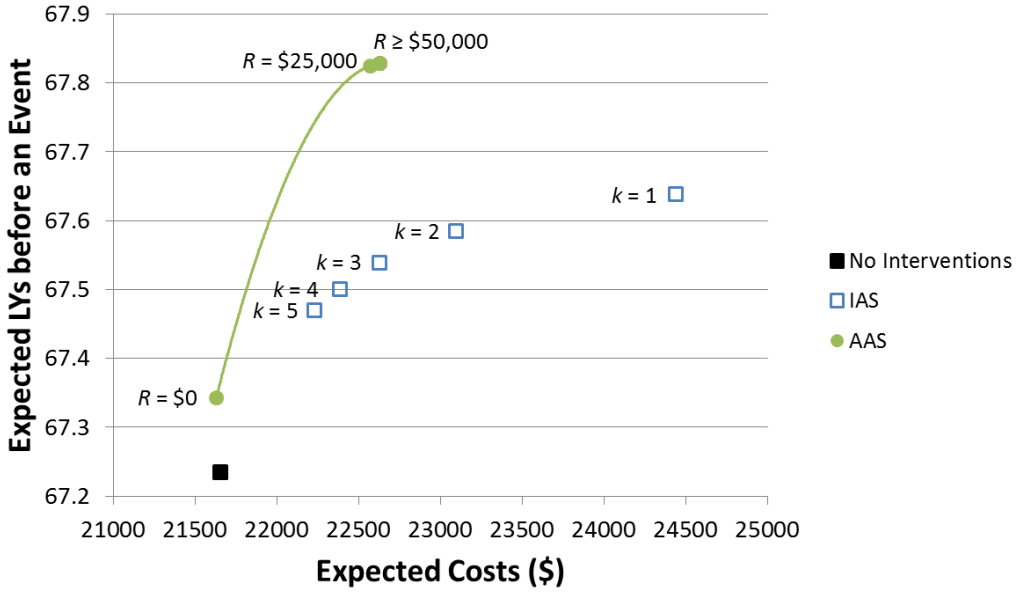

Figure 5.3: Comparison of expected LYs verses costs, as shown in Figure 5.2, for male patients.

part to unnecessary interventions for patients with HIGH adherence to treatment, highlighting the benefit of AAS. It is particularly interesting that there are major gender differences in the expected LYs before an event or death. Based on our results, we observe that males are expected to have an adverse event or death approximately 5 years earlier than females.

While AAS dominates IAS for all 16 risk states, there are significant differences in the magnitude of the differences in expected cost and LYs for patients with different risk of CHD events and stroke. Figure 5.4 presents results for females with low, medium, and high risk in a format similar to Figure 5.2. Patients with low risk can expect to have their first event or death later in life than patients with medium or high risk. Also, as a patient's risk increases, her benefit over no treatment and her benefit over IAS increases. Thus, it appears the benefit of AAS is increasing in patient risk. We also note that the expected costs and LYs are less sensitive to changes in the willingness-to-pay factor as risk increases. The observations for males are consistent with the results for females.

We performed sensitivity analysis on the type of intervention. When a perfect intervention is considered, AAS (R =\$100,000) and IAS ($k = 1$) achieve nearly the same expected LYs before an adverse health event or death, with AAS providing 0.0016 fewer LYs for females and 0.000024 fewer LYs for males. AAS results in an average reduction in costs of \$516 for females and \$635 for males. Thus, if perfect interventions were achievable, the benefit of AAS would be diminished.

5.6.2.2 Sensitivity to Cost of Intervention

We performed sensitivity analysis on the cost of interventions. When interventions are free, we observe that patients should have yearly interventions starting at age 41 ($t = 1$), the earliest possible age for interventions to occur in our model, since there is no downside for free interventions. For C^{INT} = \$61.50, \$123, or \$246, we observe female patients should have yearly interventions starting at the ages listed in Tables 5.3 and 5.4. The optimal policy for male patients follows a similar pattern to the optimal policy for female patients, but male patients

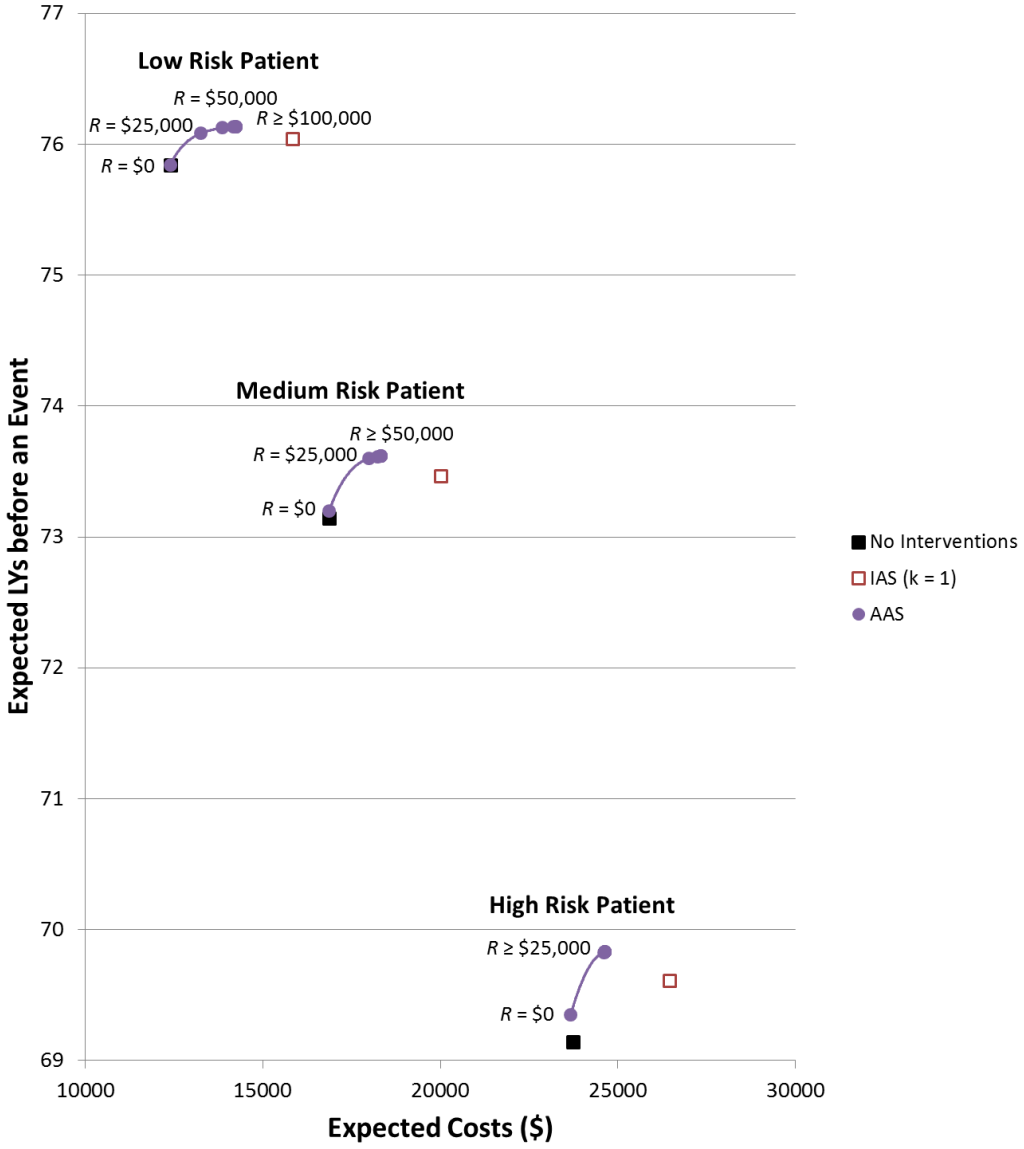

Figure 5.4: Comparison of expected LYs verses costs for medication, interventions, and treatment of events for active adherence surveillance (AAS) policies (with varying R values) and yearly inactive adherence surveillance (IAS) for female patients using imperfect interventions. Results are compared for low, medium, and high risk patients.

should start having interventions up to 10 years earlier than female patients, depending on the type and cost of intervention. The differences between the policies for male and female patients are likely due to the fact that males generally have an earlier onset of risk for cardiovascular events than females.

The optimal policy, presented in Tables 5.3 and 5.4, exhibits a control limit structure, as expected from Theorem 5.5.1. We also observe that the control limit tends to increase with respect to age with the exception that at very old ages interventions are no longer optimal. The latest age of intervention ranges from 95 to 98, depending on the patient's risk level and the cost of the intervention. We expect this is due to the end of horizon approximation in which we truncate the decision horizon at $T = 100$.

5.6.2.3 Sensitivity to Individual Patient Risk Factors

In general, female patients and patients with lower risk stop having interventions earlier due to lower risk of stroke and CHD events. The policies are very insensitive to changes in the cost of interventions, particularly for males and patients in higher risk states. We observe that the higher cost interventions have a shorter range for which it is optimal to perform the interventions; that is, the interventions start later and do not continue as late in life. The female patients have fewer interventions overall, with interventions starting later and ending

Table 5.3: Optimal ages to begin having yearly interventions for female patients using active surveillance. Imperfect (probabilistic) interventions are assumed. Note: '–' denotes it is never optimal for the patient to have interventions.

Low Risk				Medium Risk				High Risk			
	\$61.50	\$123	\$246		\$61.50	\$123	\$246		\$61.50	\$123	\$246
NON	41	41	41	NON	41	41	41	NON	41	41	41
LOW	41	41	41	LOW	41	41	41	LOW	41	41	41
MED	41	43	54	MED	41	41	42	MED	41	41	41
HIGH	–	–	–	HIGH	–	–	–	HIGH	–	–	–

earlier. This is likely due to the fact that being male is a risk factor for stroke and CHD events, the events statins help prevent.

When perfect interventions are considered, it is always optimal for male and female patients to have interventions when their adherence is less than HIGH. The use of perfect interventions for patients with HIGH adherence depends on the intervention cost and risk state. For imperfect interventions, however, patients with HIGH adherence should never have an intervention since the probability of remaining in the HIGH adherence state under an intervention is lower than the probability of remaining in the HIGH adherence state without an intervention.

5.6.2.4 Potential Yearly Benefits of AAS to the U.S. Diabetes Population

In order to estimate the benefits of AAS applied to all diabetes patients in the United States, we first estimated the prevalence of diabetes in the United States using population estimates, by age and gender, based on the 2010 U.S. Census [107], and the estimated diabetes prevalence by state and age range reported by Danaei et al. [27]. Next, we estimated the number of newly diagnosed diabetes patients for each gender, for every state and the District of Columbia, and for each age, starting at age 40. Patients were defined as *newly diagnosed* in 2010 if they were a diabetes patient at age 40 or an older patient diagnosed later in life. Patients were identified as newly diagnosed past age 40 if the population of total patients diagnosed at earlier ages was less than the diagnosed population at the given age. This accounts for increases in population

Table 5.4: Optimal ages to begin having yearly interventions for female patients using active surveillance. Perfect interventions are assumed.

Low Risk				Medium Risk				High Risk			
	\$61.50	\$123	\$246		\$61.50	\$123	\$246		\$61.50	\$123	\$246
NON	41	41	41	NON	41	41	41	NON	41	41	41
LOW	41	41	41	LOW	41	41	41	LOW	41	41	41
MED	41	41	41	MED	41	41	41	MED	41	41	41
HIGH	43	55	66	HIGH	41	43	54	HIGH	41	41	41

and diabetes prevalence with respect to age.

Table 5.5 provides a yearly estimate of expected LYs and costs over the remaining years of life for newly diagnosed diabetes patients aged 40 or older with no interventions, IAS ($k = 1$), and AAS. According to our model, the implementation of IAS ($k = 1$), compared to no interventions, would increase LYs for the U.S. population at a cost of \$6990/LY. In comparison, AAS would increase LYs over no interventions for the U.S. population at a cost of \$1455/LY. Using AAS in place of IAS ($k = 1$) would result in over 131,000 additional LYs among adults newly diagnosed with diabetes while saving over \$1.41 billion per year.

Table 5.5: Yearly costs (billions) and future LYs for newly-diagnosed diabetes patients using no adherence interventions, yearly inactive adherence surveillance (IAS, $k = 1$), and active adherence surveillance (AAS).

	Males		Females		Total Population	
	LYs	Cost (billions)	LYs	Cost (billions)	LYs	Cost (billions)
No Interventions	8,142,611	\$10.59	10,355,836	\$10.19	18,498,448	\$20.78
IAS ($k = 1$)	8,287,353	\$11.53	10,501,173	\$11.28	18,788,525	\$22.81
AAS	8,353,550	\$10.83	10,565,978	\$10.56	18,919,528	\$21.39

5.7 Conclusions

The CMS Meaningful Use initiative has the potential to encourage improved efficiency and effectiveness of healthcare delivery through the use of EHRs. Based on our results we found that the use of EHRs to improve adherence has the potential to significantly delay the onset of adverse events or death, and reduce expected costs of treatment, hospitalization, and follow-up care associated with adverse events such as stroke and CHD. From the population perspective, we found that AAS is cost effective compared with no interventions at a cost of $1455 spent per LY added prior to CHD, stroke, or death. This cost per LY added is very low with respect to commonly used thresholds [37]. In addition, AAS results in significant cost savings over IAS ($k = 1$) while providing more than 131,000 additional event-free LYs to newly diagnosed diabetes patients each year at a savings of $1.41 billion per year. These estimated annual benefits highlight the potential benefits of AAS. Our study considers the use of AAS for a subpopulation in the United States that is at a high risk of stroke and CHD events; however, AAS could be used for the broader U.S. population and for patients on other medications, yielding additional savings.

From the individual patient perspective, males receive an average of 0.19 additional LYs before an event or death over IAS ($k = 1$) at a reduction in costs of $1808, and females receive 0.17 additional LYs at a cost savings of $1727 over IAS ($k = 1$). These increases in LYs over IAS ($k = 1$) are an order of magnitude greater than the benefits seen through some prevention programs that are part of standard practice in the United States. For example, childhood vaccination against measles, mumps, and rubella results in an increase of 0.017 LYs [114]. In addition, the increase in LYs from AAS over no interventions is even greater. The benefits of AAS over IAS and no interventions increase with increasing patient risk. In other words, patients at higher risk of adverse events stand to have greater benefit from AAS.

We found the optimal policy for adherence-improving interventions to exhibit a control-limit type policy. This is consistent with the theoretical results we presented. From our numerical experiments, it appears that the control limit is increasing with respect to age. Once a patient

begins having interventions, it is generally optimal to continue having yearly interventions until very late in life. Such a simple policy is encouraging for the application of AAS system in the already complex clinical environment.

We proved structural properties related to the optimal control limit when interventions of different effectiveness are considered, and when patients of different levels of risk are considered. While we presented Theorems 5.5.2 and 5.5.3 in the context of the problem we are applying our model to (the optimal timing of adherence-improving interventions for patients with type 2 diabetes), these theoretical properties and our model are generalizable to other contexts. For example, in the context of machine maintenance, Theorem 5.5.3 could be useful in scheduling maintenance for different types of machines that have different levels of reliability.

Although the outcomes of the AAS policy dominated the easier-to-implement IAS outcomes, our model did not account for the possibility of initial set-up costs and ongoing maintenance costs for such a system. While the data our model is based on is generally available in administrative claims systems and laboratory information systems, the development of a decision support system that collects and utilizes the data would have some cost associated with instantiation of the system in a clinical environment. Although we did not consider this in our analysis, it is worth noting that our model can easily be modified to incorporate any maintenance costs that would be necessary to use AAS. In addition, our model could be used to estimate the payback time for the initial costs of the system by calculating expected return on investment of using AAS over IAS. Furthermore, CMS incentives for participation in the Meaningful Use program may offset some of the costs of implementation.

There are some practical challenges associated with the use of EHR data for applications such as we discuss. First, patients do not always stay with the same insurance provider. There may be a limited amount of time for which there is continuous information for each patient. This challenge may eventually be overcome by the development of a universal EHR. Second, our model assumes population level data can be used to estimate parameters for individual patients. In the case of adherence to medical treatments, such as statin therapy, this is reasonable because

researchers have not been able to identify ways to predict adherence on the basis of available health data. Nevertheless, the use of population level data represents a barrier to more accurate prediction of adherence that might be possible with additional data. Third, in order for AAS to be implemented at the point of care, EHRs will need to collect and compute patient information such as PDC for prescribed medications (to estimate the patients adherence level) and patient health information (such as blood pressure and cholesterol) to estimate the risk of adverse events. While we have demonstrated this is possible in this chapter, the ability to rapidly collect and combine such data presents a challenge for some health systems.

Future research could build on our model in several ways. For example, we considered interventions for a single medical treatment. Future studies could extend the current model to include the optimal timing of interventions for patients on multiple medications. This generates a number of interesting questions. For example, would there be correlations between interventions? In other words, could an intervention for one medication influence adherence to another medication? Could an intervention be designed that would simultaneously improve adherence to multiple medications? Furthermore, interesting questions arise about the relative importance of interventions. For example, our model could be amended to help prioritize interventions for different medications. Additional variations on our model could include different assumptions about the effectiveness of interventions. Although there is no evidence at present, it is possible that interventions may provide diminishing improvement to adherence over time. As we pointed out in the introduction, the recent substantial commitment of resources and efforts by the medical community to improve the current state of knowledge about medication adherence presents a number research opportunities for the OR community. Our model lays the foundation for some of these future studies.

Chapter 6

Conclusions

In this dissertation we presented two novel MDP models related to the optimal timing of medical treatment and the optimal timing of adherence-improving interventions. In the first MDP model, presented in Chapter 3, we solved a discrete-state MDP to determine the optimal sequence and timing of blood pressure and cholesterol medications over a patient's lifetime to manage the risk of stroke and CHD events. We considered the primary prevention objective of maximizing rewards for LYs minus medication costs before a patient's first adverse health event or death and the objective of maximizing rewards for QALYs minus costs of medication and treatment of events over a patient's lifetime. We showed the use of optimal guidelines over the current U.S. guidelines can provide a significant reduction in costs without reducing benefits of treatment, particularly when primary and secondary prevention is the objective. Based on our numerical experiments it appears that significant savings could be realized at the population level if the optimal guidelines were applied to all diabetes patients in the United States.

Coordination of treatment can substantially lower costs without patients having to reduce quality or length of life. For example, male patients can reduce costs by $4,573 on average per patient when using the optimal guidelines over the U.S. and international guidelines. For female patients the average per patient savings increases to $7,378. When the population level savings are considered, yearly cost savings of optimal treatment over the U.S. guidelines could

be as much as $1.4 billion while increasing lifetime QALYs by more than 40,000 QALYs.

The main limitation of the multiple medication MDP in Chapter 3 stemmed from discretization of the state space representing the patient's blood pressure and cholesterol. In Chapter 4, we explored the use of ADP methods to solve the continuous-state MDP. We focused on the use of a finite-state MDP (similar to the MDP presented in Chapter 3) and the use of basis function approximation of the value function.

From the results presented in Chapter 4, we found no significant differences among the ADP methods implemented and the U.S. guidelines, except for when very large medical QALY decrements were assumed. We conclude that the expected QALYs before an event is likely not sensitive to the ADP method used. When considering the ease of implementation, ADP approach 1 with policy mapping is a more simple approach than the other more complex basis function approximation methods. ADP 1 may be the overall best approach considering the simplicity of implementation and the fact that the ADP 1 expected event-free QALYs were within 0.1% of the best ADP approach in all experiments conducted.

In the second model presented in Chapter 5, we use an MDP to determine the optimal timing of adherence-improving interventions for patients that have already begun medical treatment. This model considers the use of AAS, interventions based on a patient's estimated adherence using his or her pharmacy claims data. We compare the optimal results for adherence-improving interventions to IAS, the use of periodically occurring interventions that are not dependent on a patient's adherence state. AAS results in significant savings over yearly IAS when imperfect interventions are considered; event-free LYs could be increased and costs could be decreased using AAS. AAS provides the majority of its benefits by identifying patients that already have HIGH adherence and would be able to forgo interventions to improve adherence.

The work in this dissertation suggests a number of avenues for further research. For the multiple medication model in Chapter 3, we could extend the model to include additional medications and treatment for other related risk factors. For example, aspirin could be added to the model since aspirin helps reduce the risk of stroke and CHD events. In addition, the model

could be extended to manage HbA1c with the use of glucose control medications. With respect to the benefit of coordinated treatment, open questions still remain. Is it possible to quantify the benefit of coordination of treatment in general? Is it more important to coordinate risk factors that affect the same type of event than risk factors for unrelated events? For example, it may be possible to show that coordinating treatment for cholesterol and blood pressure medications is more important than coordinating cancer treatment and the use of cholesterol medications.

There are many possible future directions for the ADP solution methods presented in Chapter 4. One natural future experiment would be to compare the solution of our ALP to the solution of an ALP found using one of the other commonly used methods for achieving a tractable number of constraints, such as constraint sampling or column generation. Use of these methods would allow us to provide bounds on the error associated with the approximations. In addition we plan to use other traditional sets of basis functions such as Legendre polynomials or radial basis functions.

In all of the experiments we presented in Chapter 4, we used one form of state aggregation. However, perhaps using the same basis function methods combined with other ways of aggregating the states would result in policies that would produce greater expected QALYs before an event. One additional state aggregation approach would be to use a time-dependent partition. The continuous state space in terms of LR and SBP would be subdivided into different aggregate states depending on t. Initially, cells could be grouped in the original state aggregation configuration presented in Section 4.2. Cells would then be regrouped according to some measure of "closeness" based on value function estimates. A second form of aggregation would be adaptive discretization in which an iterative procedure is used to form successively finer discretizations of the continuous state space. Another extension related to the state space would be to partition the LR and SBP state space based on the risk of stroke and CHD events. Since the optimal treatment actions for a patient in a particular LR and SBP state are likely highly dependent on the patient's probability of events, it would be a logical next step to test

the performance of guidelines based on a risk-based partitioning of the state space.

Another extension to this research would be to test the basis function methods with other objective functions, including maximizing QALYs over a patient's lifetime, maximizing rewards for QALYs minus costs of medication before an event, and maximizing rewards for QALYs minus total costs over a patient's lifetime. These objectives are associated with more complex reward functions that may have different performance than the more simple reward function used in Chapter 4.

We used approximation methods to solve the continuous-state MDP. Therefore, it would be useful to know how far the approximation is from the true optimal value function for the continuous-state problem. The generalized Jensen's inequality for a multi-variable function [54] could be used to show that the value-to-go from $t = 0$ for the finite-state MDP provides a lower bound for the value-to-go from $t = 0$ for the continuous-state MDP. In order to apply the generalized Jensen's inequality, we would need to show that the continuous-state MDP's value function is convex for each time period in terms of the patient's LR and SBP.

The intervention model presented in Chapter 5 could be extended to include stochastic health states. It could also be extended to include the optimal timing of interventions for more medications. In addition, we could consider the use of multiple interventions with varying degrees of effectiveness. These extensions to the model would improve how closely the model matches reality. Finally, if the model was extended to include the use of interventions after a patient has had a nonfatal event, we could model the effect of health shocks, including strokes and CHD events, on a patient's adherence to treatment.

Finally, this dissertation provides optimal treatment models that are presented in the context of treatment of diabetes patients. However the models are generalizable and could be applied to the optimal timing of any medical interventions for all types of patients. In general, we have shown that optimal treatment can reduce costs of medication and treatment of events without reducing a patient's quality or length of life. General insights about optimal medical treatment that could be applied to treatment of other types of diseases can be gained from the models,

theoretical insights, and experimental results presented in this dissertation.

REFERENCES

[1] Nationwide inpatient sample. Technical report, Healthcare Cost and Utilization Project, 2006.

[2] Bureau of Labor Statistics Handbook of Methods. Technical report, Bureau of Labor Statistics, 2007.

[3] National Vital Statistics Reports. Technical Report 5, National Center for Health Statistics, 2007.

[4] Red book: 2009 edition. Technical report, Thomson Healthcare, Inc., 2009.

[5] D. Adelman. Dynamic Bid Prices in Revenue Management. *Operations Research*, 55(4):647–661, 2007.

[6] O. Alagoz, L.M. Maillart, A.J. Schaefer, and M.S. Roberts. ADP: Goals, Opportunities and Principles. Technical report, National Science Foundation, 2002.

[7] O. Alagoz, L.M. Maillart, A.J. Schaefer, and M.S. Roberts. The Optimal Timing of Living-Donor Liver Transplantation. *Management Science*, 50(10):1420–1430, 2004.

[8] O. Alagoz, L.M. Maillart, A.J. Schaefer, and M.S. Roberts. Choosing Among Living-Donor and Cadaveric Livers. *Management Science*, 53(11):1702–1715, 2007.

[9] M.Q. Anderson. Monotone Optimal Preventive Maintenance Policies for Stochastically Failing Equipment. *Naval Research Logistics Quarterly*, 28(3):347–358, 1981.

[10] S. Antonopoulos. Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III) Final Report. *Circulation*, 106:3143–3421, 2002.

[11] M.J. Armstrong. Age Repair Policies for the Machine Repair Problem. *European Journal of Operational Research*, 138:127–141, 2002.

[12] R.E. Barlow and F. Proshan. *Mathematical Theory of Reliability*. Wiley, 1965.

[13] R. Bellman and S. Dreyfus. Functional Approximations and Dynamic Programming. *Rand Corporation*, pages 247–251, 1959.

[14] J.S. Benner, R.J. Glynn, H. Mogun, P.J. Neumann, M.C. Weinstein, and J. Avorn. Long-term persistence in use of statin therapy in elderly patients. *JAMA-Journal of the American Medical Association*, 288(4):455–461, 2002.

[15] O.K. Bhattacharyya, B.R. Shah, and G.L. Booth. Management of Cardiovascular Disease in Patients with Diabetes: the 2008 Canadian Diabetes Association Guidelines. *Canadian Medical Association Journal*, 179(9):920–926, 2008.

[16] D.A. Butler. A Hazardous-Inspection Model. *Management Science*, 25(1):79–89, 1979.

[17] P.A. Caetano, J.M.C. Lam, and S.G. Morgan. Toward a Standard Definition and Measurement of Persistence with Drug Therapy: Examples from Research on Statin and Antihypertensive Utilization. *Clinical Therapeutics*, 28:1411–1424, 2006.

[18] CDC. National Vital Statistics Reports, 2007.

[19] Centers for Disease Control, http://www.cdc.gov/diabetes/pubs/factsheet11.htm. *National Diabetes Fact Sheet*, 2011.

[20] P.S. Chan, B.K. Nallamothu, H.S. Gurm, R.A. Hayward, and S. Vijan. Incremental benefit and cost-effectiveness of high-dose statin therapy in high-risk patients with coronary artery disease. *Circulation*, 115:2398–2409, 2007.

[21] J. Chhatwal, O. Alagoz, and E.S. Burnside. Optimal Breast Biopsy Decision-Making Based on Mammographic Features and Demographic Factors. *Operations Research*, 58(6):1577–1591, 2010.

[22] A.V. Chobanian, G.L. Bakris, H.R. Black, W.C. Cushman, L.A. Green, and et al. The Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure. *Hypertension: Am Heart Assoc.*, 42:1206–1252, 2003.

[23] P. Clarke, A. Gray, and R. Holman. Estimating utility values for health states of type 2 diabetic patients using the EQ-5D (UKPDS 62). *Medical Decision Making*, 22(4):340–349, 2002.

[24] J.I. Cleeman, S.M. Grundy, D. Becker, L.T. Clark, R.S. Cooper, M. A. Denke, W.J. Howard, D.B. Hunninghake, D.R. Illingworth, R.V. Luepker, P. McBride, J.M. McKenney, R.C. Pasternak, N.J. Stone, L. Van Horn, H.B. Brewer, N.D. Ernst, D. Gordon, D. Levy, B. Rifkind, J.E. Rossouw, P. Savage, S.M. Haffner, D.G. Orloff, M.A. Proschan, J.S. Schwartz, C.T. Sempos, S.T. Shero, and E.Z. Murray. Executive summary of the Third Report of the National Cholesterol Education Program (NCEP) expert panel on detection, evaluation, and treatment of high blood cholesterol in adults (Adult Treatment Panel III). *JAMA-Journal of the American Medical Association*, 285(19):2486–2497, 2001.

[25] H.M. Colhoun, D.J. Betteridge, P.N. Durrington, G.A. Hitman, H.A.W. Neil, S.J. Livingstone, M.J. Thomason, M.I. Mackness, V. Charlton-Menys, and J.H. Fuller. Primary prevention of cardiovascular disease with atorvastatin in type 2 diabetes in the Collaborative Atorvastatin Diabetes Study (CARDS): multicentre randomised placebo-controlled trial. *Lancet*, 364(9435):685–696, 2004.

[26] D.M. Cutler and W. Everett. Thinking Outside the Pillbox - Medication Adherence as a Priority for Health Care Reform. *The New England Journal of Medicine*, 362:1553–1555, 2010.

[27] G. Danaei, A.B. Friedman, S. Oza, C.J.L. Murray, and M. Ezzati. Diabetes Prevalence and Diagnosis in U.S. States: Analysis of Health Surveys. *Population Health Metrics*, 7:16, 2009.

[28] D.P. De Farias and B. Van Roy. The Linear Programming Approach to Approximate Dynamic Programming. *Operations Research*, 51(6):850–865, 2003.

[29] D.P. De Farias and B. Van Roy. On Constraint Sampling in the Linear Programming Approach to Approximate Dynamic Programming. *Mathematics of Operations Research*, 29(3):462–478, 2004.

[30] B.T. Denton, M. Kurt, N.D. Shah, S.C. Bryant, and S.A. Smith. Optimizing the Start Time of Statin Therapy for Patients with Diabetes. *Medical Decision Making*, 29:351–367, 2009.

[31] Doris Duke Charitable Foundation, http://www.ddcf.org/Medical-Research/Program-Strategies/African-Health-Research/Operations-Research/. *African Health Research: Operations Research*, 2011.

[32] W. Duckworth, C. Abraira, T. Moritz, D. Reda, N. Emanuele, P.D. Reaven, F.J. Zieve, J. Marks, S.N. Davis, R. Hayward, S.R. Warren, S. Goldman, M. McCarren, M.E. Vitek, W.G. Henderson, and G.D. Huang. Glucose Control and Vascular Complications in Veterans with Type 2 Diabetes. *The New England Journal of Medicine*, 360:129–139, 2009.

[33] S.R. Eastaugh and M.E. Hatcher. Improving Compliance Among Hypertensives: A Triage Criterion with Cost-Benefit Implications. *Medical Care*, 20(10):1001–1017, 1982.

[34] R.C. Eastman, J.C. Javitt, W.H. Herman, E.J. Dasbach, A.S. Zbrozek, F. Dong, and et al. Model of Complications of NIDDM: I. Model Construction and Assumptions. *Diabetes Care*, 20(5):725–734, 1997.

[35] D. M. Eddy and L. Schlessinger. Archimedes - A Trial-validated Model of Diabetes. *Diabetes Care*, 26(11):3093–3101, 2003.

[36] E. Erkut, A. Ingolfsson, and G. Erdogan. Ambulance Deployment for Maximum Survival. *Naval Research Logistics*, 55:42–58, 2008.

[37] C. Evans, M. Tavakoli, and B. Crawford. Use of Quality Adjusted Life Years and Life Years Gained as Benchmarks in Economic Evaluations: A Critical Appraisal. *Health Care Management Science*, 7(1):43–49, 2004.

[38] A.A. Fisher and J.R. Foreit. Designing HIV/AIDS Intervention Studies: An Operations Research Handbook. Technical report, Population Council, 2002.

[39] O.H. Franco, E.W. Steyerberg, F.B. Hu, J. Mackenbach, and W. Nusselder. Associations of Diabetes Mellitus With Total Life Expectancy and Life Expectancy With and Without Cardiovascular Disease. *Arch Intern Med*, 167:1145–1151, 2007.

[40] W.T. Friedewald, R.I. Levy, and D.S. Fredrickson. Estimation of the Concentration of Low-Density Lipoprotein Cholesterol in Plasma, Without Use of the Preparative Ultracentrifuge. *Clinical Chemistry*, 18:499–502, 1972.

[41] M.R. Gold, J.E. Siegel, L.B. Russell, and M.C. Weinstein, editors. *Cost-Effectiveness in Health and Medicine*. Oxford University Press, 1996.

[42] M.R. Gold, D. Stevenson, and D.G. Fryback. HALYs and QALYs and DALYs, Oh My: Similarities and Differences in Summary Measures of Population Health. *Annu Rev Public Health*, 23:115–134, 2002.

[43] A.B. Goldfine, S. Kaul, and W.R. Hiatt. Fibrates in the Treatment of Dyslipidemias — Time for a Reassessment. *New England Journal of Medicine*, 365(6):481–484, 2011.

[44] C. A. Gorman, B. R. Zimmerman, S. A. Smith, S. F. Dinneen, J. B. Knudsen, D. Holm, B. Jorgensen, S. Bjornsen, K. Planet, P. Hanson, and R. A. Rizza. DEMS - a second generation diabetes electronic management system. *Computer Methods and Programs in Biomedicine*, 62(2):127–140, 2000.

[45] A. Gosavi. Reinforcement Learning: A Tutorial Survey and Recent Advances. *INFORMS Journal on Computing*, 21(2):178–192, 2009.

[46] I. Graham. European Guidelines on Cardiovascular Disease Prevention in Clinical Practice: Executive Summary. *Atherosclerosis*, 194:1–45, 2007.

[47] The ADVANCE Collaborative Group. Intensive Blood Glucose Control and Vascular Outcomes in Patients with Type 2 Diabetes. *The New England Journal of Medicine*, 358:2560–72, 2008.

[48] P. Harris, B. Joyner, P. Phillips, and C. Webster. Diabetes Management in General Practice. *Diabetes Australia Publication NP*, 1005, 2009.

[49] M. He, L. Zhao, and W.B. Powell. Optimal Control of Dosage Decisions in Controlled Ovarian Hyperstimulation. *Ann Oper Res*, 178:223–245, 2010.

[50] M. He, L. Zhao, and W.B. Powell. Approximate Dynamic Programming Algorithms for Optimal Dosage Decisions in Controlled Ovarian Hyperstimulation. In *Health Systems Seminar, Department of Industrial and Systems Engineering*. North Carolina State University, 2011.

[51] P.M. Ho, D.J. Magid, F.A. Masoudi, D.L. McClure, and J.S. Rumsfeld. Adherence to Cardioprotective Medications and Mortality among Patients with Diabetes and Ischemic Heart Disease. *BMC Cardiovascular Disorders*, 6(48), 2006.

[52] W.J. Hopp and S.-C. Wu. Machine Maintenance with Multiple Maintenance Actions. *IIE Transactions*, 22(3):226–233, 1990.

[53] K.W. Hsih. *Optimal Dosing Applied to Glycemic Control for Type 2 Diabetes*. PhD thesis, Princeton University, 2010.

[54] C.C. Huang, W.T. Ziemba, and A. Ben-Tal. Bounds on the Expectation of a Convex Function of a Random Variable: With Applications to Stochastic Programming. *Operations Research*, 25(2):315–325, 1977.

[55] E.S. Huang, M. O'Grady, A. Basu, and J.C. Capretta. Projecting the Future Diabetes Population Size and Related Costs for the U.S. *Diabetes Care*, 32:2225–2229, 2009.

[56] A.K.S. Jardine and J.A. Buzacott. Equipment Reliability and Maintenance. *European Journal of Operational Research*, 19:285–296, 1985.

[57] Joint British Societies 2. JBS 2: Joint British Societies' Guidelines on Prevention of Cardiovascular Disease in Clinical Practice. *Heart*, 91:1–52, 2005.

[58] L.P. Kaelbling, M.L. Littman, and A.W. Moore. Reinforcement Learning: A Survey. *Journal of Artificial Intelligence Research*, 4:237–285, 1996.

[59] W.D. Kelton, R.P Sadowski, and N.B. Swets. *Simulation with Arena*. McGraw-Hill, 2010.

[60] M. Klein. Inspection-Maintenance-Replacement Schedules under Markovian Deterioration. *Management Science*, 9(1):25–32, 1962.

[61] V. Kothari, R.J. Stevens, A.I. Adler, I.M. Stratton, S.E. Manley, H.A. Neil, and R.R. Holman. UKPDS 60 - Risk of Stroke in Type 2 Diabetes Estimated by the United Kingdom Prospective Diabetes Study Risk Engine. *Stroke*, 33(7):1776–1781, 2002.

[62] H.C. Kung, D.L. Hoyert, J.Q. Xu, and S.L. Murphy. Deaths: Final Data for 2005. *National Vital Statistics Reports*, 56(10), 2008.

[63] M. Kurt, B.T. Denton, A. Schaefer, N. Shah, and S. Smith. The Structure of Optimal Statin Initiation Policies for Patients with Type 2 Diabetes. *IIE Transactions on Healthcare*, 1(1):49–65, 2011.

[64] C.P. Lee, G.M. Chertow, and S.A. Zenios. Optimal Initiation and Management of Dialysis Therapy. *Operations Research*, 56(6):1428–1449, 2008.

[65] L.M. Maillart, J.S. Ivy, S. Ransom, and K. Diehl. Assessing Dynamic Breast Cancer Screening Policies. *Operations Research*, 56(6):1411–1427, 2008.

[66] J.E. Mason, D.A. England, B.T. Denton, S.A. Smith, M. Kurt, and N.D. Shah. Optimizing Statin Treatment Decisions for Diabetes Patients in the Presence of Uncertain Future Adherence. *Medical Decision Making*, 32(1):154 – 166, 2012.

[67] M.S. Maxwell, M. Restrepo, S.G. Henderson, and H. Topaloglu. Approximate Dynamic Programming for Ambulance Redeployment. *INFORMS Journal on Computing*, 22(2):266 – 281, 2010.

[68] J.T. McCall. Maintenance Policies for Stochastically Failing Equipment: A Survey. *Management Science*, 11(5):493–524, 1965.

[69] T. Nakagawa. Sequential Imperfect Preventive Maintenance Policies. *IEEE Transactions on Reliability*, 37(3):295–298, 1988.

[70] National Committee for Quality Assurance. *HEDIS 2007 Volume 2 Techinical Specifications*, 2007.

[71] National Institutes of Health, http://obssr.od.nih.gov/scientific_areas/health_behaviour/adherence/adherenceresearchnetwork.aspx. *Adherence Research Network*, 2011.

[72] S. Okada, T. Morimoto, H. Ogawa, M. Kanauchi, M. Nakayama, S. Uemura, N. Doi, H. Jinnouchi, M. Waki, H. Soejima, M. Sakuma, and Y. Saito. Differential Effect of Low-Dose Aspirin for Primary Prevention of Atherosclerotic Events in Diabetes Management. *Diabetes Care*, 34(6):1277–1283, 2011.

[73] L. Osterberg and T. Blaschke. Adherence to Medication. *New England Journal of Medicine*, 353:487–497, 2005.

[74] R.S. Parker, F.J. Doyle, and N.A. Peppas. The Intravenous Route to Blood Glucose Control. *IEEE Engineering in Medicine and Biology*, 20(1):65–73, 2001.

[75] J. Patrick, M.L. Puterman, and M. Queyranne. Dynamic Multipriority Patient Scheduling for a Diagnostic Resource. *Operations Research*, 56(6):1507–1525, 2008.

[76] H. Pham and H. Wang. Imperfect Maintenance. *European Journal of Operational Research*, 94:425–438, 1996.

[77] K.A. Phillips, M.G. Shlipak, P. Coxson, P.A. Heidenreich, M.G.M. Hunink, P.A. Goldman, L.W. Williams, M.C. Weinstein, and L. Goldman. Health and Economic Benefits of Increased Beta-blocker use Following Myocardial Infarction. *JAMA-Journal of the American Medical Association*, 284(21):2748–2754, 2000.

[78] W.P. Pierskalla and J.A. Voelker. A Survey of Maintenance Models: The Control and Surveillance of Deteriorating Systems. *Naval Research Logistics Quarterly*, 23:353–388, 1976.

[79] W.B. Powell. Perspectives of Approximate Dynamic Programming. *Annals of Operations Research*, pages DOI: 10.1007/s10479–012–1077–6.

[80] W.B. Powell, editor. *Approximate Dynamic Programming: Solving the Curses of Dimensionality*. Wiley, 2007.

[81] M.L. Puterman. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. John Wiley & Sons, Inc, Hoboken, New Jersey, 1994.

[82] K.L. Rascati. The $64,000 Question - What Is a Quality-Adjusted Life-Year Worth? *Clinical Therapeutics*, 28(7):1042–1043, 2006.

[83] J.N. Rasmussen, A. Chong, and D.A. Alter. Relationship Between Adherence to Evidence-based Pharmacotherapy and Long-term Mortality after Acute Myocardial Infarction. *Jama-Journal of the American Medical Association*, 297(2):177–186, 2007.

[84] P.M. Ridker, E. Danielson, F.A.H. Fonseca, J. Genest, A.M. Gotto, J.J.P. Kastelein, W. Koenig, P. Libby, A.J. Lorenzatti, J.G. MacFadyen, B.G. Nordestgaard, J. Shepherd,

J.T. Willerson, and R.J. Glynn. Rosuvastatin to Prevent Vascular Events in Men and Women with Elevated C-Reactive Protein. *New England Journal of Medicine*, 359:2195–2207, 2008.

[85] H. Robbins and S. Monro. A Stochastic Approximation Method. *Ann. Math. Statist.*, 22(3):400–407, 1951.

[86] M.W. Russell, D.M. Huse, S. Drowns, E.C. Hamel, and S.C. Hartz. Direct Medical Costs of Coronary Artery Disease in the United States. *American Journal of Cardiology*, 81(9):1110–1115, 1998.

[87] I.O. Ryzhov, K.W. Hsih, and W.B. Powell. Optimal Learning Applied to Glycemic Control in Type 2 Diabetes. In *INFORMS Healthcare Conference*, 2011.

[88] P.J. Schweitzer and A. Seidmann. Generalized Polynomial Approximations in Markovian Decision Processes. *Journal of Mathematical Analysis and Applications*, 110:568–582, 1985.

[89] N.D. Shah, J.E. Mason, M. Kurt, B.T. Denton, A.J. Schaefer, V.M. Montori, and S.A. Smith. Comparative Effectiveness of Guidelines for the Management of Hyperlipidemia and Hypertension for Type 2 Diabetes Patients. *PloS ONE*, 6(1):e16170, 2011.

[90] S.M. Shechter, M.D. Bailey, A.J. Schaefer, and M.S. Roberts. The Optimal Time to Initiate HIV Therapy Under Ordered Health States. *Operations Research*, 56(1):20–33, 2008.

[91] Y.S. Sherif and M.L. Smith. Optimal Maintenance Models for Systems Subject to Failure – A Review. *Naval Research Logistics Quarterly*, 28(1):47–74, 1981.

[92] A.A. Sherstov and P. Stone. Function Approximation via Tile Coding: Automating Parameter Choice. In *Abstraction, Reformulation, and Approximation*, 2005.

[93] C.M. Shetty and R.W. Taylor. Solving Large-Scale Linear Programs by Aggregation. *Computers and Operations Research*, 14(5):385–393, 1987.

[94] V. Snow, M.D. Aronson, E.R. Hornbake, C. Mottur-Pilson, and K.B. Weiss. Lipid Control in the Management of Type 2 Diabetes Mellitus: A Clinical Practice Guideline from the American College of Physicians. *Annals of Internal Medicine*, 140(8):644–649, 2004.

[95] M.C. Sokol, K.A. McGuigan, R.R. Verbrugge, and R.S. Epstein. Impact of Medication Adherence on Hospitalization Risk and Healthcare Cost. *Medical Care*, 43(6):521–530, 2005.

[96] R.J. Stevens, R.L. Coleman, A.I. Adler, I.M. Stratton, D.R. Matthews, and R.R. Holman. Risk Factors for Myocardial Infarction Case Fatality and Stroke Case Fatality in Type 2 Diabetes (UKPDS 66). *Diabetes Care*, 27(1):201–207, 2004.

[97] R.J. Stevens, V. Kothari, A.I. Adler, I.M. Stratton, and R.R. Holman. The UKPDS Risk Engine: a Model for the Risk of Coronary Heart Disease in Type 2 Diabetes (UKPDS 56). *Clinical Science*, 101(6):671–679, 2001.

[98] R.S. Sutton and A.G. Barto. *Reinforcement Learning: An Introduction*. MIT Press, 1998.

[99] T.O. Tengs and T.H. Lin. A Meta-analysis of Quality-of-life Estimates for Stroke. *Pharmoeconomics*, 21(3):191–200, 2003.

[100] T.O. Tengs, M. Yu, and E. Luistro. Health-related Quality of Life After Stroke - A Comprehensive Review. *Stroke*, 32(4):964–971, 2001.

[101] T. Thom, N. Haase, W. Rosamond, V.J. Howard, J. Rumsfeld, T. Manolio, Z.J. Zheng, K. Flegal, C. O'Donnell, S. Kittner, D. Lloyd-Jones, D.C. Goff, Y.L. Hong, R. Adams, G. Friday, K. Furie, P. Gorelick, B. Kissela, J. Marler, J. Meigs, V. Roger, S. Sidney, P. Sorlie, J. Steinberger, S. Wasserthiel-Smoller, M. Wilson, and P. Wolf. Heart Disease and Stroke Statistics – 2006 Update – A Report from the American Heart Association Statistics Committee and Stroke Statistics Subcommittee. *Circulation*, 113(6):E85–E151, 2006.

[102] J.W. Timbie, R.A. Hayward, and S. Vijan. Diminishing Efficacy of Combination Therapy, Response-heterogeneity, and Treatment Intolerance Limit the Attainability of Tight Risk Factor Control in Patients with Diabetes. *Health Services Research*, 45(2):437–56, 2010.

[103] The Action to Control Cardiovascular Risk in Diabetes Study Group. Effects of Intensive Glucose Lowering in Type 2 Diabetes. *The New England Journal of Medicine*, 358:2545–59, 2008.

[104] M.A. Trick and S.E. Zin. A Linear Programming Approach to Solving Stochastic Dynamic Programs. *working paper*, 1993.

[105] M.A. Trick and S.E. Zin. Spline Approximations to Value Functions: A Linear Programming Approach. *working paper*, 1995.

[106] J. Tsevat, L. Goldman, J.R. Soukup, G.A. Lamas, K.F. Connors, C.C. Chapin, and T.H. Lee. Stability of Time-tradeoff Utilities in Survivors of Myocardial Infarction. *Medical Decision Making*, 13(2):161–165, 1993.

[107] U.S. Census Bureau. 2010 Census Summary File 1 - United States. Technical report, 2011.

[108] U.S. Department of Health & Human Services, `http://healthit.hhs.gov/portal/server.pt/community/healthit\_hhs\_gov\_\_meaningful\_use\_announcement/2996`. *Electronic Health Records and Meaningful Use*, 2011.

[109] C. Valdez-Flores and R.M. Feldman. A Survey of Preventative Maintenance Models for Stochastically Deteriorating Single-Unit Systems. *Naval Research Logistics*, 36:419–446, 1989.

[110] S. Vijan and R.A. Hayward. Pharmacologic Lipid-Lowering Therapy in Type 2 Diabetes Mellitus: Background Paper for the American College of Physicians. *Ann Intern Med*, 140:650–658, 2004.

[111] C.J.C.H. Watkins. *Learning from Delayed Rewards*. PhD thesis, King's College, 1989.

[112] C.J.C.H. Watkins and P. Dayan. Technical Note: Q-Learning. *Machine Learning*, 8:279–292, 1992.

[113] A.J. Weymiller, V.M. Montori, L.A. Jones, A. Gafni, G.H. Guyatt, S.C. Bryant, T.J.H. Christianson, R.J. Mullan, and S.A. Smith. Helping Patients With Type 2 Diabetes Mellitus Make Treatment Decisions. *Archives of Internal Medicine*, 167:1076–1082, 2007.

[114] J.C. Wright and M. C. Weinstein. Gains in Life Expectancy from Medical Interventions – Standardizing Data on Outcomes. *New England Journal of Medicine*, 339:380–386, 1998.