12-2021

# Data-Driven Statin Initiation Evaluation and Optimization for Prediabetes Population

Muhenned A. Abdulsahib
*University of Arkansas, Fayetteville*

Data-Driven Statin Initiation Evaluation and Optimization for Prediabetes Population

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy in Industrial Engineering

by

Muhenned A. Abdulsahib
University of Baghdad
Bachelor of Science in Mathematics, 1999
University of Arkansas
Master of Science in Mathematics, 2014

December 2021
University of Arkansas

Shengfan Zhang, PhD
Dissertation Director

Haitao Liao, PhD
Committee Member

Chase E. Rainwater, PhD
Committee Member

Qingyang Zhang, PhD
Committee Member

**Abstract**

This dissertation develops quantitative models to support medical decision making of statin initiation considering the uncertainty in disease progression for prediabetes patients. A mathematical model is built to help medical decision-makers take action of statin initiation under uncertainty in future prediabetes progressions. The association between cholesterol drug use, such as statin, and elevating glucose level attracted considerable amounts of attention in the literature. Statin effects on glucose vary with respect to different levels of glucose. The first chapter of this dissertation introduces the problem and an overview of the tools that will be used to solve it. In the second chapter of this dissertation, we use quantile regression to investigate the statin effects on different glucose quantiles. The third chapter is devoted to address the problem of estimating transition probabilities for the prediabetes populations for different levels of lipid ratios (i.e., ratios of total cholesterol to high-density lipoprotein) from cross-sectional data. We also show the risk of prediabetes as a function of age using a Bayesian approach. These parameters are used in the implementation of the Markov Decision Process (MDP) to estimate the best point in time, i.e., age to start statin treatment, which is discussed in the fourth chapter of the dissertation. Parameter estimation plays a crucial role in our model. Therefore, our fifth problem to tackle is to provide the optimal policy for statin initiation by considering stochastic progression under uncertainty for lipid ratio transition probabilities. A robust MDP problem is formulated and an optimal policy for statin initiation is provided. Structural properties of the robust MDP problem are proved. Additionally, sufficient conditions for the existent of robust MDP are introduced.

**Acknowledgments**

First, I would like to thank my advisor Dr. Shengfan Zhang for her unlimited patience, constant support and invaluable guidance through my academic research. I owe a special thanks to Dr. Haitao Liao, Dr.Chase E. Rainwater and Dr. Qingyang Zhang for agreeing to be committee members. I appreciate the help of the staff of the industrial engineering department who made my stay comfortable at University of Arkansas.

I would like to thank the chair of the industrial engineering department, Dr. Edward Pohl for his support during my stay at the department.

Finally, I am indebted forever to my parents Abdulameer and Kidega, my brothers and sisters, and my wife and daughters for their unconditional love.

**Dedication**

I dedicate this dissertation to the memory of my father for his constant support. I also dedicate it to my mother for her unconditional love. Lastly, I dedicate it to my wife Asmaa Sadoon, and my daughters Zeinib, Rukayyah, and Maryam.

**Contents**

**List of Figures**

# Chapter 1

## Introduction

Disease progressions have long been represented using Markov chains by the medical community. Because patients react to different treatments in different ways, it is crucial to develop a policy that personalizes treatment plans. Markov decision processes (MDPs) are a valuable tool used to tackle complex problems in medical decision-making in which decision-makers are required to take a sequential decision under uncertainty. Through using the MDP tool, decision-makers can find the best action that maximizes the future rewards under uncertainty. Recently, researchers have implemented MDP to optimize treatment-related decisions for diabetes and related conditions. For example, Denton et al. [24] developed an MDP model to find the optimal time (age) to start statin treatment in the diabetes population, which considers protection from cardiovascular diseases and cost of treatment. They found that statin treatment initiation is sensitive to the cholesterol levels, gender, and risk models that are used to estimate the CVD risk. Mason et al. [61] studied optimizing the starting time for statin initiation assuming uncertain future adherence. The optimal time for interventions to improve adherence is also studied using an MDP model [60]. In another study, the optimal policy for taking the A1C test, measuring the glucose level in a three-month period, was studied using an MDP model from which the authors recommended taking the test quarterly [51]. Zhang et al. [97] developed an MDP model to optimize the best sequence for glucose medications that are used to reduce glucose levels. Although the studies mentioned above focused on diabetes and related conditions, MDP is a powerful tool used to find the optimal policy in different complicated problems, see [29] and references therein.

This dissertation focuses on prediabetes, which is considered one of the leading chronic conditions in the U.S. [34]. Prediabetes is defined as a health condition in which either impaired fasting glucose (IFG) with fasting plasma glucose is greater than 100 mg/dl and less than 126 mg/dl or impaired glucose tolerance (IGT) with 2-hour plasma glucose is greater than 140 mg/dL and less than 199 mg/dL, i.e., the state in which the criteria for diabetes mellitus are not all satisfied [72]. For every new-onset diabetes patient, they must have visited a prediabetes state before reaching a diabetes state [59]. Better management of prediabetes can delay or prevent the progression to the diabetes state, for example, lifestyle intervention may reduce the risk of progression to the diabtest state by 20% [88]. Therefore, the prediabetes individuals need to get special attention from medical decision-makers.

There are approximately 3 million newly diagnosed prediabetes cases each year according to the American Diabetes Association. More specifically, about 38% of adults in the U.S. have prediabetes [62]. Approximately 16.2% of American adults, aged above 20, use statin [40] to reduce cholesterol levels, but statin has side effects such as elevating the glucose levels [78] and/or body mass index (BMI) [84]. The 2013 cholesterol guideline by the American College of Cardiology/American Heart Association (ACC/AHA) recommended statin treatment initiation for patients with a 10-year risk of atherosclerotic vascular disease of more than 7.5%. It has been shown through simulation that the ACC/AHA guideline is expected to treat more people, save more lives, and cost less if the dis-utility associated with pill burden is small [43]. Pandya et al. [68] presented a cost-effectiveness analysis for statin use in which they found that the dis-utility associated with taking a pill daily, statin price, and the risk of statin-induced diabetes negatively impacted the cost-effectiveness of statin.

There is a strong positive relationship between elevated glucose levels and cardiovascular disease risk (CVD) [80], [56]. Defronzo et al. [23] reported individuals with prediabetes

are at high risk of diabetes and adverse cardiovascular event, stroke, and cardiovascular deaths in later life. In addition, there is a positive relationship between diabetes and impaired glucose metabolism with an increased risk of CVD mortality [19]. Braun et al. [9] suggested using glucose values instead of total cholesterol values in estimating the overall CVD mortality risk. Therefore, it is important to consider the two risk factors for CVD when a drug is recommended to reduce cholesterol levels. In other words, raising glucose levels and decreasing cholesterol levels may increase CVD risks.

The west of Scotland coronary prevention study (WOSCOPS ) presented that statin use reduces the risk of new-onset diabetes by 30% [32]. On the other hand, other studies found that using statin increases the risk of new-onset diabetes [89], [95], [21], [85], [12], [78], [76], and [93]. In addition, a recent study presented a proportional relation between the diabetes risk and length of time statin is used (hazard ratio (HR) 1.25 $<$ 1 year, HR 2.22 for $1-2$ years, and HR 2.62 $>$ 2 years) [49]. Moreover, the presence of diabetes risk factors, such as prediabetes, raises the relative risk for new-onset diabetes associated with using statin significantly [76], [13].

A multivariate quantile regression model is used to quantify the association between statin use and elevated glucose levels. This approach is used to study the effects of age, gender, body mass index on different quantiles of the conditional fasting blood glucose distribution. The benefit of this approach is that it can characterize the association between statin use and fasting blood glucose levels in a specific group, e.g., prediabetes population

Based on the above it is crucial to find a personalized decision for statin treatment in prediabetes patients. Lowering the glucose threshold for diabetic prevention gives more health benefits but raises the cost [98]. Our goal is to find the optimal policy to initiate statin treatments that considers the trade-offs between heart risks and diabetes risk for prediabetes patients. Researchers face an issue in calculating a personalize deci-

sion, which is the difficulty in estimating the exact value of the population parameters. One of these parameters is the transition probability matrix. Fortunately, we are able to estimate the transition probability matrix using the National Health and Nutrition Examination Survey (NHANES) cross-sectional survey data set. We also used Bayesian approach to estimate the risk of prediabetes using NHANES data sets.

State transition probabilities are assumed to be known with certainty in the MDP models but these probabilities are estimated from data and, hence, they are subject to uncertainty due to different reasons such as scarcity of data for some state-action pairs, measure ment errors, and environment reacts in a different way that alters state-action probabilities. The solution of dynamic problems may be different with small changes in the probability estimation. Therefore, it is crucial to come up with an optimal solution that stays optimal with small changes in the transition probabilities. In the fifth chapter, we find an optimal solution for a finite horizon Markov decision process with finite state and action spaces. A sufficient condition for the existence of optimal control limit policy is introduced.

# Chapter 2

# Quantile Regression Analysis for Glucose

## 2.1 Introduction

The association between statin use and the risk of elevating glucose levels recently has attracted researchers in the medical community [78], [13], [76], [43]. Statin use can cause elevated FBG levels [78], [76]. Both elevated cholesterol levels and elevated FBG levels are cardiovascular disease risk factors [23]. Therefore, we will investigate statin use effects on FBG levels at different quantiles of the conditional FBG distribution rather than studying the mean effect of statin use on FBG because the mean effects approach is done by averaging the effects on higher quantiles and on lower quantiles of FBG. The mean effects approach reduces the statin use effects on higher FBG levels quantiles.

Quantile regression (QR) was developed to study the relationships between predictor variables that have weak or no relationship with the mean of the response or when our interests are in the covariates effects at different quantiles of the conditional response distribution. In QR, for every quantile, there are different estimations for the regression coefficients. That is, the association relation between response and explanatory variables varies with respect to different levels of the response. Quantile regression is an important tool used to regress the dependent variable with high variance over the independent variables (predictors) [50].

On the other hand, the ordinary least squares method (OLS) is used to summarize the relationship between covariates $X$ and the conditional mean of the response. Guo et al. [42] used QR to study the association between different fasting blood glucose (FBG)

quantiles and risk factors for the Northeast Chinese population. They found that BMI has positive effects on the low and middle quantiles of FBG, but waist circumference has a positive effect on the higher quantiles of FBG. They also found that the conditional distribution of FBG varies with respect to gender. Using NHANES data (2007-2014), Shoumeng et al. [94] show there is a negative association between FBG and consumption of folic acid in male individuals with normal or high FBG. For a male individual with normal FBG, there is a negative association between FBG and calcium intake. For a female, there is a negative association between FBG and folic acid, and calcium in the individuals with normal FBG. Using QR, Hu et al. [44] studied the effects of age and gender on different FBG quantiles for adolescents aged between 12 and 20 years old in the NHANES data set. They recommended establishing standard blood glucose for adolescents that considers ages and genders. Ellerbe et al. [25] used QR to study the gestational diabetes mellitus effects on different birth weight quantiles considering maternal pregnancy, body mass index and race. Another area of QR application is health care expenditure. In terms of health condition, the frequent users of health care are very different from the average users. However, other predictors of interest for decision makers are Race/Ethnicity, gender and employment. These predictors are used to assess the difference in health care expenditure [20].

In this chapter, we use QR to study the statin effects on different glucose quantiles. This helps us to understand the effect of statin use on the normoglycemic, prediabetes, and diabetes populations. One of the advantages of using QR over OLS is QR's robust ability to account for outliers [50].

## 2.2 Quantile Regression

Consider a random variable $X$ with cumulative distribution function (CDF) $F(X)$. The $\tau$th quantile of $X$ is defined by

$$F^{-1}(\tau) = \inf\{x : F(x) \geq \tau\},$$

where $0 < \tau < 1$. Let the loss function be defined as

$$\rho_\tau(u) = u(\tau - I_{(u<0)}),$$

where $I$ is the indicator function. The quantile estimator is the value that minimizes the expected loss function

$$E\rho_\tau(X - \hat{x}) = (\tau - 1)\int_{-\infty}^{\hat{x}} (x - \hat{x})dF(x) + \tau \int_{\hat{x}}^{-\infty} (x - \hat{x})dF(x).$$

Differentiating with respect to $\hat{x}$, we get

$$0 = (\tau - 1)\int_{-\infty}^{\hat{x}} dF(x) - \tau \int_{\hat{x}}^{-\infty} dF(x) = F(\hat{x}) - \tau.$$

Due to the monotonicity of the cumulative distribution function, any solution that satisfies $\{x : F(x) = \tau\}$ is a minimum of the expected loss function.

The least-square method expresses the conditional mean of y given x as $\mu(x) = x^T\beta$ and it solves

$$\min_{\beta \in \mathcal{R}^p} \sum_{i=1}^{n} (y_i - x_i^T\beta)^2.$$

Quantile regression expresses conditional quantile function $Q_y(\tau|x) = x^T\beta(\tau)$ and solves

$$\min_{\beta\in\mathcal{R}^p} \sum_{i=1}^{n} \rho_\tau(y_i - x_i^T\beta)^2.$$

This minimization problem can be reformulated into a linear programming problem by introducing artificial and slack variables to take care of the positive and negative parts of the vector of residuals, see Eq. (1.15) [50].

### 2.2.1 Data

In this section, NHANES data is used. BMI values are classified into categories underweight, 18.5 kg/m2; normal weight, 18.5 to 25 kg/m2; overweight, 25 to 30 kg/m2; obese, 30 to 35 kg/m2; and very obese more than 35 kg/m2. The statin medications and their survey codes used in this study are LOVASTATIN : d00280, SIMVASTATIN: d00746, ROSUVASTATIN: d04851, ATORVASTATIN: d04105, PRAVASTATIN: d00348, EZIT/SIMVASTATIN: d05348, AMILop/ATROV: d05048. Table 5.1 describes the characteristic of our data set.

Table 2.1: Baseline characteristic for the study population including mean and standard deviation.

| **Population Attribute** | **Estimation (S.E.) female** | **Estimation (S.E.) male** |
|---|---|---|
| Age | 47.0 (17.1) | 47.2 ( 17.1) |
| FBG | 104(31.2) | 110 (34.5) |
| TC (No statin) | 196(41.0) | 193(40.9) |
| HDL (No statin) | 59.3(16.4) | 49.6 (14.6) |
| LDL (No statin) | 115(35.1) | 118 (35.2) |
| BMI | 29.2(7.47) | 28.3( 5.94) |
| % count | 0.523 | 0.477 |

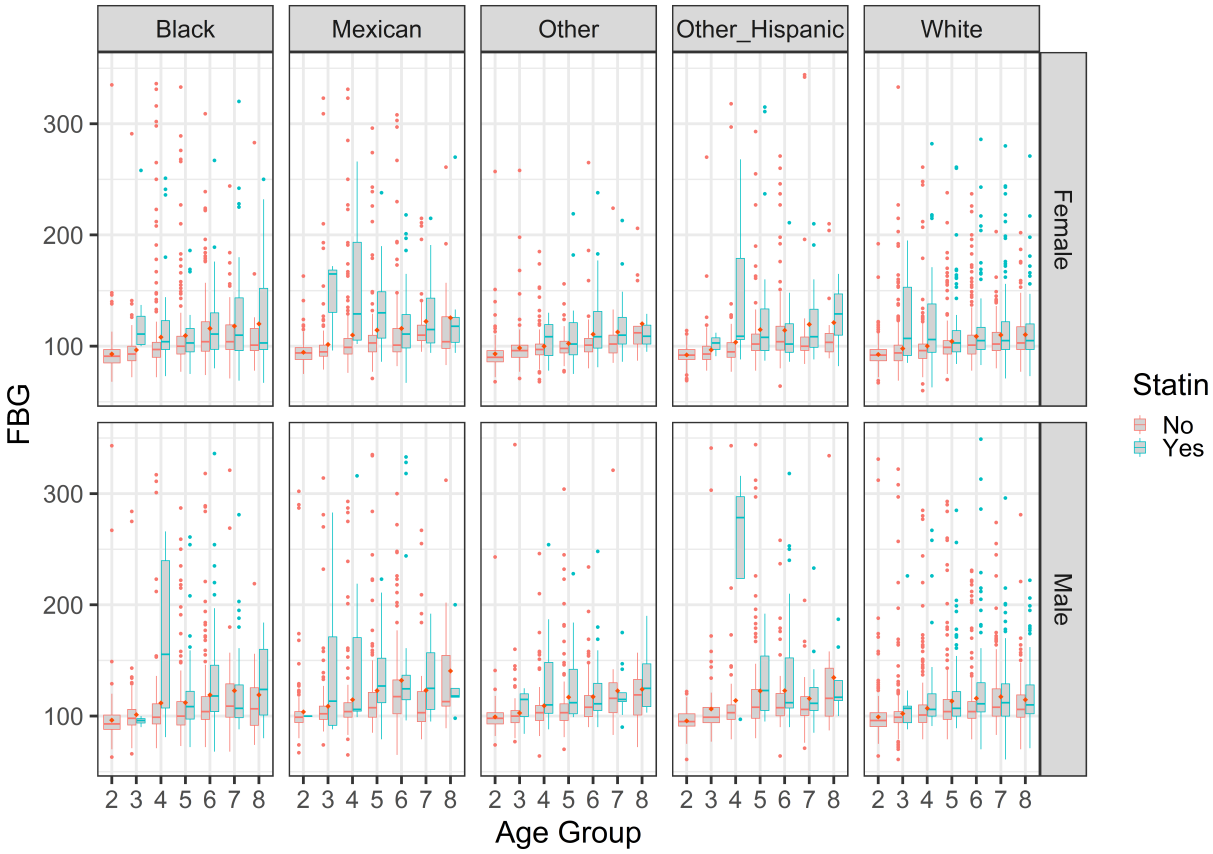


Figure 2.1: Boxplot for fasting blood glucose plotted with respect to Gender, Age, Statin uses, and Race-ethnicity (Mexican American, Other Hispanic, Non-Hispanic White, Non-Hispanic Black, Other Race - Including Multi-Racial). Age groups $2, 3, \cdots, 8$ on the x-axis represents ages from $20-29, 30-39, \cdots, 80$, respectively.

From Fig. 2.1, we can see that at younger ages cholesterol medication treatments are given only to populations with high glucose levels. After the age of 50 years, the differences in glucose levels between populations who take cholesterol medication and those who do not start to diminish until they reach a negligible value. This can be interpreted in two ways: either there is a positive correlation between glucose and cholesterol, or taking cholesterol medication can cause an elevated glucose level. However, both of these scenarios have been studied. It is shown that cholesterol medications (statins)

can cause elevating glucose levels [78], [13], [43], and there is a positive correlation between cholesterol and glucose [81].

Moreover, there is a variation in glucose distribution in terms of gender and race as a function of age. For example, Mexican females have higher variability in glucose values if compared to others. On the other hand, white populations have low variability in glucose values. In terms of age, there is a clear pattern, when age increases, glucose level increases too. In the Other Hispanic and Non-Hispanic Black groups, the statin users at age group 40 suffer from high FBG. That is, at 40 years old, these groups start to suffer from high cholesterol levels and high FBG. In the Mexican American group, the elevation in cholesterol and FBG start earlier, at the age group 30 years old. On the other hand, FBG concentrations for white males do not show a noticeable change in the FBG distribution as a function of age but show a mild increase in glucose levels. Prevalence of prediabetes in the white race and other races are higher than the rest of the Race/Ethnicity groups, but the prevalence of diabetes in the rest races is higher than the prevalence of diabetes in white and other Race/Ethnicity groups, see Fig. 2.1. This trend was observed by [63]. Variations in terms of gender and Race/Ethnicity have also been noticed in the TC level distribution. Engeda et al. [26] have noticed that the progression from ideal to intermediate TC across adulthood varied by Race/Ethnicity and sex.

### 2.2.2 Quantile Regression Equation

A multivariate quantile regression model is used to assess the characteristics of the association variability in different quantiles of the conditional distribution of the glucose levels with respect to influencing factors. The dependent variables in our model are gender, Race/Ethnicity (Mexican, Other Hispanic, Non-Hispanic White, Non-Hispanic Black, Other Race Including Multi-Racial), age, BMI, and cholesterol medication (statin)

(yes or no).

In many cases, the relationships between responses and predictor variables are complicated to the point that linear regressions are not suitable to model these relationships [10]. For example, the responses to different levels of drug doses form a nonlinear relationship. One approach used to model the nonlinear relationships is using polynomial regression. The nonlinear regression is modeled using polynomial terms of predictors.A draw back of this approach is that the degree of the polynomial forces the relationship to take a specific shape. For example, when a predictor is modeled using a quadratic term, the value of the response will approach $\pm\infty$ as values of approaches $\pm\infty$. A more robust approach used to deal with the nonlinar relationships is by modeling predictors through using spline basis expansions. This is called a spline regression, which is constructed by using different polynomials for different segments of the predictor.

Since the correlation between age and BMI is small ($\rho = 0.03$), we added both of them in the model.The regression equation is defined as follows:

$$\begin{aligned} \hat{\text{FBG}} = \beta_0 + \beta_1 \mathit{Gender} + \beta_2 \mathit{spline(Age,\ df=5)} + \beta_3 \mathit{spline(BMI,\ df=5)} + \\ \beta_4 \mathit{Statin} + \beta_5 \mathit{Race/Ethnicity} + \epsilon. \end{aligned} \tag{2.2.1}$$

Fig. 2.2 shows, at younger ages, the 90th quantile of fasting blood glucose in prediabetes statin users is higher than the 90th quantile of fasting blood glucose in non-statin users. In other words, cholesterol medications are given to prediabetes patients with higher glucose. This could be interpreted as statin use can elevate glucose levels [78], [76]. Furthermore, there is a disparity in the 90th quantile in terms of genders. The 90th quantile of glucose for females is almost constant with respect to age but the 90th

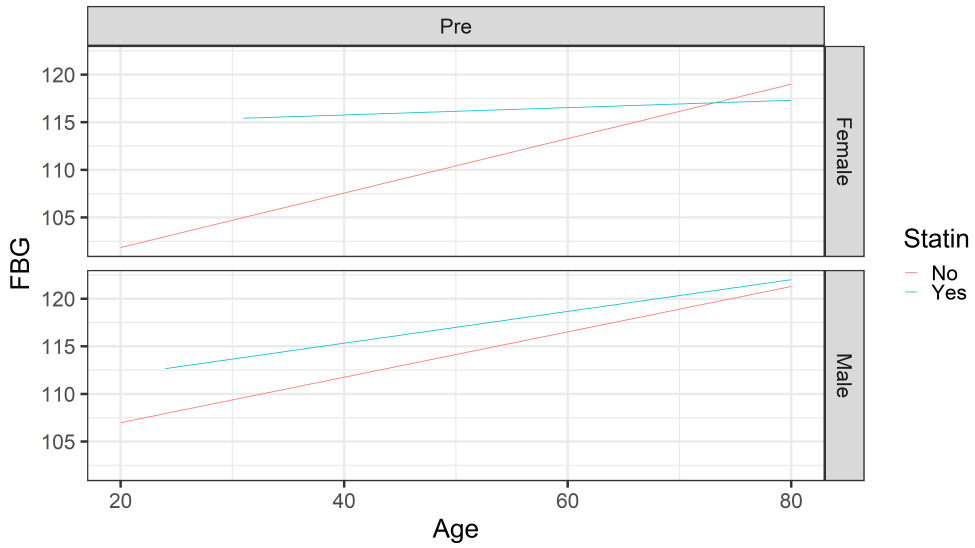


Figure 2.2: An illustration of age effects on the 90th quantile of glucose, with respect to gender and statin use, is presented for prediabetes population.

quantile of glucose for males increases with respect to age. The 90th quantile for female statin users is higher than the 90th quantile for male statin users; that is, at younger ages, statin is prescribed to females with a higher glucose level than males or statin uses have higher effects on females FBG levels than males FBG levels.

### 2.2.3 The Association Between TC and FBG

Because Statin has effect on both TC and FBG, it is crucial to study the association between TC and FBG. QR is used to study the relationship between TC and FBG. We modeled FBG as a function of TC using spline basis expansions of degree 5. The relationship between TC and fasting blood glucose levels have been studied in [87], [30], [17]. The authors found that there is a positive correlation between the two groups on

average. However, we found that the trends are different in different FBG quantiles.

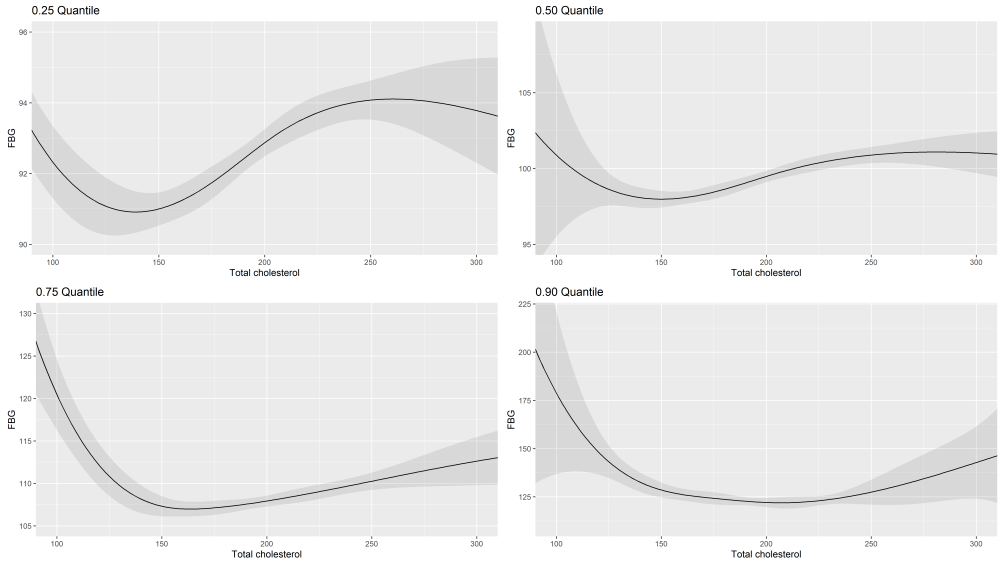


Figure 2.3: Illustration of the TC effects on glucose levels for four different quantiles of the conditional glucose distribution.

Fig. 2.3 shows TC effects on FBG at different quantiles of conditional FBG distribution. For TC levels less than around 150 mg/dL the glucose level and TC are negatively correlated. When TC values pass 160 mg/dL, the association is positively correlated for the FBG values less than 125 mg/dL. The TC effects on FBG diminishes when TC values passes 200 mg/dL. At the fourth quantile of FBG, the relationship is almost constant for TC value between 160 mg/dL and 225 mg/dL, but a positive relationship occurs after that. The lowest effects of TC on FBG occurs at TC=140 mg/dL, TC=150 mg/dL, TC=150, TC= 200 for the first, second , third, and $90th$ quantiles of FBG, respectively. This tells us that there is a disparity in the association between TC and different FBG quantiles, and studying the TC effect on the conditional mean of glucose, which is the second quantile of FBG, is not appropriate to detect this association.

## 2.3 Results

A concise summary of the model output is presented in Fig. 2.4. The marginal effects of each predictor are presented in the plot. The point estimates $\{\hat{\beta}_j(\tau) : j = 1, \ldots, 17\}$ are represented by the solid line with filled dots. The 95th percentiles are depicted by the shaded area. The resulting ordinary least squares estimates of conditional mean of glucose level may not reflect the size and nature of these effects on the lower or upper quantiles. For example, the conditional mean effects of gender (female) on the glucose level is on average about -7 mg/dL less than the conditional mean effects of gender (male), see Fig. 2.4. However, the disparity of the gender effects on lower tails is almost -4 mg/dL. At higher quantiles, the conditional mean effect of female is less than of male by, on average, approximately -9 mg/dL. The OLS method estimation is fixed across different quantiles. The QR plot shows that at lower quantiles of the conditional FBG distribution, the gender has a small effects, but at higher quantiles, the gender

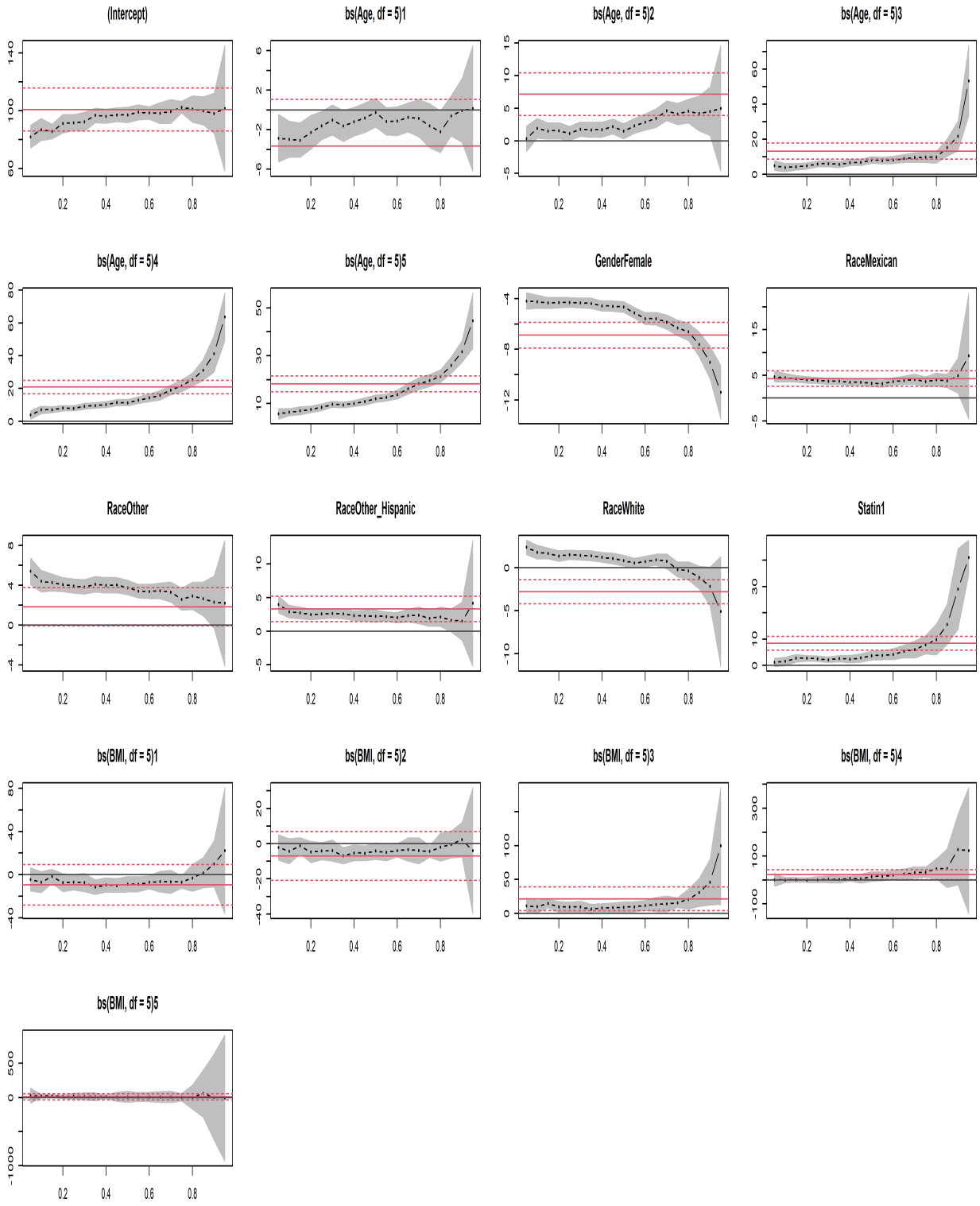


Figure 2.4: Quantile regression marginal effects results for fasting blood glucose regressed on Gender, Age, Race-Ethnicity, Statin use (0-not statin users, 1- statin users), and body mass index. The x-axis represents different glucose quantiles.

has a larger effects.

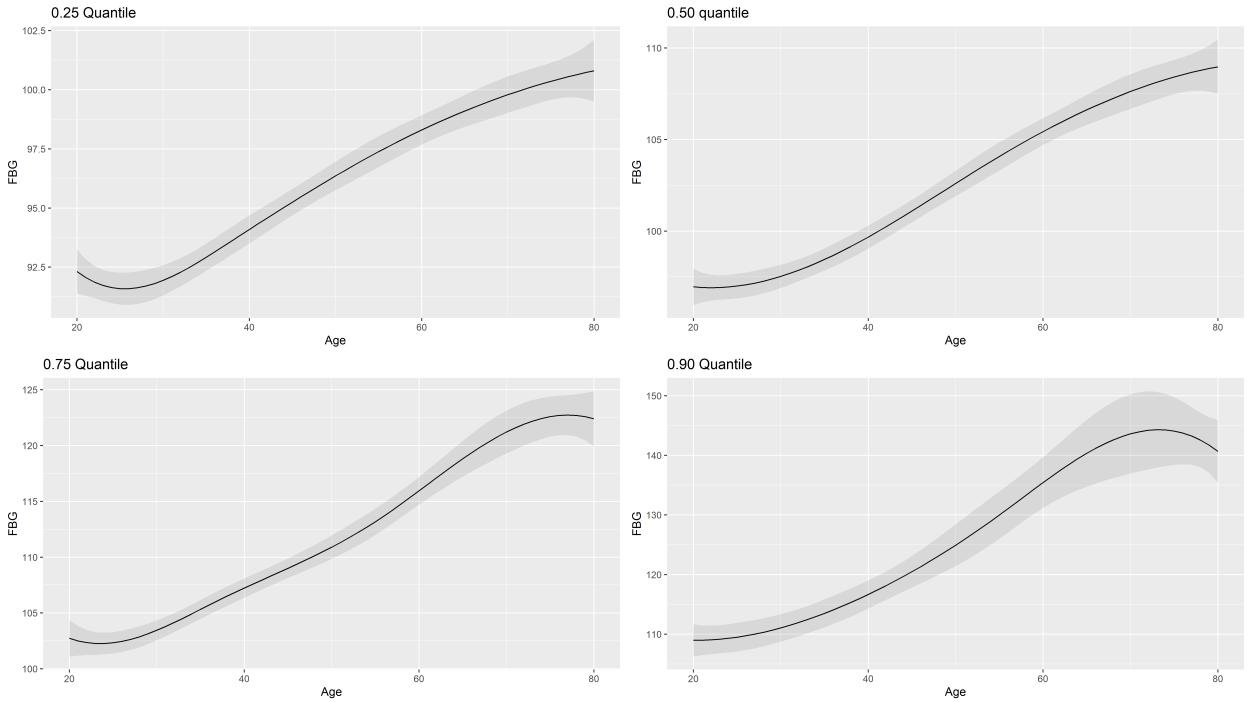


Figure 2.5: Illustration of the age effect on glucose levels for four different quantiles of the conditional glucose distribution.

Fig. 2.5 shows the age effect, which is modeled using spline basis expansions of degree 5. There is a positive association between age and FBG. However, at a lower quantile of FBG $\tau = 0.25$, there is a negative association for an individual who is between 20 and 27 years old. At higher quantiles of FBG, we do not see this trend for a young population, i.e., the age effect is almost negligible for this age group. One reason for the negative association, at younger ages, is that individuals are more active due to engaging in college life and other sport activities. When an individual's age passes 27 years old, the age effects start to increase. These effects are very small at lower quantiles, which is around 10 mg/dL. At a higher quantile, $\tau = 0.75$, age effects are small at lower ages and they starts to increase at a round 27 years old until around 70 years old, where the age effects start to diminish. Populations with high glucose levels have

a higher rate of changes with respect to age if compared to low glucose level populations. It was found that there is a significant correlation between glucose levels and age [87]. Francisco et al. [30] reported a positive association between diabetes and age in the studied population, in which age values are less than 74 years old. However, this trend is reversed after age 74 years old, see the right bottom plot in Fig. 2.5.

Race/Ethnicity is a categorical variable with five categories. Different Races/Ethnicities (Mexican, Other Hispanic, Non-Hispanic White, Other Race-Including Multi-Racial) have different associated trends with respect to Non-Hispanic Black, but overall Race-/Ethnicity associates with an increase in the glucose level if compared to Non-Hispanic Black except Non-Hispanic White. Mexican population has higher FBG levels by approximately 5 mg/dL when compared to Black population at a lower quantiles. The differences increase to 10 mg/dL at higher quantiles. At lower quantiles, Non-Hispanic white population has higher FBG than Non-Hispanic Black population, but at higher quantiles, Non-Hispanic White population has lower FBG value by 6 mg/dL.

Next, from the OLS plot, statin users have on average higher glucose levels if compared to non-statin users, which is around 10 mg/dL. When QR is used to study the association, a different trend is detected. The disparity in glucose level for statin users vs non-statin users is negligible at lower quantiles. However, statin use seems to be associated with rather large effects on FBG levels somewhat larger than 40 mg/dL for the upper quantile see Fig. 2.4. In other words, there is a strong association between high glucose levels and statin use. As is mentioned earlier using statin raises the chance of

progression to the diabetes state [78].

Table 2.2: Estimation of the 90th quantile regression parameters are presented with the standard errors and P values.

| | **Estimate** | **Std. Error** | **t value** | $Pr(>\|t\|)$ |
|---|---|---|---|---|
| **Intercept** | 98.103 | 8.301 | 11.819 | $< 0.001$ |
| **Gender Female** | -9.013 | 0.777 | -11.597 | $< 0.001$ |
| **Age1** | -0.052 | 1.951 | -0.027 | 0.97855 |
| **Age2** | 4.548 | 2.221 | 2.048 | 0.041 |
| **Age3** | 21.746 | 5.078 | 4.282 | $< 0.001$ |
| **Age4** | 41.242 | 6.803 | 6.062 | $< 0.001$ |
| **Age5** | 31.705 | 2.682 | 11.820 | $< 0.001$ |
| **Mexican** | 4.920 | 2.310 | 2.130 | 0.0332 |
| **Other** | 2.324 | 1.575 | 1.475 | 0.140 |
| **Other Hispanic** | 1.501 | 1.692 | 0.887 | 0.375 |
| **White** | -2.162 | 1.348 | -1.605 | 0.109 |
| **Statin** | 29.062 | 9.287 | 3.129 | 0.002 |
| **BM1** | 9.951 | 12.813 | 0.777 | 0.437 |
| **BMI2** | 2.273 | 5.895 | 0.386 | 0.700 |
| **BMI3** | 46.244 | 20.252 | 2.283 | 0.022 |
| **BMI4** | 127.486 | 89.331 | 1.427 | 0.154 |
| **BMI5** | 5.572 | 381.297 | 0.015 | 0.988 |

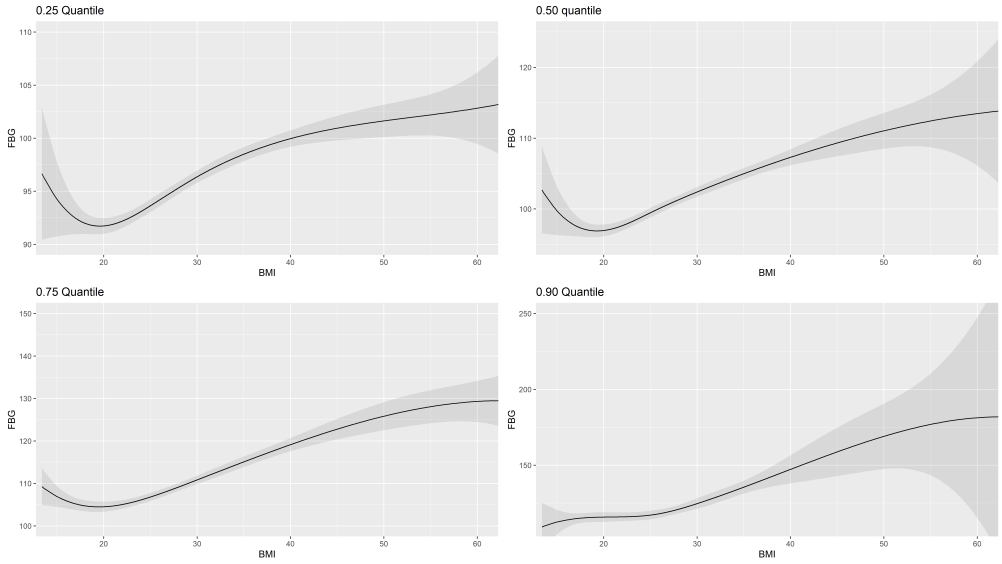


Figure 2.6: Illustration of the BMI marginal effect on the conditional FBG distribution. Different BMI quantiles have different marginal effects on the conditional FBG distribution.

BMI is modeled using spline basis expansions. According to NHANES data for 2007, 63% of Americans were in the overweight category and 26% are in the obese category [22]. For a population with low BMI ( BMI $\leq$ 20 kg/m$^2$), the association between FBG and BMI is negative for the first three FBG quantiles ($\tau = 0.25, \tau = 0.5, \tau = 0.75$). At a higher quantile ($\tau = 0.90$) of FBG, the association is flat for the population with a BMI value less than 25 kg/m$^2$. The positive association between FBG and BMI starts to occurs after an individual BMI passes the cut point around 20 kg/m$^2$. The marginal BMI effects on FBG starts to diminish when BMI values pass 60 kg/m$^2$. This might be due to the drop off in the population size after this value, see Fig. 2.6. Different studies report the BMI effects on FBG [27], [47]. Our results suggest that the ideal BMI values is around 20 kg/m$^2$. The CDC classification of ideal BMI is between 18.5 kg/m$^2$. and 24.9 kg/m$^2$ [14].

### 2.3.1 Race/Ethnicity and Gender Effects on Glucose

Recently, Race/Ethnicity effects on cardiovascular health have attracted researchers attention, see [28] and references therein. They found that cardiovascular health for non-Hispanic Asian Americans is better than the cardiovascular health of non-Hispanic Whit. The difference diminishes at lower BMI.

In this section, our goal is to estimate the effects of gender, age, statin, and Race/Ethnicity on FBG at different quantiles of the conditional FBG distribution. The number of predictors in the model is reduced for simplicity purposes. The model is defined as follows:

$$\hat{\text{FBG}} = \beta_0 + \beta_1 \textit{Gender} + \beta_2 \textit{spline(Age, df=5)} + \beta_4 \textit{Statin} + \beta_5 \textit{Race/Ethnicity} + \epsilon. \quad (2.3.1)$$

The association between age and glucose levels in different quantiles of glucose varies with respect to race and gender. On average, in all Race/Ethnicity groups males have higher FBG levels than females, see Fig. 2.7, $\cdots$, Fig. 2.10. Statin users on average have higher FBG levels than non-statin users. This observation is consistent with what we have found in the boxplot, Fig. 2.1. Moreover, the association between age and FBG at higher quantiles is greater than the association at lower quantiles because the estimation of age coefficient, at lower quantile $\tau = 0.25$, is less than the estimation of age coefficient at higher quantile, $\tau = 0.90$. This association was noticed in the adolescent population, see [44].

The lowest age effects occur in females and males from the group of Other Race-include Multi-racial. The highest age effects occur in the female and males group of Other Hispanic. This may interpret the reason why the cardiovascular health of non-Hispanic Asian Americans is better than the cardiovascular health of non-Hispanic White [28].

Table 2.3: Parameter estimates for the model(2.3.1).

| | $\tau = 0.25$ | $Pr(>\|t\|)$ | $\tau = 0.50$ | $Pr(>\|t\|)$ | $\tau = 0.75$ | $Pr(>\|t\|$ | $\tau = 0.90$ | $Pr(>\|t\|)$ |
|---|---|---|---|---|---|---|---|---|
| **Intercept** | 85 | $< 0.001$ | 89.795 | $< 0.001$ | 96.168 | $< 0.001$ | 101.902 | $< 0.001$ |
| **Age1** | -0.661 | 0.472 | 1.147 | 0.196 | 0.294 | 0.816 | 0.270 | 0.911 |
| **Age2** | 3.208 | $< 0.001$ | 4.013 | $< 0.001$ | 5.713 | $< 0.001$ | 10.913 | $< 0.001$ |
| **Age3** | 6.600 | $< 0.001$ | 9.679 | $< 0.001$ | 11.849 | $< 0.001$ | 27.186 | $< 0.001$ |
| **Age4** | 10.223 | $< 0.001$ | 14.098 | $< 0.001$ | 24.171 | $< 0.001$ | 50.364 | $< 0.001$ |
| **Age5** | 9.472 | $< 0.001$ | 12.300 | $< 0.001$ | 20.3005 | $< 0.001$ | 31.910 | $< 0.001$ |
| **RaceMexican** | 4.186 | $< 0.001$ | 3.433 | $< 0.001$ | 2.604 | $< 0.001$ | 6.187 | 0.007 |
| **RaceOther** | 2.133 | $< 0.001$ | 1.432 | $< 0.001$ | -0.679 | 0.284 | -2.667 | 0.138 |
| **RaceOther Hispanic** | 2.394 | $< 0.001$ | 1.673 | $< 0.001$ | 0.912 | 0.255 | 1.02055 | 0.633 |
| **RaceWhite** | 0.737 | 0.029 | 0.205 | 0.572 | -1.549 | $< 0.001$ | -4.000 | 0.008 |
| **Gender(Male)** | 4.396 | $< 0.001$ | 4.600 | $< 0.001$ | 6.153 | $< 0.001$ | 7.021 | $< 0.001$ |
| **Statin(statin users)** | 2.79102 | $< 0.001$ | 3.804 | $< 0.001$ | 9.204 | $< 0.001$ | 28.676 | 0.036 |

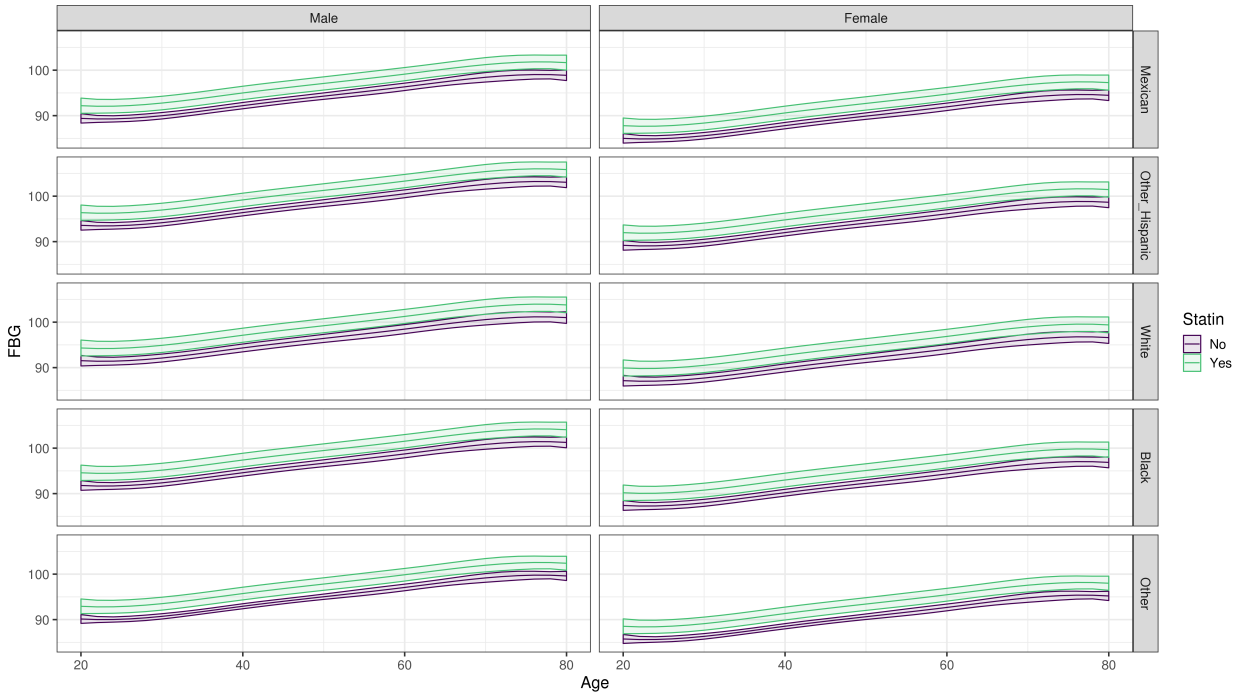


Figure 2.7: Illustration of the age effect with respect to Race/Ethnicity (Mexican American, Other Hispanic, Non-Hispanic White, Non-Hispanic Black, Other Race-Including Multi-Racial) and gender for the first quantile ($\tau = 0.25$)of the conditional glucose distribution.

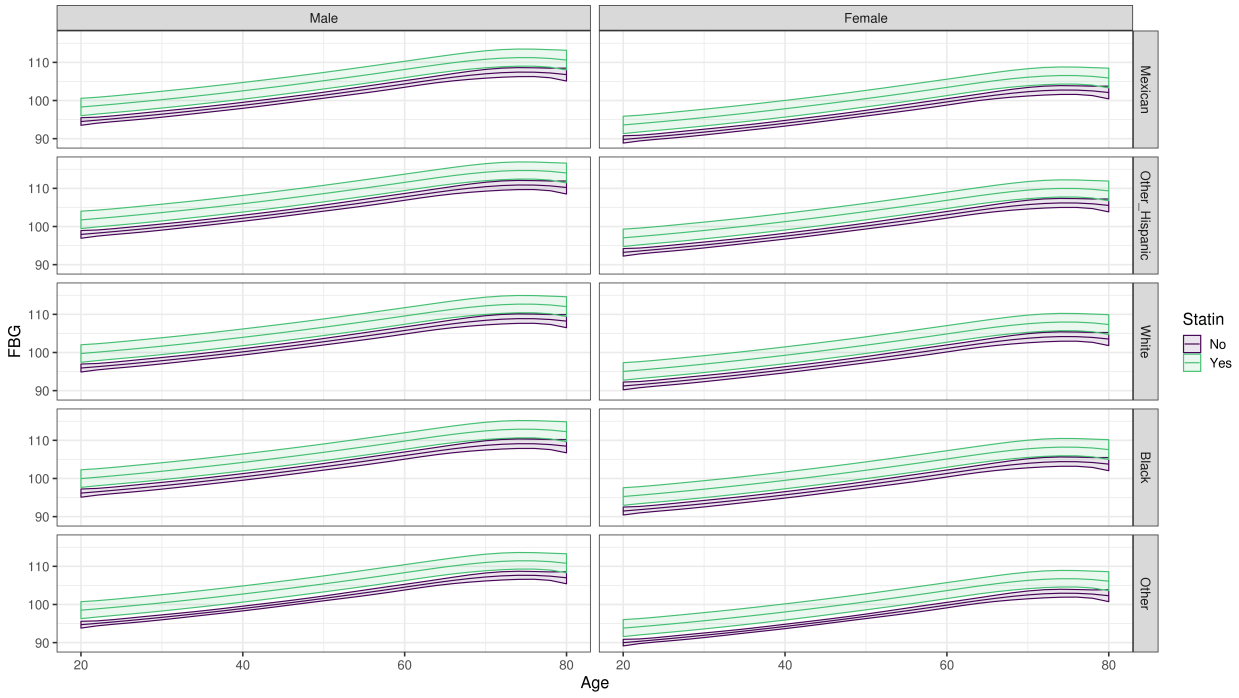


Figure 2.8: Illustration of the age effect with respect to Race/Ethnicity (Mexican American, Other Hispanic, Non-Hispanic White, Non-Hispanic Black, Other Race-Including Multi-Racial) and gender for the second quantile ($\tau = 0.5$)of the conditional glucose distribution.

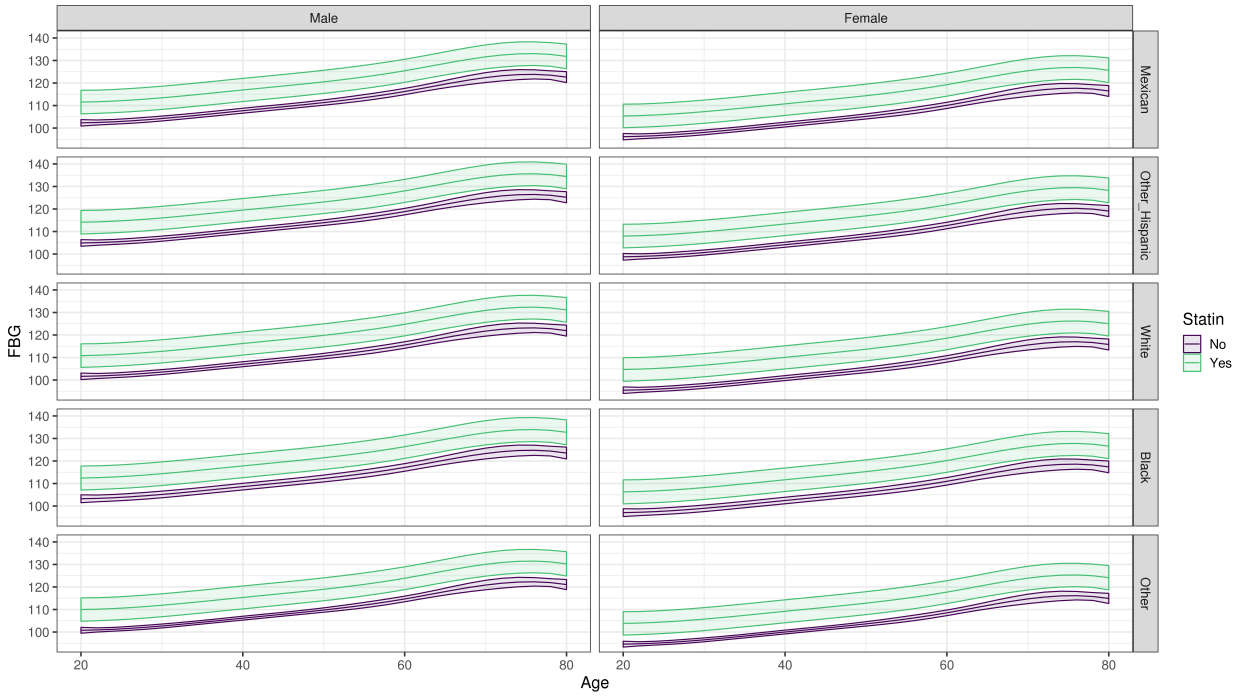


Figure 2.9: Illustration of the age effect with respect to Race/Ethnicity (Mexican American, Other Hispanic, Non-Hispanic White, Non-Hispanic Black, Other Race - Including Multi-Racial) and gender for the third quantile ($\tau = 0.75$)of the conditional glucose distribution.

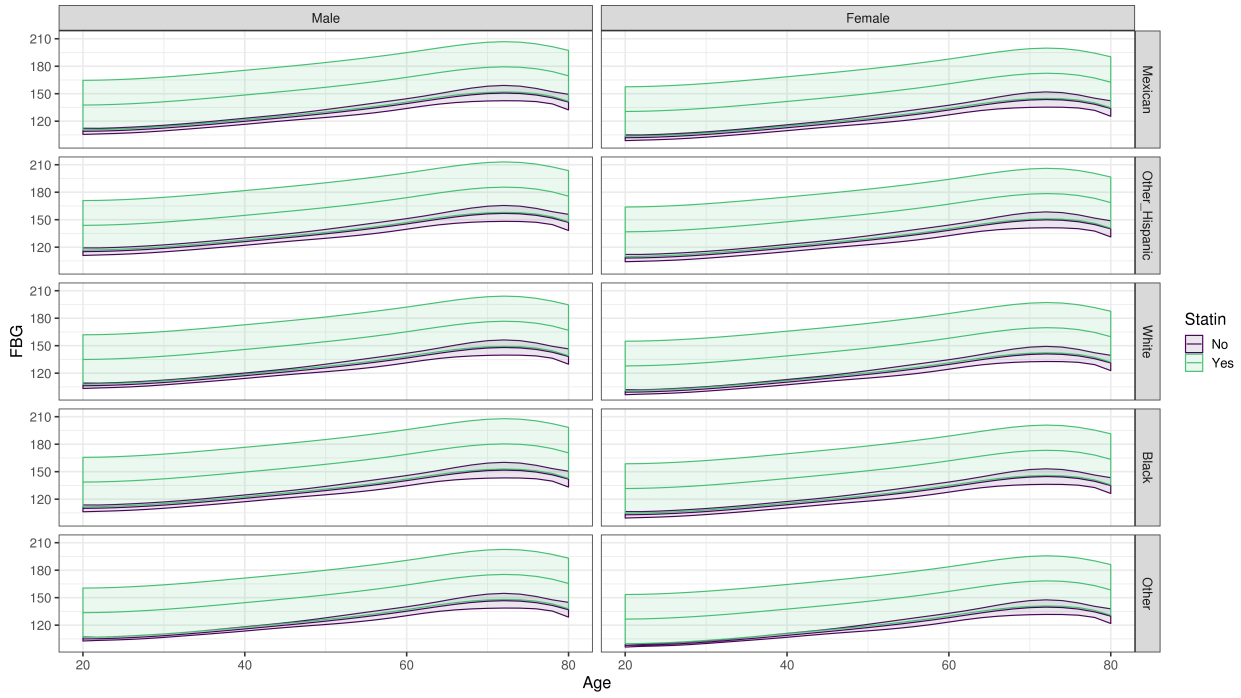


Figure 2.10: Illustration of the age effect with respect to Race/Ethnicity (Mexican American, Other Hispanic, Non-Hispanic White, Non-Hispanic Black, Other Race - Including Multi-Racial) and gender for the 90th quantile of the conditional glucose distribution.

## 2.4 Conclusion

Multivariate quantile regression is used to study the effects of different variables on fasting blood glucose levels. A positive relation between FBG and age is found. Furthermore, statin effects on glucose at the lower quantiles are very small if compared to the higher quantiles of FBG, see Fig. 2.4. Therefore, it is crucial to find a policy that maximizes the benefit of using statin and minimizes the side effects.

We also studied the association between FBG and TC using QR. Our study showed that there is negative association between TC and FBG for the value of TC less than

140 mg/dL for the FBG value less than 105 mg/dL. At higher values of FBG,that is greater than 105 mg/dL, the lowest value of FBG occurs at TC value around 160 mg/dL. In conclusion, TC values around 150 mg/dL is an ideal value for TC.

Additionally, BMI effects on FBG have been presented. Our results state that the lowest BMI effects on FBG occurs at values of BMI around 20 BMI, which are the recommended BMI values. When BMI values passes 20 mg/dL, it is positively associates with FBG.

Race/Ethnicity effects on FBG vary with respect to age. The lowest age effects on FBG occurs in female White and Other Race-Including Multi-Racial . The highest effects of age occurs in the male Mexican and black.

## Chapter 3

## Characterizing Prediabetes Risk and Lipid Ratio Progression Using NHANES Cross-Sectional Survey Data

Approximately 17.8 million deaths in 2017 were caused by cardiovascular (CVD) diseases globally, and it is estimated that by 2030, more than 22.2 million deaths will be caused by CVD [92]. In 2011, annual costs for CVD and stroke were $320.1 billion, which is more than the annual cost of cancer. This includes $195.6 billion in direct costs (health care costs), and the cost of future productivity loss is $124.5 [64]. One of the CVD risk factors is abnormal lipid ratio [96]. Lowering low-density lipoprotein (LDL), using statin, reduces the risk of cardiovascular diseases even in a population with no CVD [96]. It is known that diabetes is a risk factor for CVD; however, it is shown that elevated glucose is also a risk factor for CVD even for prediabetes [4]. Nielson et al. [66] found that individuals with prediabetes compared to normoglycemic have higher incident rates of myocardial infarction, acute coronary syndrome, and a greater number of new prescriptions for nitrates by $53.9\%, 18.6\%, 26.4\%$, respectively. An independent predictor of deaths and cardiovascular events is high volatility of fasting blood glucose and total cholesterol levels, systolic blood pressure, and body mass index [48].

Statin use is associated with a high risk of new-onset diabetes in normoglycemic (HR 1.19, 95% CI, 1.05 to 1.35) and in prediabetes patients (HR 1.24 95% CI, 1.11 to 1.38). On the other hand, overall mortality risks decrease in both normoglycemic (HR 0.70; 95 % CI, 0.66 to 0.80) and impaired fasting glucose (IFG) patients (HR 0.77, 95 % CI, 0.64 to 0.91) with statin use [13]. For patients with one or more risk factors for diabetes, statin use is associated with a 39% reduction in the primary endpoints (the HR

0.61) and a 28% (HR 1.28) increase in diabetes [76]. Pandya et al. [68] suggested that the dis-utility associated with taking a pill daily, statin price, and the risk of statin-induced diabetes negatively impacted the cost-effectiveness of statin use.

Several studies estimated the association of high-level glucose risk factors, such as waist circumference, and cholesterol, with the conditional mean of glucose using ordinary least squares (OLS) or logistic regression models. The drawbacks in these studies are ignoring the dispersion of the association of covariates with different quantiles of the dependent variables (glucose levels).

In the second section, we estimate transition probabilities for different cholesterol levels using cross-sectional survey data. There are different models that are used to estimate transition probabilities from cross-sectional data; see [70] and [26].

### 3.1 Estimating Transition Probabilities

One of the issues in chronic diseases modeling, using the Markovian chain, is estimating the transition probability of a patient moves from one state of a categorical risk factor to another state of a different risk factor in the next decision epoch (or next year). There is vast literature about estimating transition probability for multi-state Markov models from longitudinal data. However, this is not the case with cross-sectional data. There are some methods that are developed to estimate transition probability from repeated cross-sectional data using age-specific prevalence data, see for an example [69] and [70]. Pelzer et al. [69] utilize a Bayesian approach to estimate the transition probabilities, but the restriction with using this method is that it works well only for risk factors that can be categorized into two states, for example, two states of lipid ratios, high and low. Another approach to estimate transition probabilities is through estimating the odd of transition to a particular state against staying at the same state. Van

de Kassteele et al. [90] introduced a method to calculate net transitions, which are net inflow or outflow into a particular risk factor at a particular time point. Net transitions are different from transition probabilities. In this approach, a transportation problem is formulated to transform the proportions from time $t$ to time $t+1$. The drawback of this approach is that it is sensitive to the cost matrix which is hard to estimate. Engada et al. [26] implemented this method by using longitudinal data to estimate the cost matrix and then used this cost matrix for larger populations.

In this section, we present a continuous-time multi-state model that is used to estimate transition probabilities which was developed by Ardo Van Den Hout [91]. We denote age in years by $t$, and the lipid ratio states are denoted by $LR1, LR2, LR3$ in ascending order. Transitions between lipid ratio sates are defined by the hazard model

$$q_{(r,s)}(t) = \exp(\beta_{(r,s)} + \xi_{(r,s)}t), \qquad (r,s) \in (1,2),(2,1),(2,3),(3,2), \tag{3.1.1}$$

where $\beta_{(r,s)}$ is transition hazards, and $\xi_{(r,s)}$ is the age effect. Maximum likelihood is used to estimate the model parameters. The likelihood function of a multinomial distribution for the frequency at a point in time (age $t$) conditional on the proportions at a previous point in time (age $t-1$) is constructed. The distribution of frequencies $\omega_t = (\omega_1(t), \omega_2(t), \omega_3(t))$ at time $t$ conditioning on row vector of proportions $p_{t-1} = (p_1(t-1), p_2(t-1), p_3(t-1))$ at $(t-1)$ is formulated as multinomial distribution; that is,

$$\omega_t | p_{t-1}, P(t-1,t), m_t \sim \text{Multi}(p_{t-1}P(t-1,t), m_t), \tag{3.1.2}$$

where $P(t-1,t) = [\exp^{q_{rs}}]$ is a transition matrix for one period, and the number of the multinomial trials is denoted by $m_t$. The sum of the observed frequencies for each of the lipid ratio levels at time $t$ is the number of multinomial trials. The probability mass function of multinomial distribution is denoted by $f$. The likelihood function is defined

as

$$L(\theta|data) = \sum_{t=1}^{20} \log(f(\omega_t|p_{t-1}, m_t, \theta)),$$

where $\theta$ is a vector of parameters of interest. The model is implemented for a population with age values between 40 and 60 years. This age range is selected to ensure we get an accurate picture of all cholesterol level distributions because the high cholesterol levels have a negative impact on an individual life expectancy and, hence, the proportion of high cholesterol levels population disappear from the higher ages. The time scale $t$ represents the age after converting it to start from 0, i.e., in this scenario $(t - 40)$.

Due to limited information that is available on individual transitions in the cross-sectional data, our model is set to restrict the number of variables that are needed to be estimated. This can be done by assuming we have only two different processes: one for moving forward and the other for moving backward. In this scenario, we have only two independent parameters: one for the hazards of transitioning forward and the other for the hazards of transitioning backward. Hence, the parameters that are needed to be estimated are $\beta_{12} = \beta_{23} = \beta_F$, for the hazards of moving forward and $\beta_{21} = \beta_{32} = \beta_B$ for the hazards of moving backward. Due to the positive association between lipid ratio and age [38], we extended our model to include the age effect for the risk of moving forward through different lipid ratio levels: $LR1, LR2, LR3$. The age effect for the hazard of moving forward is denoted by $\xi_f$, and these forward effects are assumed to be equal; $\xi_{12} = \xi_{23} = \xi_F$. The age effect for the hazard of moving backward is assumed to be zero as it does not improve Akaike's information criterion (AIC) value, i.e., $\xi_{21} = \xi_{32} = 0$.

$$q_{12}(t) = \exp(\beta_F + \xi_F t)$$

$$q_{21}(t) = \exp(\beta_B)$$

$$q_{23}(t) = \exp(\beta_F + \xi_F t)$$

$$q_{32}(t) = exp(\beta_B),$$

### 3.1.1 Introduction to Data

The data used in this study is National Health and Nutrition Examination Survey data (NHANES) [16]. It is available for free for the public with no ethics application required. The survey examines a nationally representative sample of the U.S. population. Every two years a new cycle of the survey is implemented. The survey focuses on a variety of health and nutrition measurements. In this study, we cumulated 6 cycles of NHANES data (2007–2018). There are around 12,000 records. Population ages in this study are more than 20 years old and less than 80 years old.

Fig. 3.1 shows TC distribution with respect to age, gender, and Race/Ethnicity using box plot. From this visualization, we can see different trends for the mean of the TC as a function of age. Mean values of TC for different races/ethnicities and gender start low and then increase to the peak. Then they start to decline. The peak point vary with respect to Race/Ethnicity and gender. For example, the peak point for a white female American is around the age group 5 , while for a male is a round the age group 4. In all Race/Ethnicity groups, the average TC passes 200 mg/dL in males earlier than in

females.

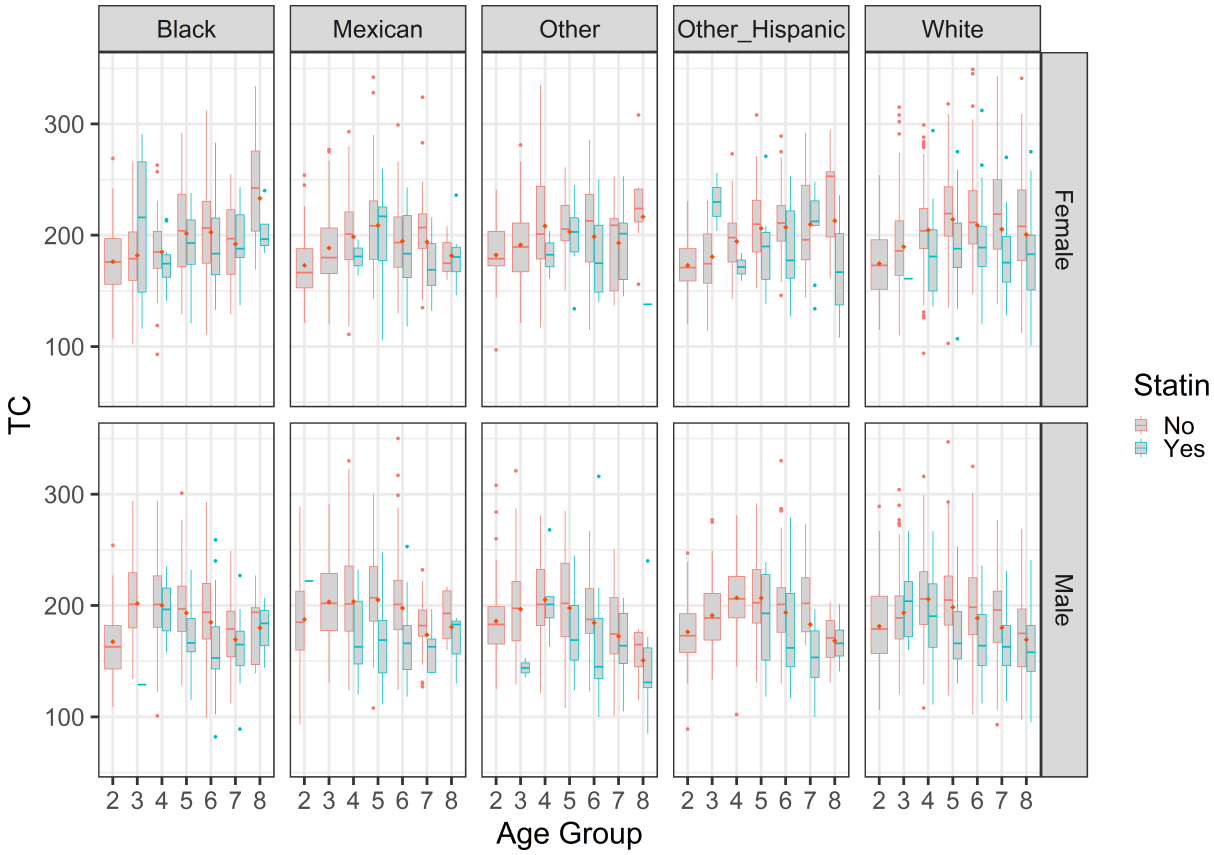


Figure 3.1: A Boxplot for TC is presented for prediabetes population with respect to Gender, Age, Statin uses, and Race-ethnicity (Mexican American, Other Hispanic, Non-Hispanic White, Non-Hispanic Black, Other Race - Including Multi-Racial). Age groups $2, 3, \cdots, 80$ on the x-axis represents ages from $20-29, 30-39, \cdots, 80$, respectively.

### 3.1.2 Results

The above multi-state model is fitted on NHANES cross-sectional survey data to estimate transition probabilities. The model is fitted to the prediabetes population whose ages range from 40 to 60 years. Four subgroups of the prediabetes population are considered: male statin users, male non-statin users, female statin users, and female non-statin users. The likelihood function is optimized using the *optim* function in the statis-

tical software R.

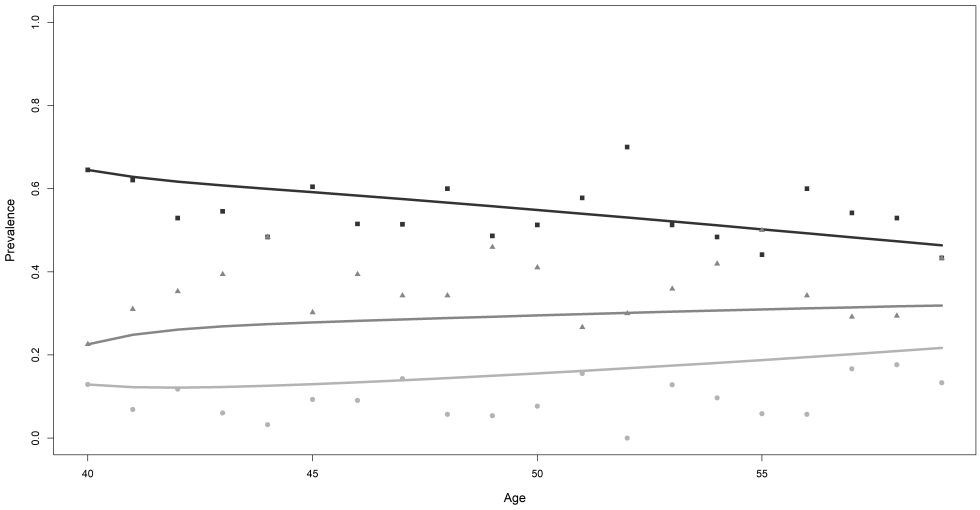


Figure 3.2: An illustration of the observed prevalence and the estimated distribution of the prevalence of the three categories of lipid ratios for the NHANES prediabetes males non-statin users.

Fig. 3.2 shows under conditioning on the prevalence at a particular age, the model is able to predict the three-stat process. With only a few parameters to estimate (three parameters), the model is efficient in capturing the trend in the state transitions with respect to age given the prevalence at a specific age, such as 40.

Table 3.1: Estimations of the model parameters are presented for females and males prediabetes non-statin users using the NHANES data.

females and males prediabetes statin users.

| **Parameters** | **Estimation (S.E.) for a female** | **Estimation (S.E.) for a male** |
|---|---|---|
| $\hat{\beta}_F$ | -1.86 (0.54) | -1.58 (0.53) |
| $\hat{\beta}_B$ | -1.06 (0.43) | -1.13 (0.3) |
| $\hat{\xi}_F$ | 0.03 ( 0.03) | 0.02 (0.03) |

Table 3.2: Estimations of the model parameters are presented for the females and males prediabetes statin users.

| **Parameters** | **Estimation (S.E.) for a female** | **Estimation (S.E.) for a male** |
|---|---|---|
| $\hat{\beta}_F$ | -1.98 (0.6) | -1.58 (0.53) |
| $\hat{\beta}_B$ | -0.63 (0.55) | -1.13 (0.3) |
| $\hat{\xi}_F$ | 0.02 ( 0.03) | 0.02 (0.03) |

Table 3.3: The fitted model AIC ($-2\log$(likelihood)) value for the four prediabetes subgroups using the NHANES data

| **Parameters** | **Male** | **Female** |
|---|---|---|
| Non-statin users | 231.78 | 166.87 |
| Statin users | 200.98 | 136.5 |

The transition probability matrix for a female prediabetes non-statin user at age 50 years old is given in the matrix Eq. (3.1.3). The first row of the matrix shows the probability of moving from $LR1$ to $LR1$, $LR2$, and $LR3$. Similarly, the second and third-row represents the probability of moving from $LR2$ and $LR3$ to $LR1$, $LR2$, and $LR3$. The transition probabilities for female statin users, male non-statin users, and male statin users are given in Eq. (3.1.4), Eq. (3.1.5), and Eq. (3.1.6), respectively.

$$P_{fns} = \begin{bmatrix} 0.84 & 0.14 & 0.01 \\ 0.24 & 0.62 & 0.13 \\ 0.04 & 0.23 & 0.73 \end{bmatrix} \tag{3.1.3}$$

$$P_{fs} = \begin{bmatrix} 0.87 & 0.12 & 0.01 \\ 0.36 & 0.55 & 0.1 \\ 0.09 & 0.30 & 0.61 \end{bmatrix} \tag{3.1.4}$$

$$P_{mns} = \begin{bmatrix} 0.81 & 0.17 & 0.02 \\ 0.22 & 0.61 & 0.17 \\ 0.04 & 0.21 & 0.75 \end{bmatrix} \tag{3.1.5}$$

$$P_{ms} = \begin{bmatrix} 0.84 & 0.15 & 0.02 \\ 0.31 & 0.56 & 0.13 \\ 0.07 & 0.28 & 0.65 \end{bmatrix} \tag{3.1.6}$$

The transition probabilities of lipid ratios for female non-statin users, male non-statin users, female statin users, and male statin users are plotted in Fig. 3.3, Fig. 3.4, Fig. 3.5, Fig. 3.6, respectively. The first row of the panel represents, as a function of age, the transition probability of moving from LR1 to each of the three lipid ratios LR1, LR2, and LR3. The columns show the probability of moving to a single state from the three levels of lipid ratios. For example, at age 50, the probability of staying at the same state LR1 in the next year is 0.84, the probability of moving from LR1 to LR2

in the next year is 0.24, and the probability of moving from LR1 to LR3 is 0.04.

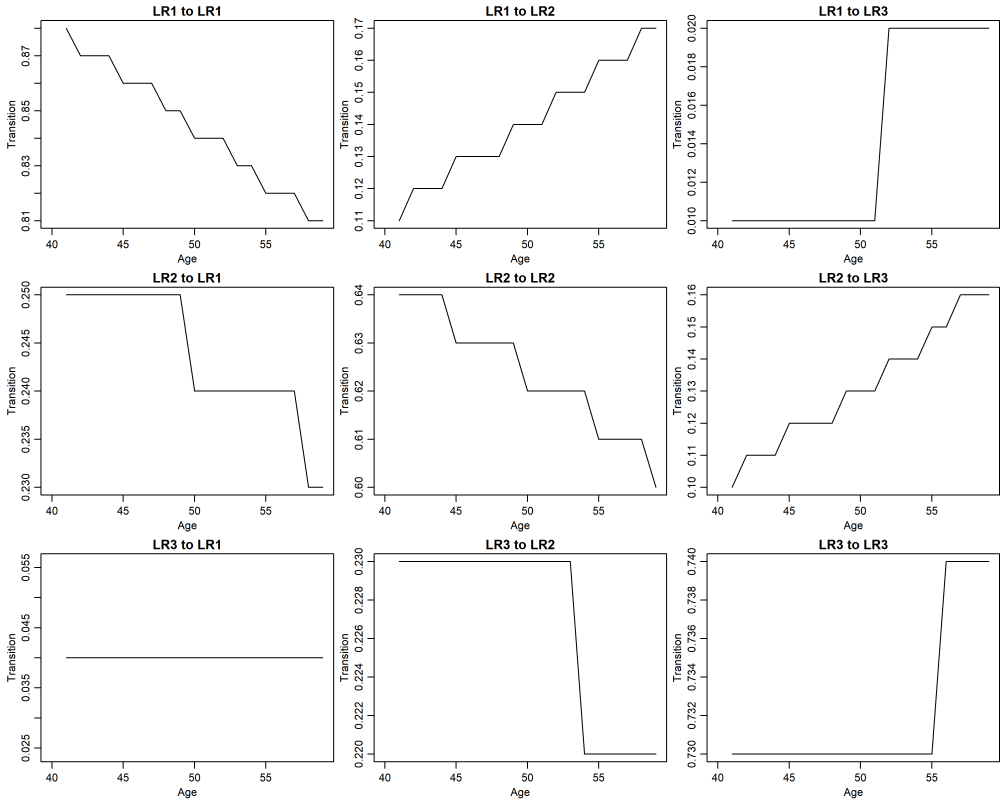


Figure 3.3: An illustration of the transition probabilities for the three categories of lipid ratios for a female non-statin user.

Fig. 3.3 shows there is a positive relationship between age and the risk of moving to higher cholesterol states. For example, at age 50, the risk of moving from $LR1$ to $LR2$ is 0.14, while at a higher age (60 years old) the risk becomes around 0.17. The probability of an individual staying at the same LR level in the next year is higher than the

probability of moving to a higher LR level or a lower LR level.

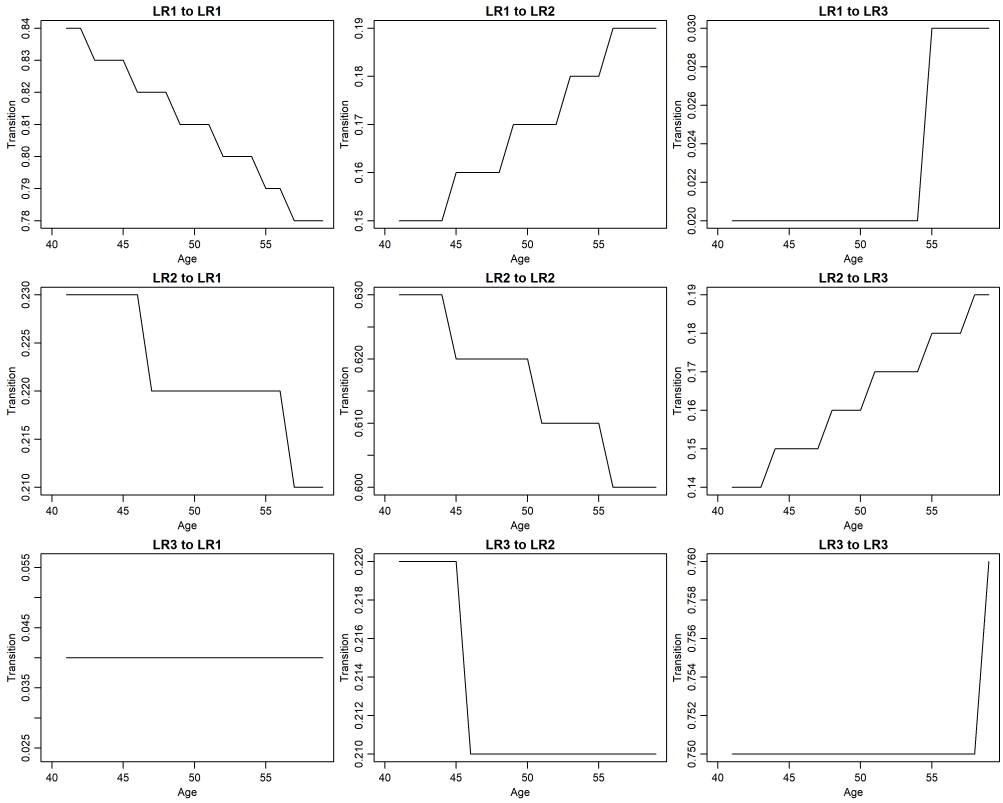


Figure 3.4: An illustration of the transition probabilities for the three categories of lipid ratios for a male non-statin user.

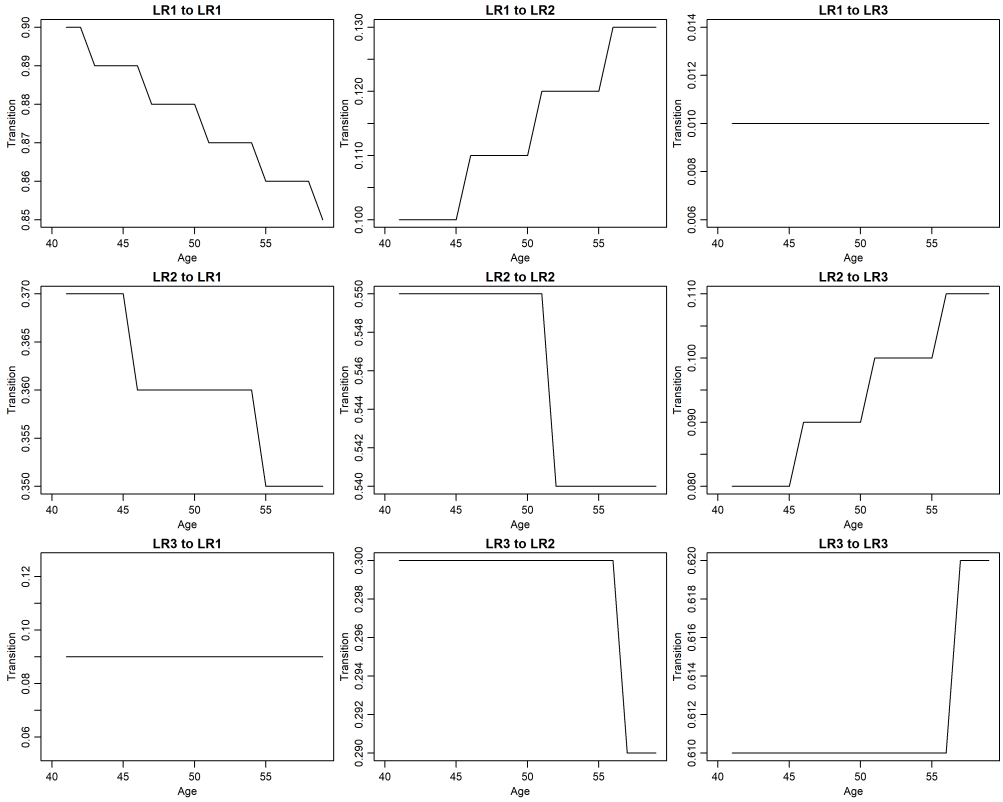


Figure 3.5: An illustration of the transition probabilities for the three categories of lipid ratios for a female statin user.

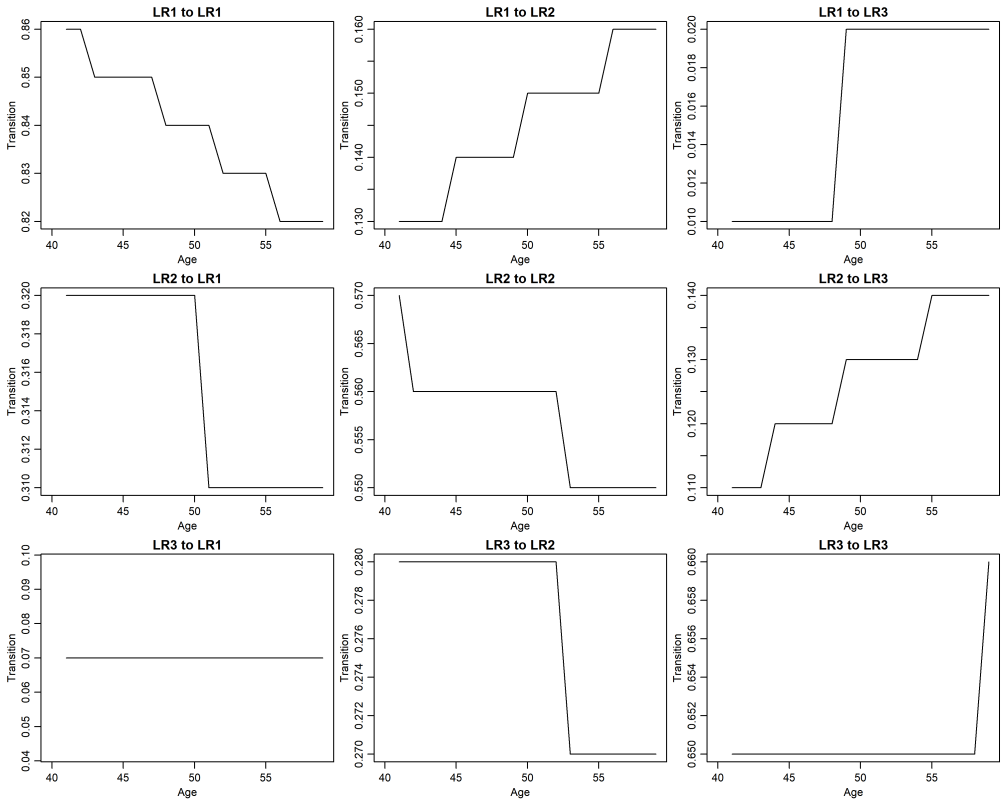


Figure 3.6: An illustration of the transition probabilities for the three categories of lipid ratios for male statin users.

### 3.1.3 Conclusion

A multi-state model is used to estimate the transition probability matrix for the three levels of lipid ratios. The transition probabilities among different lipid ratio levels for males are higher than the transition probabilities for females. Engeda et al. [26] show the net transition probabilities among different total cholesterol levels vary with respect to Race/Ethnicity and sex. As expected, statin therapy reduces the probability of progression to a higher lipid ratio level, i.e., the probability of progression from LR1 to LR2 and LR2 to LR3 are higher for non-statin users if compared to statin users.

Adding age effects to the model has positively impacted model performance in terms of AIC value.

Our model is sensitive to the observed proportions at the starting age because the probability of each outcome, category of lipid ratio, to occur at age $t$ depends on the observed proportions at age $(t-1)$. This dependent relationship goes to the starting age. In our setting, the starting age is 40 years old.

### 3.2 Using Bayesian Approach to Estimate The Risk of Prediabetes

In this section, we estimate the risk of prediabetes using the Markov Chain Monte Carlo (MCMC) approach to estimate the parameters of the logistic regression function. One advantage of using Bayesian inference over the maximum likelihood approach to estimate model parameters is that in Bayesian inference we are concerned with sampling from the posterior distribution of the parameters given the data and prior density. The mean or mode of the posterior distribution of the parameters is used as an estimation for the population parameters, while in the maximum likelihood approach we are concerned with calculating the maximum for the likelihood function, which might have multiple local maximums that make it intractable to solve. Therefore, the Bayesian inference approach is more accurate in this case.

We consider the prediabetes population in the NHANS data set. A logistic regression model is used to estimate the probability of transitioning from a normoglycemic to a prediabetes state. The basic idea of inferencing parameters of a population is rooted in Bayes' Theorem (Thomas Bayes, 1701-1761). The Mathematical formulation for the Bayes' Theorem is as follows:

$$\pi(\theta|D) = \frac{L(D|\theta)\pi(\theta)}{\pi(D)},$$

where $\theta$ is a parameter of interest, $D$ is the data, $L(D|\theta)$ is the likelihood function of $D$ given $\theta$, and $\pi(\theta)$ is the probability distribution of $\theta$ (prior). For $D = \{y_1, y_2, \cdots, y_n\}$, where $y_i \overset{iid}{\sim} f(.|\theta)$, the likelihood function is defined as: $L(D|\theta) = \Pi_{i=1}^{n} f(y_i|\theta)$.

In Bayesian inference, we are interested in finding the posterior distribution of $\theta$ given the data i.e., $\pi(\theta|D) = \Pi_{n=1}^{n} f(y_i|\theta)\pi(\theta)$. Information about $\theta$ can be found from analyzing the posterior distribution; for example, we may use the mean= $\int \theta\pi(\theta|D)$ as a good estimator for $\theta$. When $\theta$ is a vector of variables, the calculation of the mean becomes intractable. One approach used to solve this integral is the Monte Carlo method [77]. Using the strong law of large numbers, the mean can be computed using the formula $\frac{\sum_{i=1}^{n} Y_i}{n} \to \mu$, where $Y_i$ are independent and identically distributed random variables sampled from the target distribution (in our scenario posterior distribution.) When the posterior distribution is unknown in closed form, then the MCMC method is used to sample from the posterior distribution. MCMC method can be run using different algorithms such as Metropolis-Hasting algorithm (MH) ([54], [54], Gibbs sampler [37].

Next, we present a brief description of the MH algorithm, which we are going to use to sample from our posterior. At iteration $i$, we propose an initial value for $\theta$ and a proposal distribution $q(x)$. The next proposed value for $\theta$ is drawn from the proposal distribution conditioning on the previous value $q(\theta^{proposed}|\theta^i)$. The acceptance ratio is computed according to the following formula:

$$R = min\{1, \frac{p(\theta^{proposed})q(\theta^i|\theta^{propsed})}{p(\theta^i)q(\theta^{proposed}|\theta^i)}\}.$$

For a uniform random number $u$ generated from the interval (0,1), the next value $\theta^{i+1} = \theta^{proposed}$ if $u \leq R$, otherwise $\theta^{i+1} = \theta^i$. There is a trade-off between the acceptance rate and the variance of the proposed distribution. With high variance, the acceptance rate is low, but with low variance, the acceptance rate is higher. However, choosing a small

value for the variance reduces the chance to investigate the whole parameter space. Therefore, it is preferable to have the acceptance rate in the range between 20-40% [36]. Moreover, it is recommended that the initial draws are thrown to reduce the effects of the initial values. Additionally, to reduce the effect of correlation a thinning process is recommended.

### 3.2.1 Model Formulation

A logistic function takes values between 0 and 1 for all values in the domain (Age). An MCMC is used to maximize the probability of the parameter given the data, that is, the logistic regression parameters $\alpha$, and $\beta$.

$$P(\text{Pre}|\text{Age}) = \frac{1}{1 + e^{\,\alpha+\beta Age}}$$

.

Since we do not have information about the distributions of the model parameters $\alpha$ and $\beta$, we assume they are coming from the normal distribution, which has two parameters $\mu$ and $\sigma$. These parameters play an important role in the model. For example, if the parameters that we are interested in are positive, then we need to choose a value for $\sigma$ that restricts our parameters from taking negative values, provided that $\mu$ is taking a positive value. A larger value of $\sigma$ indicates the data is spread out a lot. The other parameter $\mu$ defines the location of the distribution. So, the normal distribution is defined as follows:

$$f(\alpha|\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\alpha-\mu}{\sigma}\right)^2}.$$

### 3.2.2 Posterior Probability of Prediabetes given Age

We have all the pieces for the posterior probability and can now put them together. The logistic function describes the transition from normoglycemic to prediabetes state, but we do not know the parameters $\beta$ and $\alpha$. The aim is to find the parameters of the logistic function which maximizes the likelihood of the observed data. The parameters are assumed to come from a normal distribution defined by a mean denoted by $\mu$, and a variance denoted by $\tau$. The MCMC algorithm will sample values of $\mu$ and $\tau$ for both $\alpha$ and $\beta$ to try and maximize the parameters of the logistic function given the data.

The data is connected to the parameters through a Bernoulli Variable.

### 3.2.3 Bernoulli Variable

A Bernoulli variable is a discrete random variable that takes a value of either 0 or 1. In our problem, we model normoglycemic or prediabetes as a Bernoulli variable where normoglycemic is 0 and prediabetes is 1. The Bernoulli random variable $pre$ for prediabetes depends on the age, in a manner defined by the logistic function:

$$pre_i \sim \text{Ber}(\ p(t_i)\ ), \quad i = 1, \cdots, N,$$

where $p(t_i)$ is the logistic function with the independent variable time. Therefor, the probability of becoming prediabetes is defined as follows:

$$P(\text{pre}|t_i) = \text{Ber}(\frac{1}{1 + e^{(\beta t_i + \alpha)}})$$

The goal of MCMC is to find the $\alpha$ and $\beta$ parameters using the data and assuming

normal priors.

### 3.3 Results

The Metropolis-Hastings algorithm is used to sample from the posterior distribution. The code is implemented using Python Package PyMC3, which was developed by Google. Fig. 3.7 shows the trace plot for the samples drawn from the posterior distribution for $\beta$ and $\alpha$. The MCMC algorithm is expected to converge at the true values as the number of samples increases. It is recommended to do burn in to the early samples. In the problem, we have drawn 10000 samples and 5000 samples are burned in. Only the last 5000 are used to estimate $\beta$ and $\alpha$ values.

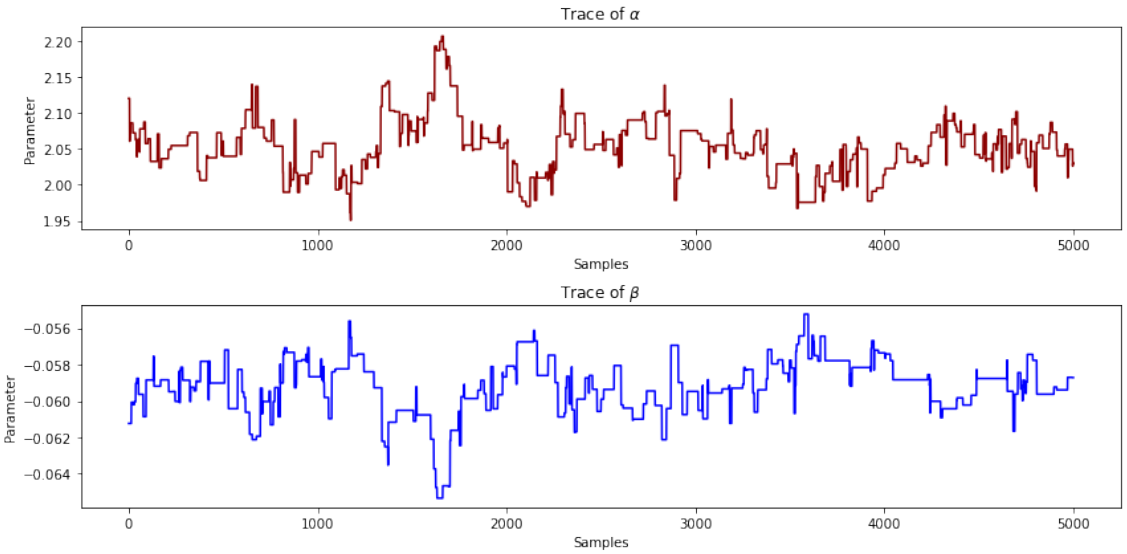


Figure 3.7: Trace of $\alpha$ and $\beta$. Sampled from the posterior distribution using Metropolis Hastings algorithm.

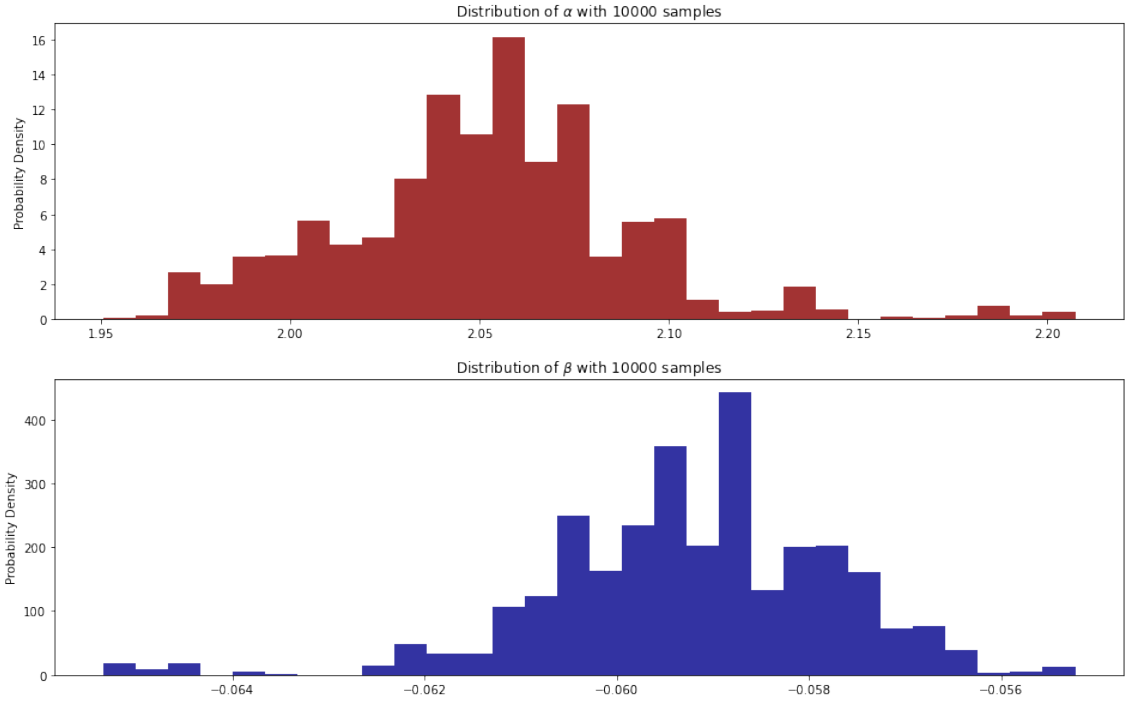


Figure 3.8: Posterior distribution of $\alpha$ and $\beta$.

Fig. 3.8 shows the resulting parameter estimates for $\alpha$ and $\beta$ are 2.049, and -0.059, respectively. These values correspond to modes of the density distributions. These estimations for $\alpha$ and $\beta$ are being used to estimate the risk of prediabetes as a function of

age, see Fig (3.9).

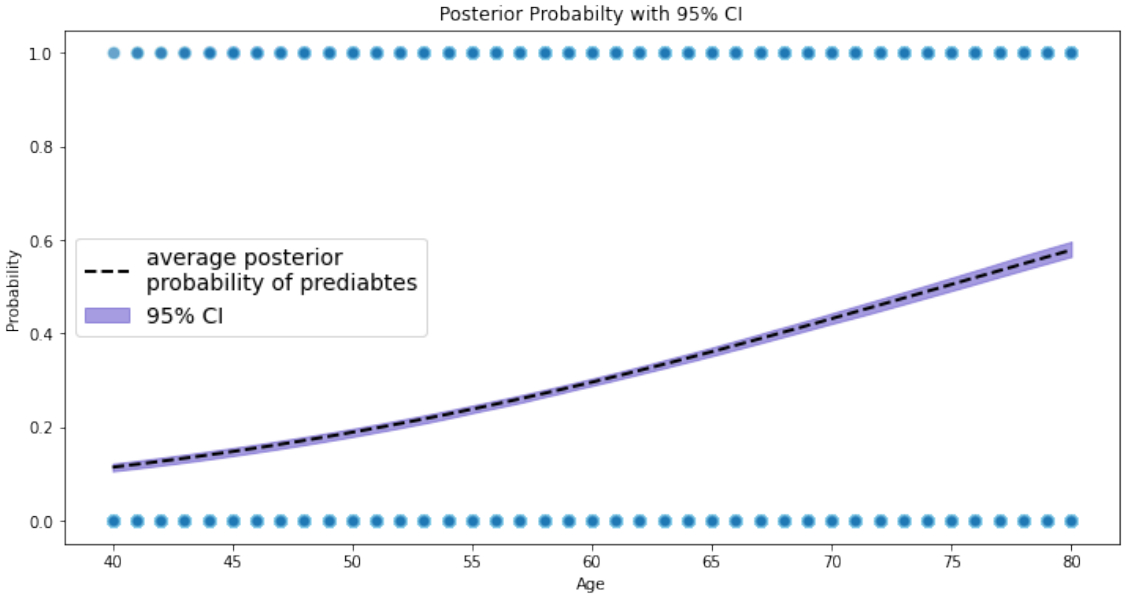


Figure 3.9: Illustration for the risk of transitioning from normoglycemic to prediabetes state with respect to age.

Fig. 3.9 shows the risk of prediabetes increases when age increase. For example, at age 50 years old, the prediabetes risk is around 20%, which increases up to 55%, at age 80 years old. It is shown that the prevalence of prediabetes in the age group 24-44 years, 45-64 years, and $\geq$ 65 years are 28.2 (CI: (24.4-32.4)), 44.9 (CI: (37.6-52.4)) and 49.5 (CI: (43.4-55.6)), respectively [63]. The trend in the prevalence is consistent with our model estimation for the results of prediabetes risks.

### 3.3.1 Discussion

According to the National Diabetes Statistics report [31], there are approximately 32 million individuals with diabetes in the US population, of which approximately 12 million instances are over 65 years old. This number constitutes about 25% of this age

group category. From Fig 3.9, we can see after age 64, approximately 25% of this age group is estimated to have prediabetes. This similarity in the population behavior is acceptable as the prediabetes population is converting to diabetes population with a rate of 51% in 10 years [65].

# Chapter 4

## A Markov Decision Process Approach for Statin Initiation in the Prediabetes Population

Both high glucose and high cholesterol levels are risk factors for heart diseases [18], and aging is a risk factor for diabetes and heart diseases [79], too. However, statin can be used to reduce the level of cholesterol but it has side effects such as raising the glucose levels [78] and body mass index [84]. Compared to the normoglycemic patients, the relative risks for coronary heart disease (CHD) and stroke in prediabetes patients are 1.10 and 1.06, respectively [45]. The relative risk associated with using statin for new-onset diabetes in the prediabetes population is 1.28 compared to normoglycemic [76]. According to an observational study, it was found that the existence of type 2 diabetes mellitus elevates CVD risk by two to four folds [3]. Therefore, it is important to consider the risk of moving to the diabetes state when cholesterol medication (statin) is given to prediabetes individuals. During 2011-2012, more than 27.9% of the US population whose their ages were more than 40 years old reported using a cholesterol-lowering medication [41]. Of this number, 83% reported using statin for cholesterol-lowering treatment, and the remaining 10% and 7% reported using both a statin and a non-statin and a non-statin, respectively.

According to the American College of Cardiology and the American Heart Association (ACC/AHA) guideline, Cholesterol medication is recommend if the estimated 10-year atherosclerotic cardiovascular disease risk is $\geq 7.5\%$ [83]. Our goal is to find the optimal decision strategy to initiate statin treatment that maximizes the discounted reward. A discrete-time finite horizon Markov decision process (MDP) model is developed to tackle

this problem for the prediabetes populations. Our model optimizes the starting time for statin treatment by considering the risk of moving to a terminal event (including CHD, stroke, death from causes other than CHD or stroke) or the diabetes state. A backward value iteration technique is used to find the optimal policy that provides the maximum value of accrued rewards.

## 4.1 Overview of Markov Decision Process

An MDP is a tool used for sequential decision-making under uncertainty. MDPs are categorized into two groups: the finite horizon and the infinite horizon MDP. Each category has its own analytical properties and solution algorithm. The optimal solution of finite horizon MDPs are approaching the optimal solution of infinite horizon MDPs if the planning horizons increase and both rewards and transition probabilities are stationary. MDPs are categorized as discrete-time if decisions are taken only at discrete time intervals, and continuous-time if decisions occur at any time [75]. The important components of MDPs are states, actions, the transition probabilities between states, and rewards. The information that is needed to decide how the system will evolve is encoded in states. Under Markovian assumption, the system's next state only depends on the current state and action, i.e., the past states do not have an effect on system evolvement. Our goal is to find the best action that maximizes the total expected rewards.

A patient's health condition at a given age is described by lipid ratio, systolic blood pressure, prediabetes, and gender. These are risk factors for diabetes, CHD, and stroke. We stop following a patient after an individual reaches an absorbing state. Rewards associated with moving to a living state are based on quality-adjusted life years gained. A patient accrues lump-sum rewards (i.e., total quality-adjusted life years) if they enters the diabetes state or accrued zero rewards if he/she enters one of the other termi-

nal states. The systolic blood pressure is needed as a covariate in the Framingham risk equations [2] which is used to estimate CHD and stroke risks for patients. The systolic blood pressure is modeled as a function of age, and it is estimated by fitting a cubic spline on the prediabetes population. Since statin reduces the lipid ratio and increases the risk of diabetes, we modeled the change over time in lipid ratio as a discrete-time Markov process [35] and moving stochastically to the diabetes state. At a given age $t$, a decision-maker chooses between either initiating statin treatment or waiting for the next decision epoch. At age $t+1$, a patient moves to one of the living states or to an absorbing state Fig. 4.1.

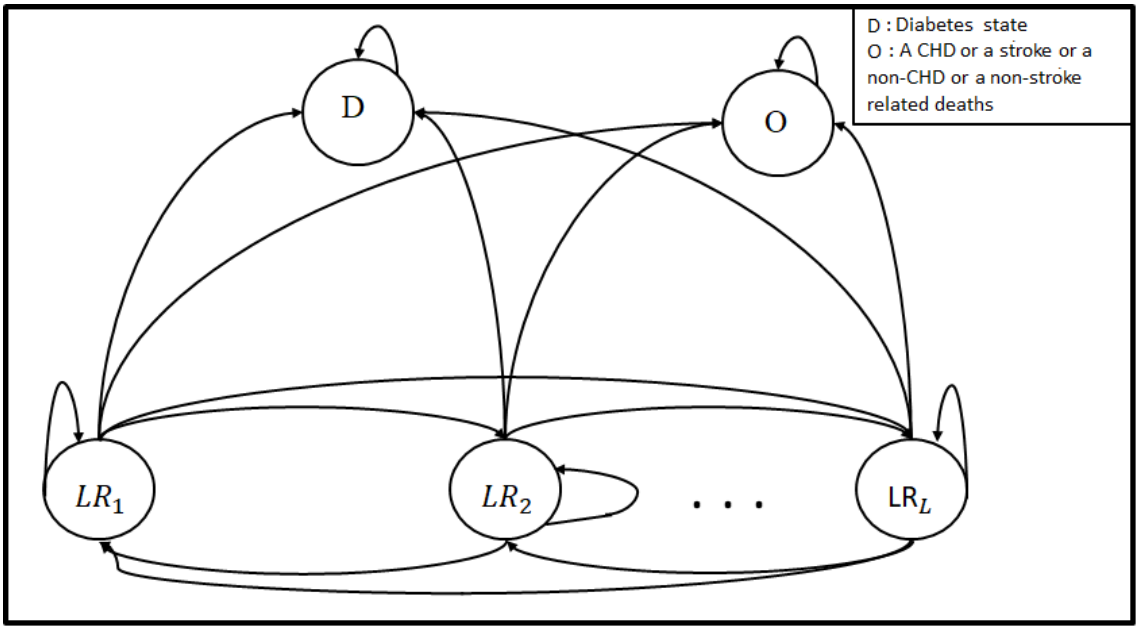


Figure 4.1: Illustration of the Markov model representing transitions among lipid ratio (LR) states, and absorbing states.

### 4.1.1 Model Formulation

Due to statin benefits for reducing the low-density lipoprotein and its side effects, the decision to initiate statin is revised annually starting from some age $t$. The set of decision epochs is denoted by $T$. A finite-horizon model is presented. At the end of the

decision horizon, which is the epoch $N-1$, the decision is made according to the best action that maximizes the reward, and the model parameters are assumed to be stationary beyond this epoch. That is, all decision epoch beyond $N-1$ are computed using $N-1$ epoch's parameters. Moreover, patients accrue lump-sum rewards instead of following them after the decision horizon $N-1$, see [24], [60]. If the decision is to initiate a treatment, then the treatment is used for the rest of the patient's life, otherwise, no treatment for the rest of the patient's life. It is also assumed that, once the decision is taken to initiate statin, the patients will continue using statins for the rest of their life, and the patient will use statins with perfect adherence. We describe the MDP components for our problem in the following sections.

### 4.1.2 States

The model has $L$ living states denoted by $\mathcal{L} = \{LR1, LR2, LR3, \ldots, LRL\}$ which are levels of lipid ratio ordered from low to high epoch and two absorbing states denoted by $\mathcal{D} = \{O, D\}$ which are fatal stroke or cardiovascular heart disease or deaths from other causes denoted by $O$, and the diabetes state denoted by $D$. At age $t$, we have $\bar{\mathcal{L}} = \{LR1, LR2, LR3, \ldots, LRL\} \cup \mathcal{D}$ states. The lipid ratio states have deterministic values of systolic blood pressure for a patient.

### 4.1.3 Actions

At any decision epoch $t \in T = \{1, 2, 3, \ldots, N\}$, the set of available actions are either to initiate statin treatment or to wait until the next decision epoch, which is after one year. We denote the action of starting treatment by 1 and the action to wait by 0. The set of available actions are denoted by $\mathcal{A} = \{0, 1\}$.

### 4.1.4 Transition Probabilities

To implement the MDP model, we need to estimate the transition probability from a living state at age $t$ to another living state at age $t+1$ and from a living state to the set of absorbing states $\mathcal{D}$. The probability of CHD or stroke events at state $s_t$ occurs with a probability $p_t^{CS}(s,a)$. The probability of fatal CHD or stroke at time $t$ is $(1-F_t)\,p_t^{CS}\,(s,a)$, where $F_t$ denotes for non-CHD or non-stroke related deaths and $a \in \mathcal{A}$. Therefore, the probability of moving to the absorbing state $O$ under treatment status $a \in \mathcal{A}$ at time $t$ is computed as follows: $p_t\,(O \mid s,a) = F_t + (1-F_t)\,p_t^{CS}\,(s,a)$. The probability of moving to the diabetes state is denoted by $\mathcal{W}_t\,(s' \mid s,a)$. The probability of moving to other living states in the next epoch is equal to the probability of not incurring a terminal events (absorbing states) $1-(p_t\,(O \mid s,a) + p_t\,(D \mid s,a))$ times the probability of moving to another living state $s'$ at age $t+1$ denoted by $q_t(s' \mid s,a$. The transition probability function is defined as follows:

$$p_t(s' \mid s,a) = \begin{cases} [1-(p_t\,(O \mid s,a) + p_t\,(D \mid s,a))]q_t(s' \mid s,a), & s,s' \in \mathcal{L} \\ F_t + (1-F_t)\,p_t^{CS}\,(s,a)\,, & s' = O, s \in \mathcal{L} \\ \mathcal{W}_t\,(s' \mid s,a)\,, & s' = D, s \in \mathcal{L} \\ 1, & s = s' \in \mathcal{D} \\ 0. & \text{Otherwise} \end{cases} \tag{4.1.1}$$

### 4.1.5 Rewards

The immediate reward for the patient in a state $s_t$ under treatment status $a_t = 0$ is $r_t\,(s_t,a_t) = 1$ QALY [24], and under treatment status $a_t = 1$ is $r_t\,(s_t,a_t) = 1-k$ QALY, where $k$ is pill disutility which includes dissatisfaction due to using statin daily and other pill side effects such as raising the chance of obesity during one year of life. Due to the low cost of statin, we are not considering stain cost in the reward func-

tion. The immediate reward accrued for a patient entering an absorbing state, a CHD or a stroke or a non-CHD or a non-stroke related death, under either of the actions is $r_t(s_t, a_t) = 0$. This assumption ensures a patient accrued a maximum number of QALYs prior to arriving in the absorbing state. However, the rewards associated with the diabetes states are the lump-sum rewards estimated using Kurt et al.'s work [53].

### 4.1.6 Optimization of Treatment Decision

The optimal policy of statin initiation considers trade-offs between the benefits of statin use and its side effects. At epoch $t$ and state $s$, we denote for the maximum total expected discounted rewards by $u_t^*(s)$ which is collected from time $t$ to the end of the time horizon. It can be calculated by using the following recurrence relation

$$\nu_t^*(s) = \max\{r_t(s_t, 0) + \lambda \sum_{s' \in \mathcal{L}} p_t(s' \mid s, 0)\, \nu_{t+1}^*(s'),\ \mu_t(s)\}, \qquad s \in \bar{\mathcal{L}}, \quad t = 1, \ldots, N-1, \tag{4.1.2}$$

where $\mu_t(s)$ is the patient's post-treatment discounted rewards computed in a separated Markov model. The value of $\mu_t(s)$ is the value of accrued rewards under treatment initiation in the state s at the time of epoch $t$ for the rest of life period, i.e.,

$$\mu_t(s) = r_t(s_t, 1) + \lambda \sum_{s' \in \mathcal{L}} p_t(s' \mid s, 1)\, \mu_{t+1}(s')\}, \qquad s \in \bar{\mathcal{L}}, \quad t = 1, \ldots, N-1. \tag{4.1.3}$$

At epoch $N$, the values of $\nu_N^*, \mu_N(s)$ are computed using $p_{N-1}$ values as follows:

$$\nu_N^*(s) = \max\{r_{N-1}(s_t, 0) + \lambda \sum_{s' \in \mathcal{L}} p_{N-1}(s' \mid s, 0)\, \nu_N^*(s'),\ \mu_N(s)\}, \qquad s \in \bar{\mathcal{L}}, \tag{4.1.4}$$

$$\mu_N(s) = r_t\left(s_t, 1\right) + \lambda \sum_{s' \in \mathcal{L}} \; p_{N-1}\left(s' \mid s, 1\right) \mu_N\left(s'\right)\}, \qquad s \in \bar{\mathcal{L}}. \tag{4.1.5}$$

A decision-maker chooses to start treatment if the value of $\mu_t(s)$ is greater than the value of delaying the treatment for one more decision epoch.

Since we stop following patients after age $N$, the decision at this epoch is used to find the action that gives us the higher rewards value for the whole post-horizon rewards. The calculation for the post-decision horizon rewards (PDHR) is done using the following formula:

$$E\left[PDHR \mid a_N, s_N\right] = max_{a_N \in \mathcal{A}} \; r_N\left(s_N, a_N\right) \sum_{i=1}^{M-N} i[1 - p\left(L \mid s_N, a_N\right] p\left(L \mid s_N, a_N\right)^{i-1} \lambda^i, \tag{4.1.6}$$

where $p(L \mid s_N, a_N) = 1 - p(O \mid s_N, a_N)$ denotes the probability of staying alive for one decision epoch after post-decision horizon $N$. The maximum age is denoted by M years.

## 4.2 Numerical Studies

We implement the above model for female and male patients with prediabetes, and evaluate the optimal policy with sensitivity for them. At a given age t, there are three lipid ratio states: low ($1 < LR1 \leq 3.8$), medium ($3.8 < LR2 \leq 5.5$), and high ($5.5 < LR3 \leq 8.7$), a diabetes state and a terminating state $O$. The probabilities of moving from living states to a fatal CHD or stroke or non-CHD or non-stroke related deaths are calculated using the Framingham risk equations [2]. The parameters that are needed in this risk equation are estimated from NHANES data sets [16]. Since the Framingham risk equations consider either diabetic or non-diabetic patients, we multiplied the risk of CHD or stroke by relative risk for prediabetes patients, which are 1.1 [45], 1.21 [57], re-

spectively. We assume the whole prediabetes population will move to the diabetes state at age 80. This is a plausible assumption because the probability of a prediabetes patient moving to the diabetes state is 51% in 10 years [65], and our decision horizon is around 40 years. Next, we used cubic splines on NHANES data to estimate patient systolic blood pressure as a function of age. The systolic blood pressure values are needed to estimate CHD and CVD risks using the Framingham risk equation. The probabilities of moving to the diabetes state are estimated using [58]. We computed these values by taking the average of rate of moving to the diabetes state in the group whose FBG is between 100 mg/dL and 110 mg/dL, and the group whose FBG is between 110 mg/dL and 125 mg/dL. The probabilities of moving to other living states in the next decision epoch are given in matrices Eq (3.1.4), Eq (3.1.3), Eq (3.1.5), and Eq (3.1.6).

The lump-sum reward for a patient entering the diabetic state is estimated using the values in Table 2 in [53]. We discretized the lipid ratio levels in [53] into three levels and assigned the average of the maximum expected QALY prior to the first terminal events to the first lipid ratio level [1, 3.8], similarly assigned the maximum expected QALY to the second and third lipid ratio intervals. The use of statin can reduce the lipid ratio levels by 0.19815 [53]. Under statin treatment, a transition probability for a patient to move to the CHD or stroke state is estimated using the Framingham risk equation after adjusting for the statin treatment effects. Under statin treatment, the diabetes risk is estimated by multiplying the diabetes risk by the relative risk associated with statin use which is 1.28 [32]. The mortality rates are from National Center for Health Statistics (NCHS) [15]. Table 5.2 summarizes the parameters that are used

in our model.

Table 4.1: Parameters Used in the Model

| Gender | Disease Progression | Parameters | Sources |
|---|---|---|---|
| Female | Risk of Progression to the diabetes state | 0.028 | Levitzky et al. [58] |
| | Coronary heart disease | Farmingham CHD Risk Equation | Anderson et al. [2] |
| | Stroke Risk | Farmingham Stroke Risk Equation | Anderson et al. [2] |
| | Diabetes Relative Risk Associated With using statin | 1.28 | Freeman et al. [32] |
| | Transition probabilities among Lipid ratios | | Eq. (3.1.3) & Eq. (3.1.4) |
| Male | Risk of Progression to the diabetes state | 0.024 | Levitzky et al. [58] |
| | Coronary heart disease | Farmingham CHD Risk Equation | Anderson et al. [2] |
| | Stroke Risk | Farmingham Stroke Risk Equation | Anderson et al. [2] |
| | Diabetes Relative Risk Associated With using statin | 1.28 | Freeman et al. [32] |
| | Transition probabilities among Lipid ratios | | Eq. (3.1.5) & Eq. (3.1.6) |
| | pill disutility | 0.02 | Pignone et al. [73] |

### 4.2.1 Results

As shown in Figure (4.2), our model recommends early statin use for medium and high lipid ratio for 0.98 QALY for statin use and 1 QALY for non-statin users. For the low

lipid ratio state, statin initiation is recommended for females who are 50 years old and males who are 55 years old. For states with a tie between starting a treatment or waiting for the next decision epoch, we picked an action that is similar to the action that was taken in the previous states. The existence of optimal control limit policy is proved in [53]. The existence of optimal control limit policy grants that the treatment is initiated whenever LR is greater than some threshold. The maximum expected QALYs gained are given in Table 5.2.

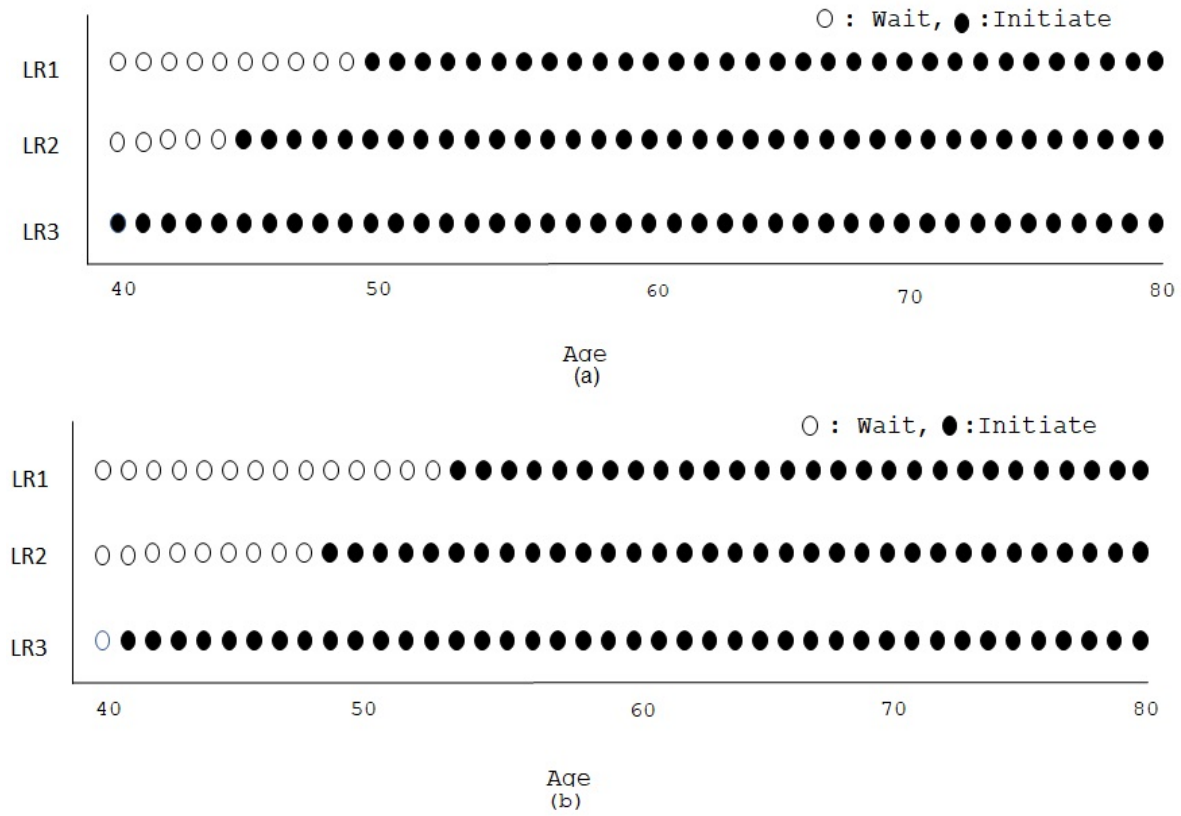


Figure 4.2: Optimal time for statin initiation is depicted with respect to the three lipid ratio levels,(a) for a female and (b) for a male. The disutility used in this implementation is 0.02 and the diabetes relative risk associated with statin use is 1.28.

Table 4.2: The maximum expected QALYs prior to the first terminal events for prediabetes patients are depicted for 0.98 QALY for statin use and 1 QALY for no-statin.

| **Gender** | **Age** | **LR1** | **LR2** | **LR3** |
|---|---|---|---|---|
| female | | | | |
| | 40 | 21.440 | 21.386 | 21.324 |
| | 50 | 18.617 | 18.594 | 18.554 |
| | 60 | 15.488 | 15.454 | 15.395 |
| | 70 | 11.925 | 11.877 | 11.795 |
| | 80+ | 8.153 | 7.443 | 6.628 |
| Male | | | | |
| | 40 | 20.400 | 20.332 | 20.239 |
| | 50 | 17.499 | 17.448 | 17.410 |
| | 60 | 14.241 | 14.199 | 14.145 |
| | 70 | 10.374 | 10.318 | 10.248 |
| | 80+ | 5.697 | 5.097 | 4.403 |

### 4.2.2 Sensitivity Analysis

There are different estimates for the diabetes risk associated with statin use [76], [78], [49]. We implement the model with different values for the relative risk associated with statin. With a higher value for the relative risks associated with statin 1.5, the statin use is not recommended for a low-level lipid ratio state. Furthermore, the risk of progression to the diabetes state in two years has been estimated by Tabak et al. [86] in the range from 5% to 10%, but we use the relative risks values 2.8% for a female and 2.4% for a male, which are taken from Levitzky et al. [58]. Using the above parameters, our model does not recommend statin use for LR1 only at a later age. This is because of the increased risk of moving to the diabetes state.

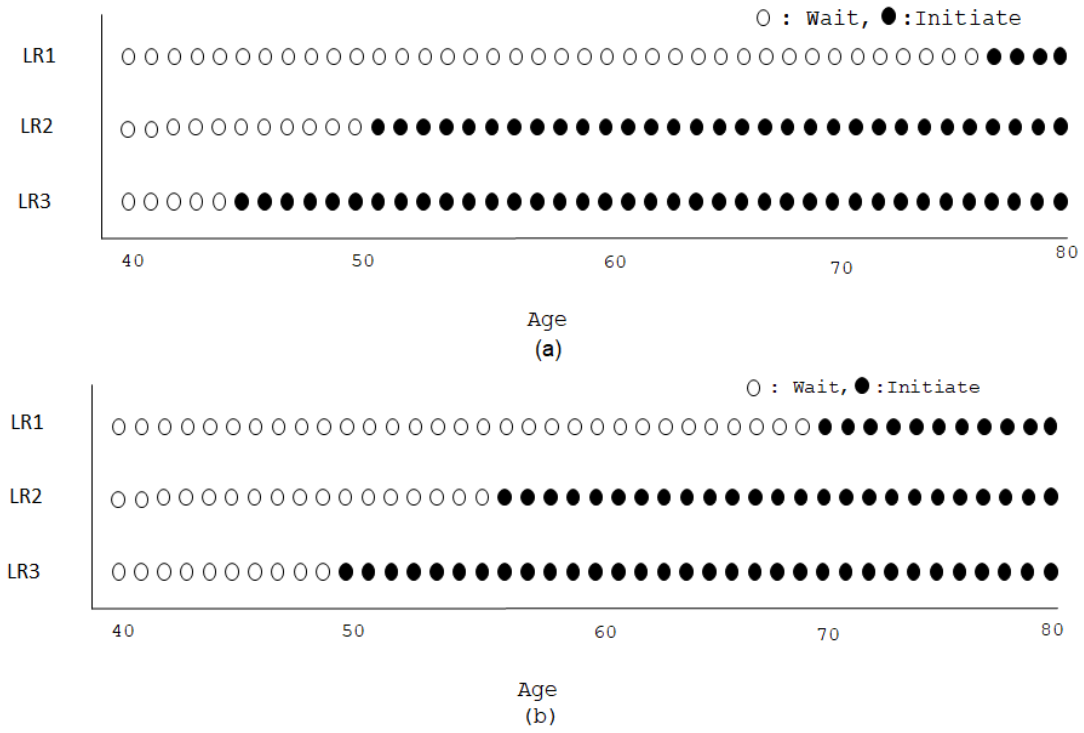


Figure 4.3: Optimal times for statin initiation are depicted with respect to the three lipid ratio levels. The relative risk for diabetes associated with using statin is 1.5. The QALY value is 0.96 for one year of statin use.

Table 4.3: The maximum expected QALYs prior to the first terminal events for prediabetes patients are depicted for 0.96 QALY for statin use and 1 QALY otherwise. The relative risk of diabetes associated with using statin is 1.5.

| **Gender** | **Age** | **LR1** | **LR2** | **LR3** |
|---|---|---|---|---|
| female | | | | |
| | 40 | 21.201 | 21.124 | 21.011 |
| | 50 | 18.281 | 18.187 | 18.138 |
| | 60 | 15.154 | 15.105 | 15.039 |
| | 70 | 11.738 | 11.670 | 11.579 |
| | 80+ | 7.746 | 7.071 | 6.297 |
| Male | | | | |
| | 40 | 20.156 | 20.076 | 19.968 |
| | 50 | 17.169 | 17.055 | 16.922 |
| | 60 | 13.889 | 13.806 | 13.747 |
| | 70 | 10.158 | 10.098 | 10.021 |
| | 80+ | 5.412 | 4.842 | 4.183 |

Due to different estimates for the relative risk associated with statin use, a sensitivity analysis is performed to investigate its effects on the optimal policy. When QALY associated with statin use is decreased from 0.98 QALY to 0.96 QALY, the optimal policy does not recommend statin use for a low-level lipid ratio. In addition, the maximum expected QALYs prior to the first terminal events have also increased.

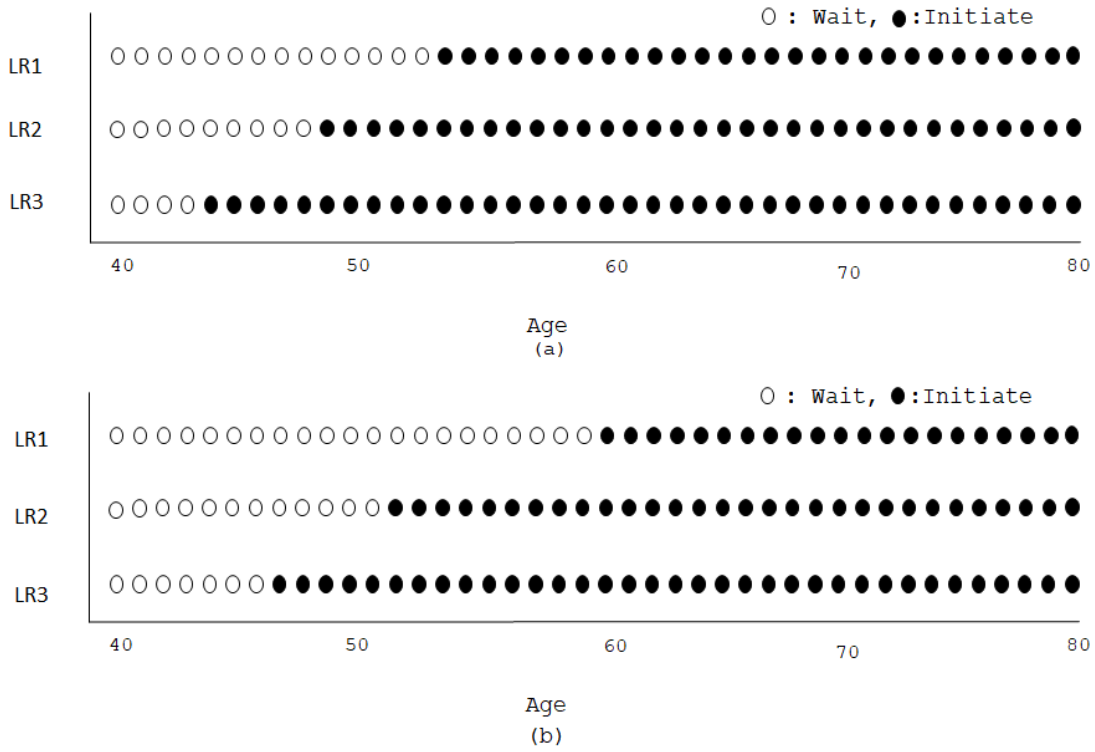


Figure 4.4: Optimal time for statin initiation is depicted with respect to the three lipid ratio levels. The disutility used in this implementation is 0.04, and the diabetes risk associated with using statin is 1.28.

Table 4.4: The maximum expected QALYs prior to the first terminal events for prediabetes patients are depicted for 0.96 QALY for statin use and 1 QALY for no-statin.

| **Gender** | **Age** | **LR1** | **LR2** | **LR3** |
|---|---|---|---|---|
| female | | | | |
| | 40 | 21.285 | 21.213 | 21.106 |
| | 50 | 18.401 | 18.346 | 18.307 |
| | 60 | 15.282 | 15.248 | 15.189 |
| | 70 | 11.789 | 11.741 | 11.660 |
| | 80+ | 7.746 | 7.071 | 6.297 |
| Male | | | | |
| | 40 | 20.252 | 20.174 | 20.067 |
| | 50 | 17.304 | 17.211 | 17.151 |
| | 60 | 14.028 | 13.985 | 13.931 |
| | 70 | 10.235 | 10.180 | 10.110 |
| | 80+ | 5.412 | 4.842 | 4.183 |

## 4.3 Conclusion

Statin recommendations are sensitive to the dis-utility associated with its use and its side effects [68], [43]. Our model consists of two parts. First, we compute the value function at a given age $t$ as a statin treatment initiated for the rest of his/her life. Then we use these values to find the best action either to start or delay the treatment depends on the action that maximizes the value function in Eq (4.1.2). Ridker et al. [76] have suggested that the benefit of using statin overweigh the negative side effects. However, our policy shows that statin is not recommended only at late a age for patients with a low level of lipid ratio with respect to some scenarios, for example, the case in which pill reward associated with statin use is 0.96 QALY, see Fig. 4.4. Moreover, the risk of a new-onset diabetes has a crucial role in delaying a statin initiation treatment because statin use expedites the progression to the diabetes state and; consequently, moving to the absorbing state. When the relative risk for a new-onset diabetes asso-

ciated with statin use is raised from 1.28 to 1.5, statin is not recommended for a low level-lipid ratio except for the old-aged population. In this model, we have applied 1.5 relative risks, and it shows a higher impact on our policy, see Fig. 4.3. As previously mentioned, the relative risk reaches 2.62 for using statin for more than two years.

Limitations in our study are due to different factors. First, we did not consider patients adherence to the treatment in our modeling. Low adherence can reduce the progression rate to diabetes state. Second, treatment doses can impact on progression rate to the diabetes state and; consequently, the value of the QALYS gained. Lastly, we did not consider different races in our numerical example. Different races have different CVD risk values.

# Chapter 5

# A Robust Markov Decision Process Approach for Statin Initiation Decision in the Prediabetes Population

## 5.1 Introduction

Dynamic programming is generalized to be more robust to account for the variation in the model parameter estimation. Robust dynamic programming focuses on mitigating the effects of parameter uncertainty while avoiding conservative policies. In other words, robust dynamic programing uses a "max-min" approach to find a policy that maximizes the worst-case policy performances for different transition probability matrices in an ambiguity set. One approach to model the uncertainty parameters is called stochastic programming in which the parameters are modeled using a probability distribution. For example, stochastic programming is used to model the demand for energy. The demands vary with respect to weather, economic conditions and other factors. Therefore, the demand can be modeled using the normal distribution [71]. Another approach to tackle optimization under uncertainty is to ensure the constraints are satisfied with a specific probability, say $1-\epsilon$. This approach is called chance constraint [74]. The decision-makers are able to choose risk levels by specifying the probability for the constraints to be satisfied. The formulation of chance programming is as follows:

$$\begin{aligned} &\underset{x}{\text{minimize}} && f(x) \\ &\text{subject to} && \\ & && P\{\xi \in \Theta |\, G(x,\xi) \leq 0\} \geq 1-\epsilon, \end{aligned} \tag{5.1.1}$$

where the decision variables are denoted by $x$; the objective function is denoted by $f(x)$; the set of constraints is denoted by $G(x)$; the uncertainty set of a given probability distribution $P$ of the random variable $\xi$ is denoted by $\Theta$. The constraint in Eq. (5.1.1) ensures that the decision variables $x$ satisfy the constraint with a probability $1 - \varepsilon$. One of the limitations of this approach is computational complexity due to the non-convexity of the feasible region even a convexity of the feasible region with respect to $x$ is satisfied for some realization of $\xi$ [74].

Robust optimization (RO) is another approach to tackle uncertainty in the parameters. The first main contribution to RO is by Soyster [82] who developed robust linear programming. Robust optimization has many applications in portfolio optimization [39], healthcare [97], and many other fields of operations research; see Ben-Tal et al. [7] and the references therein.

Robust Markov decision processes (RMDPs) have been introduced in the work of Iyengar [46] and Nilim et al. [67]. The development of RMDP is rooted in robust optimization. RMDPs are developed to address the variability that arises in estimating a transition probability matrix for MDPs. Ambiguity in parameter estimation plays a crucial rule in determining the optimal policy for RMDP. Transition probabilities are subject to variability due to the environment, missing data, measurement errors, estimation errors, or some other reasons. RMDPs are also categorized into the finite-horizon and the infinite-horizon problems, which are the same as MDP. A finite-horizon RMDP is represented by the tuple:

$$(\tau, \mathcal{S}, \mathcal{P}_t^a, \mathcal{A}_s, r_t(s,a)),$$

where the set of decision epochs is denoted by $\tau = \{1, 2, \cdots, T\}, \quad T < \infty$; the set of states is denoted by $\mathcal{S} = \{1, 2, \cdots, S\}$; the collection of transition probability matrices at time $t$ under action $a$ is denoted by $\mathcal{P}_t^a$; the set of actions that are allowed at state $s$ is $\mathcal{A}_s$; and the accrued reward at state $s$ under action $a$ at time $t$ is denoted by

$r_t(s,a)$. At each time $t$ and action $a$, we only know the transition probability $P^a_{s,t}$ lies in a given subset $\mathcal{P}^a$ of matrices that satisfy Markov transition probabilities (rows summing to one and nonnegative entries.)

RMDPs are hard to tackle for a general uncertainty set, and the computational complexity is more than NP-hard [55]. However, under some restrictions on the uncertainty set the problems are solvable. An example of the restriction on the uncertainty set is *a rectangular property* [67], [46]. A set $\mathcal{P}^a_t$ satisfies the rectangular property if for every action $a \in \mathcal{A}$, $\mathcal{P}^a_t$ can be written as:

$$\mathcal{P}^a_t = \mathcal{P}^a_{1,t} \times \mathcal{P}^a_{2,t} \times \cdots \times \mathcal{P}^a_{n,t},$$

where $\mathcal{P}^a_{i,t}$ denotes the uncertainty set of the ith row of the transition probability matrix $P^a_t$. In other words, rectangularity means the transition probability for each state-action pair is independent of the others. The set $\mathcal{P}^a_t$ can be considered as a set of confidence for the transition matrices. Depending on the uncertainty of transition matrices, there are two models: *stationary uncertainty model* and *time-varying uncertainty model*. In the first model the transition probabilities are fixed with respect to time, while in the second model, they are varying with respect to time. The second model is used in our problem modeling.

Solution methods for the RMDP is developed by Nilim et al. [67] and Iyenger [46] under rectangular property for the uncertainty set. The robust value function for state $s$ at time $t$ is given by the following Bellman equation:

$$\nu_t(s) = \max_{a \in \mathcal{A}_s} \{r_t(s,a) + \min_{p \in \mathcal{P}^a_{s,t}} \sum_{s' \in \mathcal{S}} p(s')\nu_{t+1}(s')\} \quad \text{for} \quad t = 1, \cdots, N-1 \quad \text{and} \quad s \in \mathcal{S}, \tag{5.1.2}$$

and

$$\nu_T(s) = r_T(s_T) \quad \forall s_T \in \mathcal{S}, \tag{5.1.3}$$

where $p$ is a specific row of a particular transition matrix in the uncertainty set of this row which is $\mathcal{P}^a_{i,t}$. The rewards that a patient accrued when they arrives at the end of the decision horizon is denoted by $r_T(s_T)$. The minimization problem part in the Eq. (5.1.2) is called the inner problem which is denoted by $\sigma^a_{t,s}$. The goal of the inner problem is to find the minimum solution on the uncertainty set of a specific row of a transition probability matrix. The construction of the uncertainty sets $\mathcal{P}^a_t$ by using a likelihood constraint approach approach grantees that the uncertainty sets represent a precise description of statistical uncertainty on the transition matrix. Then the solution of the interior problem is given by the bisection method [67].

The following algorithm is used to solve RMDP:

**Algorithm 1** Robust Finite-Horizon Dynamic Programming Algorithm, Adapted from [67]

**Require:** Initialize the value function of all states to its terminal value $\hat{\nu}_T = r_T(s_T)$
  **while** $t \neq 0$ **do**
    for every state $s_t \in S$ and action $a \in \mathcal{A}$, use the bisection algorithm to solve the inner problem and update $\hat{\nu}_{t-1}(s_t) = \max_{a \in \mathcal{A}} \{ r_t(s_t, a) + \lambda \hat{\sigma}^a_{t,s} \ \}$

    the optimal action is the one that maximizes $\hat{\mathcal{A}}(s_t) \leftarrow \operatorname{argmax}_{a \in \mathcal{A}} \{ r_t(s_t, a) + \lambda \hat{\sigma}^a_{t,s} \}$

    Replace $t$ by $(t-1)$
  **end while**

A robust medical decision model is introduced in this chapter to help decision-makers optimize statin initiation treatment for prediabetes populations. An optimal policy is computed using robust backward value iteration to solve the Markov decision process problems [67]. Our model is developed considering the uncertainty in the transition probability matrix. A sensitivity analysis for the optimal policy is studied for the re-

wards and pill dis-utility.

## 5.2 Robust Markov Decision Process Formulation

A finite-horizon, discrete-state robust MDP is formulated to find the optimal statin initiation policies for prediabetes patients. The structure of this model is similar to the structure of the MDP model in Chapter 4, but the RMDP approach is different. In this chapter, our goal is to find the optimal policy that remains optimal even with a small perturbation in the transition probabilities, i.e., optimizing the worst-case scenario. Let the set of available actions at time $t$ for all states be denoted by $\mathcal{A}_t = \{0, 1\}$. A decision-maker chooses one of the two available actions at any decision epoch $t \in \{1, 2, 3, \ldots, N\}$. If a decision-maker chooses action 0, statin initiation is postponed to the next decision epoch, and if action 1 is chosen, then treatment is started for the rest of the decision horizon. At each decision epoch $t$, the system occupies one of the $L$ states that are classified according to the metabolic state. The living states are denoted by $\mathcal{L} = \{LR1, LR2, LR3, \ldots, LRL\}$, and the absorbing states are denoted by $\mathcal{D} = \{O, D\}$ which are fatal stroke or cardiovascular heart disease or deaths from other causes denoted by $O$, and the diabetes state is denoted by $D$. At age $t$, we have $\bar{\mathcal{L}} = \{LR1, LR2, LR3, \ldots, LRL\} \cup \mathcal{D}$ states.

The transition probabilities that are needed in our implementation of a robust MDP model are transition from a living state at age $t$ to other living states at age $t + 1$, and from a living state at age $t$ to the set of absorbing states $\mathcal{D}$. At epoch $t$ and state $s$ the probability of CHD or a stroke event is denoted by $p_t^{CS}(s, a)$. The probability of fatal CHD or stroke is $(1 - F_t)\, p_t^{CS}(s, a)$, where $F_t$ denotes for non-CHD or non- stroke related deaths and $a \in \mathcal{A}$. Therefore, the probability of moving to the absorbing state $O$ under treatment status $a \in \mathcal{A}$ at time $t$ is computed as follows: $p_t(O \mid s, a) =$

$F_t + (1 - F_t)\, p_t^{CS}(s, a)$. The probability of moving to the diabetes state is denoted by $\mathcal{W}_t(s' \mid s, a)$, and the probabilities of moving to other living states in the next epoch are equal to the probability of not incurring terminal events $1-(p_t(O \mid s, a)+p_t(D \mid s, a))$ times the probability of moving to another living state $s'$ at age $t+1$ (refer to Eq. (4.1.1)).

The immediate reward for a patient in a state $s_t$ under treatment status $a_t = 0$ is $r_t(s_t, a_t) = 1$ QALY [24], and under treatment status $a_t = 1$ is $r_t(s_t, a_t) = 1 - k$ QALY, where $k$ is pill dis-utility. The immediate reward accrued for a patient entering an absorbing state, a CHD or a stroke or a non-CHD or a non-stroke related death, under either of the actions is $r_t(s_t, a_t) = 0$. However, the rewards associated with the diabetes state is the lump sum rewards estimated using Kurt et al.'s work [53].

Our goal by solving the RMDP problem is to find the optimal policy that collects the largest number of rewards. At epoch $t$ and state $s$, we denote the maximum total expected discounted rewards prior to the death by $u_t^*(s)$, which are collected from state $t$ to the end of the time horizon. It can be calculated by using the recurrence relation given in the following equation:

$$\nu_t(s) = \max\{r_t(s,0) + \min_{p \in \mathcal{P}^0_{s,t}} \sum_{s' \in \mathcal{S}} p(s')\nu_{t+1}(s'), \mu_t(s)\} \quad \text{for} \quad t = 1, \cdots, N-1 \quad \text{and} \quad s \in \mathcal{S}, \tag{5.2.1}$$

where

$$\mu_t(s) = r_t(s_t, 1) + \min_{p \in \mathcal{P}^1_{s,t}} \sum_{s' \in \mathcal{S}} p(s')\mu_{t+1}(s') \quad \text{for} \quad t = 1, \cdots, N-1 \quad \text{and} \quad s \in \mathcal{S}. \tag{5.2.2}$$

The patient's post-treatment discounted rewards $\mu_t(s)$ is computed in a separated Markov model. The value of $\mu_t(s)$ (in Eq.(5.2.2)) is an estimation of accrued rewards under

treatment initiation in state $s$ at time epoch $t$ for the remaining life period. The decision-maker chooses to start treatment if the value of $\mu_t(s)$ is greater than the value of delaying the treatment for one more decision epoch. Since we stop following patients after age $T$, the decision at epoch $N-1$ is to find the action that gives us the higher rewards value for the whole post-horizon rewards. The calculations for the post-decision horizon rewards (PDHR) are completed using Eq.(4.1.6).

## 5.3 Structural Properties

A sufficient condition for the existence of an optimal control limit policy of a stationary infinite-horion MDP was proved by Alagoz et al. [1]. Kurt et al. [53] extended the result to a non-stationary finite- horizon MDP. In this section, we introduce a sufficient condition for the existence of the optimal control limit policy of the RMDP. We followed a similar technique that developed by Kurt et al. [53]. By using of the monotonicity of the value function, and consequently the expected benefit-loss function we introduce a threshold for the optimal control limit policy of an RMDP.

First we present the definition of the increasing failure rate (IFR) of a transition probability matrix [6].

**Definition 5.3.1** ([6]). *Let $B = [b_{ij}]_{n\times n}$ be a stochastic matrix. If $\sum_{j=1}^{l} b(j|i)$ is nonincreasing in $i$ for every $l \in \{1, \cdots, n\}$, then $B$ is said to have the IFR property.*

Next, we define Assumption 1 as follows.

- If a matrix satisfies the IFR property, then it satisfies the following condition: for $\ell > s$, we have

$$\min_{p\in\mathcal{P}^a_{s,t}} \sum_{s'\in\mathcal{S}} p(s')\nu_{t+1}(s') \geq \min_{p\in\mathcal{P}^a_{\ell,t}} \sum_{s'\in\mathcal{S}} p(s')\nu_{t+1}(s'). \tag{5.3.1}$$

This assumption may be interpreted as the worst case-scenario for a sicker patient is more probable to become worse. This assumption is proved for an MDP case in [75] (Proposition 4.7.3), but for an RMDP we make it as an assumption to prove the structural property of an optimal policy.

The expected benefit loss caused by postponing treatment initiation until the next decision epoch is denoted by $B_t(s)$, which is defined as follows

$$B_t(s) = \mu_t(s) - \lambda \min_{p \in \mathcal{P}^0_{s,t}} \sum_{s' \in \mathcal{S}} p(s')\mu_{t+1}(s'). \tag{5.3.2}$$

**Theorem 5.3.1.** *If Assumption 1 is satisfied, then the optimal value function $\nu_t(s)$ is nonincreasing in $s \in \mathcal{L}$.*

*Proof.* For $\ell \geq s$, from Assumption 1, and $r_t(s,1)$, the post-treatment rewards are nonincreasing in $s$, i.e., $\mu_t(s) \geq \mu_t(\ell)$. Moreover, from Assumption 1 and $r_t(s,0)$ being nonincreasing in $s$, we can see the first term in the maximization function is also nonincreasing. Hence, the $\nu_t(s)$ is nonincreasing in $s$. □

At epoch $t = N$ and iteration $k+1$, the values of $\nu_N^{k+1}(s)$ and $\mu_N^{k+1}(s)$ of the value iteration algorithm function are defined as follows:

$$\nu_N^{k+1}(s) = \max\{r_{N-1}(s,0) + \min_{p \in \mathcal{P}^0_{s,N-1}} \sum_{s' \in \mathcal{S}} p(s')\nu_N^k(s'), \mu_N^{k+1}(s)\},$$

and $\mu_N^{k+1}(s) = r_{N-1}(s,1) + \min_{p \in \mathcal{P}^1_{s,N-1}} \sum_{s' \in \mathcal{S}} p(s')\mu_N^k(s)$.

**Theorem 5.3.2.** *Let $B_t(s)$ be nondecreasing in $s$ for all $t$. Then there exists $s_t*$ such that if $a_t(s*) = 1$, then for $s <= s*, a_t(s) = 0$, and for $s > s*, a_t(s) = 1$, for all $t$.*

*Proof.* It is enough to show $\rho_t(s) = \nu_t(s) - \mu_t(s)$ is nonincreasing in $s \in \mathcal{L}$ for all $t \in T$,

because the value of $\nu_t(s)$ can be one of the two values of the maximizing function in Eq. (5.2.1). If $\nu_t(s) = \mu_t(s)$, then $\rho_t(s) = 0$, and the nonincreasing property of $B_t(s)$ ensures the action will be the same for the remaining $t$ values. The proof of this theorem is accomplished by an induction on value iteration's iterates, $k$. Consider the following cases:

- Case I: Let $t_0$ denote any epoch different than epoch $N$.
  - If $\nu_{t_0}^{k+1}(s+1) = \mu_{t_0}(s+1)$, then the function $\rho_t(s)$ is nonincreasing. It can be seen by just plugging the values of $\nu_{t_0}$, i.e.,

$$\nu_{t_0}^{k+1}(s) - \mu_{t_0}(s) \geq \nu_{t_0}^{k+1}(s+1) - \mu_{t_0}(s+1).$$

  - Otherwise, the value function $\nu_{t_0}^{k+1}$ takes the first argument in Eq. (5.2.1), i.e., $\nu_{t_0}^{k+1}(s+1) \geq \mu_{t_0}(s+1)$. We assume the following inequality is correct for some iteration $k \geq 0$, $\nu_{t_0}^{k}(s) - \mu_{t_0}(s) \geq \nu_{t_0}^{k}(s+1) - \mu_{t_0}(s+1)$, and prove it correct at higher iterations of the value iteration algorithm $\nu_{t_0}^{k+1}(s) - \mu_{t_0}(s) \geq \nu_{t_0}^{k+1}(s+1) - \mu_{t_0}(s+1)$.

    From Eq. (5.2.1), the value function $\nu_{t_0}^{k+1}(s) = r_{t_0}(s,0) + \min\limits_{p \in \mathcal{P}_{s,t_0}^0} \sum_{s' \in \mathcal{S}} p(s')\nu_{t_0}^{k}(s')$. Let us consider

$$\begin{aligned}
\nu_{t_0}^{k+1}(s) - \nu_{t_0}^{k+1}(s+1) &\overset{?}{\geq} \min_{p \in \mathcal{P}_{s,t_0}^0} \sum_{s' \in \mathcal{S}} p(s')\nu_{t_0}^{k}(s') - \min_{p \in \mathcal{P}_{s+1,t_0}^0} \sum_{(s+1)' \in \mathcal{S}} p((s+1)')\nu_{t_0}^{k}((s+1)') \\
&\overset{??}{\geq} \min_{p \in \mathcal{P}_{s,t_0}^0} \sum_{s' \in \mathcal{S}} p(s')\mu_{t_0}^{k}(s') - \min_{p \in \mathcal{P}_{s+1,t_0}^0} \sum_{s' \in \mathcal{S}} p(s')\mu_{t_0}^{k}(s+1') \\
&\overset{???}{\geq} \mu_{t_0}(s) - \mu_{t_0}(s+1).
\end{aligned} \tag{5.3.3}$$

    The first inequality comes from plugging in the values of $\nu_{t_0}^{k+1}(s)$ and $\nu_{t_0}^{k+1}(s+$

1). The second inequality is due to the fact that TPM is IFR and satisfies Assumption 1:

$$\min_{p\in\mathcal{P}^a_{s,t_0}} \sum_{s'\in\mathcal{S}} p(s')(\nu^k_{t_0}-\mu_{t_0})(s') \geq \min_{p\in\mathcal{P}^a_{s+1,t_0}} \sum_{(s+1)'\in\mathcal{S}} p((s+1)')(\nu^k_{t_0}-\mu_{t_0})((s+1)'). \tag{5.3.4}$$

The third inequality in Eq. (5.3.3) comes from the induction step. From the non-increasing property of $B_{t_0-1}(s)$, we get that the induction inequality is satisfied.

- Case II: Let $t$ denote the last epoch ($t = N$).

  We assume $\nu^k_N(s) - \mu_N(s) \geq \nu^k_N(s+1) - \mu_N(s+1)$ is nonincreasing in iteration $k \geq 0$ and prove the relation is correct in a higher iteration of the value iteration algorithm, i.e., $\nu^{k+1}_N(s) - \mu_N(s) \geq \nu^{k+1}_N(s+1) - \mu_N(s+1)$.

  From Eq.(5.2.1), if $\nu^{k+1}_N(s) = \mu_N(s)$, then the relation is satisfied. Next, if $\nu^k_N(s) = r_{N-1}(s,0) + \min_{p\in\mathcal{P}^0_{s,N-1}} \sum_{s'\in\mathcal{S}} p(s')\nu_{k+2}(s')$, we consider

$$\begin{aligned}
\nu^{n+1}_N(s) - \nu^{n+1}_N(s+1) &\overset{?}{\geq} \min_{p\in\mathcal{P}^0_{s,N-1}} \sum_{s'\in\mathcal{S}} p(s')\nu^n_N(s') - \min_{p\in\mathcal{P}^0_{s+1,N-1}} \sum_{(s+1)'\in\mathcal{S}} p((s+1)')\nu^n_N((s+1)') \\
&\overset{??}{\geq} \min_{p\in\mathcal{P}^0_{s,N-1}} \sum_{s'\in\mathcal{S}} p(s')\mu^n_N(s') - \min_{p\in\mathcal{P}^0_{s+1,N-1}} \sum_{s'\in\mathcal{S}} p(s')\mu^n_N(s+1') \\
&\overset{???}{\geq} \mu_{N-1}(s) - \mu_{N-1}(s+1).
\end{aligned} \tag{5.3.5}$$

By plugging in the values of $\nu^{n+1}_N(s)$, and $\nu^{n+1}_N(s+1)$, we get the first inequality. The second inequality is due to the fact that TPM is IFR and satisfies Assumption 1:

$$\min_{p\in\mathcal{P}^a_{s,t}} \sum_{s'\in\mathcal{S}} p(s')(\nu^n_N - \mu_N)(s') \geq \min_{p\in\mathcal{P}^a_{s+1,t}} \sum_{(s+1)'\in\mathcal{S}} p((s+1)')(\nu^n_N - \mu_N)((s+1)'). \quad (5.3.6)$$

The third inequality in Eq. (5.3.5) comes from the induction step. From the non-increasing property of $B_{N-1}(s)$, we obtain that the induction inequality is satisfied.

□

Our next theorem is to show the existence of an optimal policy.

**Theorem 5.3.3.** *For all $s \in \mathcal{L}$, the value function $\nu_t(s)$ is nonincreasing in $t; t \in T$.*

*Proof.* The result is proved using backward induction. Since the value function $\nu_t(s)$ consists of two parts, we prove the nonincreasing property for both. First, we consider post-treatment reward $\mu_t(s)$. Assume the relation is true for some epochs, say $k+2 \leq N-1$, i.e. $\mu_{k+1} \geq \mu_{k+2}$. Let $\ell \in \mathcal{L}'$. We need to show $\mu_k(\ell) - \mu_{k+1}(\ell) \geq 0$. Consider the value of $\mu_k(\ell) - \mu_{k+1}(\ell,$

$$\begin{aligned}
\mu_k(\ell) - \mu_{k+1}(\ell) &= \min_{p\in\mathcal{P}^1_{\ell,k}} \sum_{s'\in\mathcal{S}} p(s')\mu_{k+1}(s') - \min_{p\in\mathcal{P}^1_{\ell,k+1}} \sum_{s'\in\mathcal{S}} p(s')\mu_{k+2}(s') \\
&\overset{\dagger}{\geq} \min_{p\in\mathcal{P}^1_{\ell,k}} \sum_{s'\in\mathcal{S}} p(s')(\mu_{k+1}(s') - \mu_{k+2}(s')) \\
&\overset{\dagger\dagger}{\geq} 0.
\end{aligned} \quad (5.3.7)$$

The first inequality, with †, is correct because, at higher epoch $k+1$, the risk of moving to the absorbing state is higher than the risk of moving to the absorbing state at

epoch $k$. The second inequality, with ††, comes from the induction step. Using the same technique, we can show the value function is nonincreasing. □

**Theorem 5.3.4.** *Assume $B_t(s)$ is nonincreasing in $t \in T'$ and for all $s \in \mathcal{L}'$. Then, if a treatment is recommended to initiate in epoch $t$, it is recommended for the rest of the patient's life, i.e., if $a_t(s) = 1 \to a_{t+1}(s) = 1$.*

*Proof.* Consider $\rho_t(s) = \nu_t(s) - \mu_t(s)$. The value of $\nu_t(s)$ takes one of the two values of the maximizing function in Eq. (5.2.1). If $\nu_t(s) = \mu_t(s)$, then $\rho_t(s) = 0$. Therefore, it is enough to show $\rho_t(s)$ is non-increasing in $t \in T$ for all $s \in \mathcal{L}$. The proof is done using backward induction on $t$.

Suppose the relation is correct for some $k+1 < N$, i.e., $\nu_{k+1}(s) - \mu_{k+1}(s) \geq \nu_{k+2}(s) - \mu_{k+2}(s)$ for some fixed state $s$. We want to show the relation is true for $\nu_k(s) - \nu_{k+1}(s) \geq \mu_k(s) - \mu_{k+1}(s)$. Consider the following cases:

- Case I: For some state s, if $\nu_{k+1}(s) = \mu_{k+1}(s)$, then the relation is satisfied by the fact that $\nu_k(s) \geq \mu_k(s)$.
- Case II: Otherwise, the value function takes $\nu_{k+1}(s) = r_t(s,0) + \min_{p \in \mathcal{P}^0_{s,k+1}} \sum_{s' \in \mathcal{S}} p(s')\nu_{k+2}(s')$. Consider

$$\begin{aligned}
\nu_k(s) - \nu_{k+1}(s) &\overset{?}{\geq} \min_{p \in \mathcal{P}^0_{s,k}} \sum_{s' \in \mathcal{S}} p(s')\nu_{k+1}(s') - \min_{p \in \mathcal{P}^0_{s,k+1}} \sum_{s' \in \mathcal{S}} p(s')\nu_{k+2}(s') \\
&\overset{??}{\geq} \min_{p \in \mathcal{P}^0_{s,k}} \sum_{s' \in \mathcal{S}} p(s')(\nu_{k+1}(s') - \nu_{k+2}(s')) \\
&\overset{???}{\geq} \min_{p \in \mathcal{P}^0_{s,k}} \sum_{s' \in \mathcal{S}} p(s')(\mu_{k+1} - \mu_{k+2}) \\
&\overset{????}{\geq} \mu_{k+1}(s) - \mu_k(s).
\end{aligned} \tag{5.3.8}$$

The first inequality is just plugging the values of $\nu_k$, and $\nu_{k+1}$. The second inequality is due to the fact that at epoch $k+1$ the probability of moving to an absorbing state is higher than the probability of moving to an absorbing state at epoch $k$. The third inequality comes from the induction step.

Next, from the non-increasing property of $B_k(s)$, we have

$$B_{k+1}(s) - B_k(s) \geq 0.$$

Substituting values of $B_k \& B_{k+1}$, from Eq. ( 5.3.2), in the above equation, we get

$$\mu_{k+1}(s) - \lambda \min_{p \in \mathcal{P}^0_{s,k+1}} \sum_{s' \in \mathcal{S}} p(s')\mu_{k+2}(s') - \mu_k(s) - \lambda \min_{p \in \mathcal{P}^0_{s,k}} \sum_{s' \in \mathcal{S}} p(s')\mu_{k+1}(s') \geq 0.$$

Rearrange, we obtain

$$\mu_{k+1}(s) - \mu_k(s) \geq \lambda \min_{p \in \mathcal{P}^0_{s,k+1}} \sum_{s' \in \mathcal{S}} p(s')\mu_{k+2}(s') + \lambda \min_{p \in \mathcal{P}^0_{s,k}} \sum_{s' \in \mathcal{S}} p(s')\mu_{k+1}(s').$$

This shows the fourth inequality is satisfied. Hence, the induction step is satisfied.

□

In the next theorem, we present a threshold on this disutility value ($\tau$) for statin treatment to be initiated. The MDP version of Theorem (5.3.5) was proved in [52]. However, we prove the RMDP version.

**Theorem 5.3.5.** *Let the dis-utility associated with statin use be $\tau$, then there exists $\varepsilon_1, \varepsilon_2$ such that if $\tau \geq \varepsilon_1$, the best action is to wait for the next decision epoch, and if $\tau \leq \varepsilon_2$, the best action is to start medication.*

*Proof.* For $t \in [0, T)$,

$$\begin{aligned}
\nu_t^1(s) - \nu_t^0(s) &= E[R(s,1,j) + \gamma V_{t+1}(j)] - E[R(s,0,j) + \gamma V_{t+1}(j)] \\
&= \min_{p \in \mathcal{P}_{s,t}^1} \sum_{j=1}^{3} ((1-\tau) + \gamma V_{t+1}(j)) p_{ij} - \min_{p \in \mathcal{P}_{s,t}^0} \sum_{j=1}^{3} (1 + \gamma V_{t+1}(j)) p_{ij} \\
&= \min_{p \in \mathcal{P}_{s,t}^1} \sum_{j=1}^{3} (1 + \gamma V_{t+1}(j)) p_{ij} - \min_{p \in \mathcal{P}_{s,t}^0} \sum_{j=1}^{3} (1 + \gamma V_{t+1}(j)) p_{ij}(1) - \tau \\
&= \min_{p \in \mathcal{P}_{s,t}^1} \sum_{j=1}^{3} \gamma V_{t+1}(j) p_{ij} - \min_{p \in \mathcal{P}_{s,t}^0} \sum_{j=1}^{3} \gamma V_{t+1}(j) p_{ij} - \tau
\end{aligned} \tag{5.3.9}$$

The sign of $\nu_t^1(s) - \nu_t^0(s)$ depends on the dis-utility $\tau$. We can choose $\tau \geq \varepsilon_1$ such that $\nu_t^1(s) - \nu_t^0(s) \leq 0$, and the best action is to wait until the next decision epoch. Similarly, we can choose $\tau \leq \varepsilon_2$, such that $\nu_t^1(s) - \nu_t^0(s) \geq 0$, and the best action is to initiate treatment. □

## 5.4 Numerical Studies

We implement the model above for a patient with prediabetes, and the optimal policy with sensitivity is evaluated. At a given age $t$, there are three lipid ratio states: low ($1 < LR1 \leq 3.8$), medium ($3.8 < LR2 \leq 5.5$), and high ($5.5 < LR3 \leq 8.7$), a diabetes state and a terminating state $O$. The probabilities of moving from living states to a fatal CHD or stroke or non-CHD or non-stroke related deaths are calculated using the Framingham risk equations [2] and the NHANES data set [16]. Since the Framingham risk equations consider either diabetic or non-diabetic patients, we multiplied the risk of CHD or stroke by relative risk for prediabetes patients, which are 1.1, 1.06, respectively [45]. Moreover, we use cubic splines on the NHANES to estimate a patient systolic blood pressure as a function of age. The probability of moving to the diabetes state is calculated using the work of Levitzky et al. [58]. The probabilities of moving

to other living states are estimated using the results of Section 3.1. We let $P_{ns}$ and $P_s$ denote the transition probability matrices for different levels of lipid ratio for untreated and treated populations, respectively, and are estimated as shown in the third chapter.

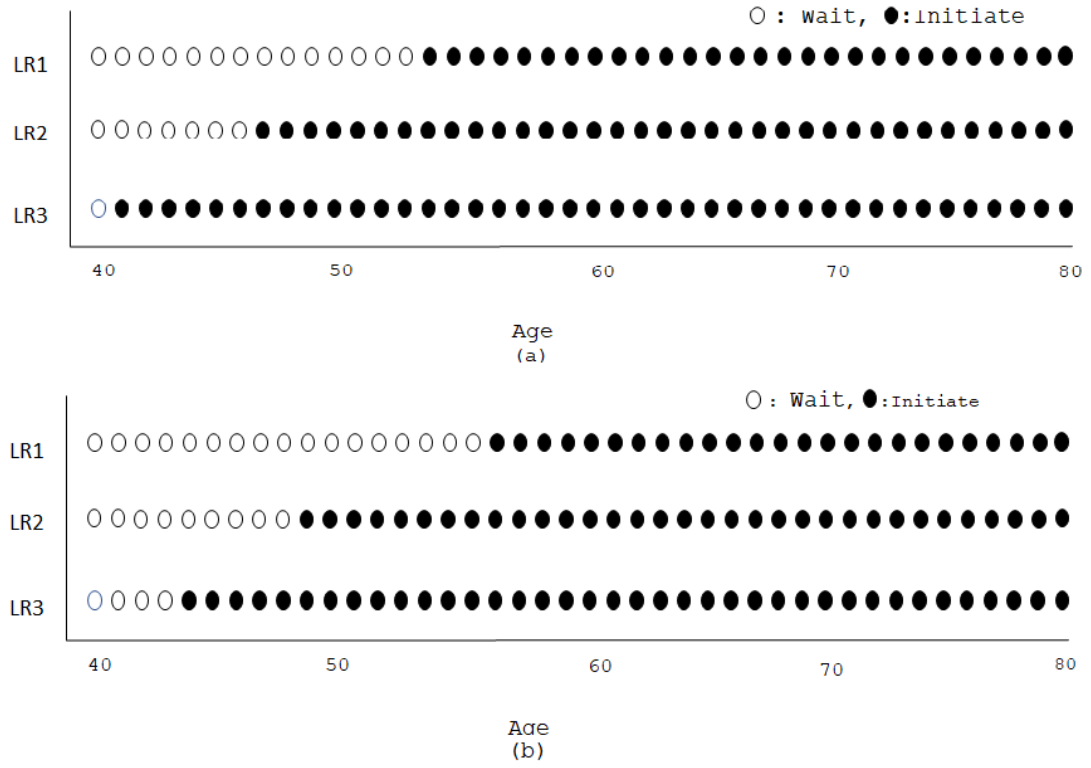


Figure 5.1: A robust MDP policy is presented with respect to the three lipid ratio levels. The disutility used in this implementation is 0.02.

Fig. 5.2 presents a worst-case scenario policy under pill dis-utility 0.02. Statin initiation is not recommended for low lipid ratio level but statin initiation is recommended at older age around 50 years old for medium lipid ratio. This also can be seen from sensitivity analysis in Chapter Four where the higher diabetes risk associated with statin negatively impacted statin recommendation. The Robust MDP implementation decreases the values of accrued QALYs. This justifies that the MDP policy is better than Robust MDP policy in terms of QALYs accrued. The maximum expected QALYs is presented

in Table 5.1.

Table 5.1: The maximum expected QALYs prior to the first terminal events for prediabetes patients are depicted. The rewards are 0.98 QALY for statin use and 1 QALY, otherwise. The relative risk for diabetes associated with using statin is 1.28.

| **Gender** | **Age** | **LR1** | **LR2** | **LR3** |
|---|---|---|---|---|
| female | | | | |
| | 40 | 17.736 | 17.717 | 17.677 |
| | 50 | 15.538 | 15.515 | 15.491 |
| | 60 | 13.090 | 13.076 | 13.041 |
| | 70 | 10.464 | 10.444 | 10.400 |
| | 80+ | 7.746 | 7.071 | 6.297 |
| Male | | | | |
| | 40 | 17.133 | 17.112 | 17.071 |
| | 50 | 15.120 | 15.079 | 15.056 |
| | 60 | 12.580 | 12.551 | 12.517 |
| | 70 | 9.631 | 9.617 | 9.577 |
| | 80+ | 5.412 | 4.842 | 4.183 |

Table (5.1) displays the maximum expected QALYs prior to the first terminal event using the robust MDP model. At the same scenario, a reward of 0.98 QALY for statin use and 1 QALY for non-statin use is assigned for one year period. The MDP approach accrued a higher value of QALYs than RMDP. For example, at age 40 with a low lipid ratio, the maximum expected QALYs accrued using the MDP approach is 21.440 QALYs, while the maximum expected QALYs accrued using robust MDP approach is 17.736 QALYs. The following table explains the remaining tasks in each chapter and the estimated time to finish them.

## 5.5 Sensitivity

A sensitivity analysis is implemented to study the effects of a small change in QALY values and the relative risk associated with statin use on the optimal policy. The pill disutility can take values more than 0.02 because of the other consequences of using statin like cancer [8], adverse effects on lever [11]. In this implementation, we raised the disutility associated with using statin from 0.02 to 0.04, and the relative risk of diabetes associated with using statin is raised from 1.28 to 1.5.

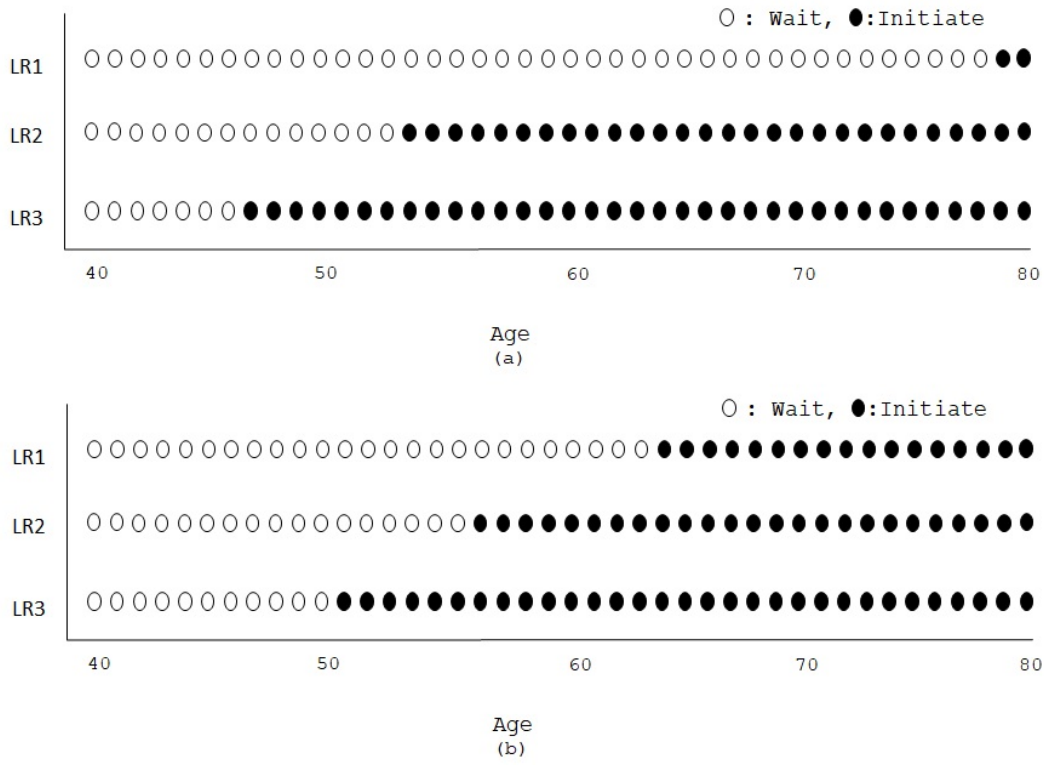


Figure 5.2: A robust MDP policy is presented with respect to the three lipid ratio levels. The disutility used in this implementation is 0.04, and the relative risk associated with statin use for diabetes is 1.5.

Table 5.2: The maximum expected QALYs prior to the first terminal events for prediabetes patients are depicted for 0.96 QALY for statin use and 1 QALY for no-statin. The relative risk associated with using statin is 1.5.

| **Gender** | **Age** | **LR1** | **LR2** | **LR3** |
|---|---|---|---|---|
| female | | | | |
| | 40 | 17.111 | 17.106 | 17.088 |
| | 50 | 14.990 | 14.980 | 14.953 |
| | 60 | 12.639 | 12.622 | 12.583 |
| | 70 | 10.176 | 10.152 | 10.103 |
| | 80+ | 7.746 | 7.071 | 6.297 |
| Male | | | | |
| | 40 | 16.532 | 16.512 | 16.471 |
| | 50 | 14.338 | 14.297 | 14.229 |
| | 60 | 11.856 | 11.800 | 11.763 |
| | 70 | 9.034 | 9.009 | 8.964 |
| | 80+ | 5.412 | 4.842 | 4.183 |

## 5.6 Conclusion

In this study, we developed a model that introduce a policy that consider the trade off between the benefits of CVD risk reductions and side effects of using statin for prediabetes population. Using RMDP model, we are able to find an optimal policy that is robust to a small changes in the model parameters, which are transition probabilities. The QALYS accrued using this model are lower because we are maximizing the rewards against the worst case scenario. In general, females are accruing higher number of QALYs then males. This could be because the average FBG and LDL for females are less than that for males, see table 5.1. Another reason for the differences in the accrued QALs is women have lower CVD risks than men [33]. The structural properties provided a sufficient condition for the existence of optimal control limit policy. Additionally, our rectangularity approach to formulate the uncertainty set allows to the worst

case scenario to occur simultaneously at the same time. This is considered one of the limitation of this approaches.

Future work is to add adherence factor to the model, and study its effects on the optimal policy. Race/Ethnicity need to be consider because the progression rates to the diabetes states and to the CHD or stroke risk state vary with respect to different race/ethnicity groups.

# Chapter 6

## Conclusion

Studying prediabetes disease has attracted researchers' attention recently. Prediabetes is a state with a high risk of developing diabetes, and complications of diabetes such as early nephropathy, small fiber neuropathy, early retinopathy and risk of macrovascular diseases [5], [23].

In the first chapter, we estimated the transition probabilities among different lipid ratio levels for prediabetes population using cross-sectional NHANES data sets. Additionally, using Bayesian approach, we estimated the risk of prediabetes as a function of age. Our result stats that as age increases the prediabtes risk is also increase. This result is used in modeling the risk of prediabetes in our MDP model as a function of age.

A multivariate quantile regression model is used to assess the characteristics of the association in different quantiles of the conditional distribution of the FBG. The dependent variables in our model are gender, Race/Ethnicity (Mexican, Other Hispanic, Non-Hispanic White, Non-Hispanic Black, Other Race - Including Multi-Racial), age, BMI, total cholesterol, and statin use (yes or no). Our model shows there is an association between statin use and FBG. In other words, statin users have higher FBG levels by approximately on average 50 mg/dL higher. However, OLS model is not able to detect this association because at lower quantile the effect of glucose is negligible. On average, the OLS shows statin users have higher glucose by 10 mg/dL. Gender effects on different FBG quantiles are investigated. At lower quantiles of FBG, the FBG for femals is less than FBG for males by 4 mg/dL. This dispersion reaches up to 8 mg/dL on average at higher quantiles. There is no association between FBG and TC at lower FBG

quantiles, but there is an association at higher FBG quantiles. The age and BMI effects on FBG are approximately similar.

An MDP model is used to introduce an optimal policy for statin treatment initiation. A sensitivity analysis is implemented on model parameters. Our findings are consistent with results in [68], and [43]. We found that whenever the risk of diabetes increases the statin initiation delays. More over, pill disutility also delays statin initiation treatments.

Our MDP model is sensitive to small changes in transition probabilities. We develop an RMDP model to tackle this issue. In the RMDP model, we found an optimal policy that stays optimal for small perturbations in the transition probability matrix. We also introduce a sufficient condition for the existence of the optimal control limit policy.

Our feature works is to study the uncertain adherence of the treatment effects on the optimal policy. The adherence has been study for diabetes population in the work of Mason et al. [61]. Moreover, We also interested in studying the effects of different statin drugs. It was reported that different statin drugs have different effects on cholesterol levels.

## References


[1] Oguzhan Alagoz, Heather Hsu, Andrew J Schaefer, and Mark S Roberts, *Markov decision processes: a tool for sequential decision making under uncertainty*, Medical Decision Making **30** (2010), no. 4, 474–483.

[2] Keaven M Anderson, Patricia M Odell, Peter WF Wilson, and William B Kannel, *Cardiovascular disease risk profiles*, American heart journal **121** (1991), no. 1, 293–298.

[3] Canadian Diabetes Association et al., *2008 clinical practice guidelines for the prevention and management of diabetes in canada* (2008).

[4] Beverley Balkau, Gang Hu, Qing Qiao, Jaakko Tuomilehto, Knut Borch-Johnsen, K Pyorala, DECODE Study Group, European Diabetes Epidemiology Group, and et. al., *Prediction of the risk of cardiovascular mortality using a score that includes glucose as a risk factor. the decode study*, Diabetologia **47** (2004), no. 12, 2118.

[5] Nidhi Bansal, *Prediabetes diagnosis and treatment: A review*, World journal of diabetes **6** (2015), no. 2, 296.

[6] RE Barlow, F Proschan, and LC Hunter, *Mathematical theory of reliability john wiley and sons inc*, New York **4** (1965), 927–929.

[7] Aharon Ben-Tal, Laurent El Ghaoui, and Arkadi Nemirovski, *Robust optimization*, Princeton university press, 2009.

[8] Denise M Boudreau, Onchee Yu, and Jeanene Johnson, *Statin use and cancer risk: a comprehensive review*, Expert opinion on drug safety **9** (2010), no. 4, 603–621.

[9] Julia Braun, Matthias Bopp, and David Faeh, *Blood glucose may be an alternative to cholesterol in cvd risk prediction charts*, Cardiovascular diabetology **12** (2013), no. 1, 1–6.

[10] Peter Bruce, Andrew Bruce, and Peter Gedeck, *Practical statistics for data scientists: 50+ essential concepts using r and python*, O'Reilly Media, 2020.

[11] Rossana M Calderon, Luigi X Cubeddu, Ronald B Goldberg, and Eugene R Schiff, *Statins in the treatment of dyslipidemia in the presence of elevated liver aminotransferase levels: a therapeutic dilemma*, Mayo clinic proceedings, 2010, pp. 349–356.

[12] Aleesa A Carter, Tara Gomes, Ximena Camacho, David N Juurlink, Baiju R Shah, and Muhammad M Mamdani, *Risk of incident diabetes among patients treated with statins: population based study*, Bmj **346** (2013).

[13] M Regina Castro, Gyorgy Simon, Stephen S Cha, Barbara P Yawn, L Joseph Melton, and Pedro J Caraballo, *Statin use, diabetes incidence and overall mortality in normoglycemic and impaired fasting glucose patients*, Journal of general internal medicine **31** (2016), no. 5, 502–508.

[14] Centers for Disease Control and Prevention (CDC)., *Healthy weight, nutrition, and physical activity*, 2021. `\url{https://www.cdc.gov/healthyweight/assessing/bmi/adult_bmi/index.html}`,.

[15] Centers for Disease Control and Prevention (CDC). National Center for Health Statistics (NCHS), *National center for health statistics*, 2017. `\url{https://www.cdc.gov/nchs/nvss/deaths.htm}`,.

[16] Centers for Disease Control and Prevention (CDC). National Center for Health Statistics (NCHS)., *National health and nutrition examination survey data. hyattsville*, 2017. `https://www.cdc.gov/nchs/nhanes/index.htm`.

[17] Jin-Biou Chang, Nain-Feng Chu, Jhu-Ting Syu, An-Tsz Hsieh, and Yi-Ren Hung, *Advanced glycation end products (ages) in relation to atherosclerotic lipid profiles in middle-aged and elderly diabetic patients*, Lipids in health and disease **10** (2011), no. 1, 228.

[18] Chee W Chia, Josephine M Egan, and Luigi Ferrucci, *Age-related changes in glucose metabolism, hyperglycemia, and cardiovascular risk*, Circulation research **123** (2018), no. 7, 886–904.

[19] Emerging Risk Factors Collaboration et al., *Diabetes mellitus, fasting blood glucose concentration, and risk of vascular disease: a collaborative meta-analysis of 102 prospective studies*, The Lancet **375** (2010), no. 9733, 2215–2222.

[20] Benjamin Lê Cook and Willard G Manning, *Measuring racial/ethnic disparities across the distribution of health care expenditures*, Health services research **44** (2009), no. 5p1, 1603–1621.

[21] Jill P Crandall, Kieren Mather, Swapnil N Rajpathak, Ronald B Goldberg, Karol Watson, Sandra Foo, Robert Ratner, Elizabeth Barrett-Connor, and Marinella Temprosa, *Statin use and risk of developing diabetes: results from the diabetes prevention program*, BMJ Open Diabetes Research and Care **5** (2017), no. 1, e000438.

[22] Lester R Curtin, Leyla K Mohadjer, Sylvia M Dohrmann, Deanna Kruszon-Moran, Lisa B Mirel, Margaret D Carroll, Rosemarie Hirsch, Vicki L Burt, and Clifford L Johnson, *National health and nutrition examination survey: sample design, 2007-2010.*, Vital and health statistics. Series 2, Data evaluation and methods research **160** (2013), 1–23.

[23] Ralph A DeFronzo and Muhammad Abdul-Ghani, *Assessment and treatment of cardiovascular risk in prediabetes: impaired glucose tolerance and impaired fasting glucose*, The American journal of cardiology **108** (2011), no. 3, 3B–24B.

[24] Brian T Denton, Murat Kurt, Nilay D Shah, Sandra C Bryant, and Steven A Smith, *Optimizing the start time of statin therapy for patients with diabetes*, Medical Decision Making **29** (2009), no. 3, 351–367.

[25] Caitlyn N Ellerbe, Mulugeta Gebregziabher, Jeffrey E Korte, Jill Mauldin, and Kelly J Hunt, *Quantifying the impact of gestational diabetes mellitus, maternal weight and race on birthweight via quantile regression*, PloS one **8** (2013), no. 6, e65017.

[26] Joseph C Engeda, Katelyn M Holliday, Shakia T Hardy, Sujatro Chakladar, Dan-Yu Lin, Gregory A Talavera, Barbara V Howard, Martha L Daviglus, Amber Pirzada, and Pamela J et al. Schreiner, *Transitions from ideal to intermediate cholesterol levels may vary by cholesterol metric*, Scientific reports **8** (2018), no. 1, 1–8.

[27] Mohammad Faheem, Saqib Qureshi, Jabar Ali, Hameed Hameed, Zahoor Zahoor, Farhat Abbas, Adnan Mahmood Gul, and Mohammad Hafizullah, *Does bmi affect cholesterol, sugar, and blood pressure in general population?*, Journal of Ayub Medical College Abbottabad **22** (2010), no. 4, 74–77.

[28] Jing Fang, Zefeng Zhang, Carma Ayala, Angela M Thompson-Paul, and Fleetwood Loustalot, *Cardiovascular health among non-hispanic asian americans: Nhanes, 2011–2016*, Journal of the American Heart Association **8** (2019), no. 13, e011324.

[29] Eugene A Feinberg and Adam Shwartz, *Handbook of markov decision processes: methods and applications*, Vol. 40, Springer Science & Business Media, 2012.

[30] Francisco J Félix-Redondo, Maria Grau, and Daniel Fernández-Bergés, *Cholesterol and cardiovascular disease in the elderly. facts and gaps*, Aging and disease **4** (2013), no. 3, 154.

[31] Centers for Disease Control, Prevention, et al., *National diabetes statistics report, 2020*, Atlanta, GA: Centers for Disease Control and Prevention, US Department of Health and Human Services (2020), 12–15.

[32] Dilys J Freeman, John Norrie, Naveed Sattar, R Dermot G Neely, Stuart M Cobbe, Ian Ford, Christopher Isles, A Ross Lorimer, Peter W Macfarlane, James H McKillop, and et. al., *Pravastatin and the development of diabetes mellitus: evidence for a protective treatment effect in the west of scotland coronary prevention study*, Circulation **103** (2001), no. 3, 357–362.

[33] Zujie Gao, Zengsheng Chen, Anqiang Sun, and Xiaoyan Deng, *Gender differences in cardiovascular disease*, Medicine in Novel Technology and Devices **4** (2019), 100025.

[34] Alan J Garber, Yehuda Handelsman, Daniel Einhorn, Donald A Bergman, Zachary T Bloomgarden, Vivian Fonseca, W Timothy Garvey, James R Gavin, George Grunberger, and Edward S et. al. Horton, *Diagnosis and management of prediabetes in the continuum of hyperglycemia—when do the risks of diabetes begin? a consensus statement from the american college of endocrinology and the american association of clinical endocrinologists*, Endocrine practice **14** (2008), no. 7, 933–946.

[35] Alan M Garber, Richard A Olshen, Heping Zhang, and ES Venkatraman, *Predicting high-risk cholesterol levels*, International Statistical Review/Revue Internationale de Statistique (1994), 203–228.

[36] Andrew Gelman, Walter R Gilks, and Gareth O Roberts, *Weak convergence and optimal scaling of random walk metropolis algorithms*, The annals of applied probability **7** (1997), no. 1, 110–120.

[37] Stuart Geman and Donald Geman, *Stochastic relaxation, gibbs distributions, and the bayesian restoration of images*, IEEE Transactions on pattern analysis and machine intelligence **6** (1984), 721–741.

[38] MENARD M GERTLER, STANLEY MARION GARN, and EF Bland, *Age, serum cholesterol and coronary artery disease*, Circulation **2** (1950), no. 4, 517–522.

[39] Laurent El Ghaoui, Maksim Oks, and Francois Oustry, *Worst-case value-at-risk and robust portfolio optimization: A conic programming approach*, Operations research **51** (2003), no. 4, 543–556.

[40] David C Grabowski, Darius N Lakdawalla, Dana P Goldman, Michael Eber, Larry Z Liu, Tamer Abdelgawad, Andreas Kuznik, Michael E Chernew, and Tomas Philipson, *The large social value resulting from use of statins warrants steps to improve adherence and broaden treatment*, Health affairs **31** (2012), no. 10, 2276–2285.

[41] Qiuping Gu, *Prescription cholesterol lowering medication use in adults aged 40 and over: United states, 2003-2012*, US Department of Health and Human Services, Centers for Disease Control and ..., 2014.

[42] Xin Guo, Li Shen, Jing Dou, Yaogai Lv, Anning Zhang, Fanchao Shi, Zhiqiang Xue, Yaqin Yu, Lina Jin, and Yan Yao, *Associations of fasting blood glucose with influencing factors in northeast china: a quantile regression analysis*, International journal of environmental research and public health **14** (2017), no. 11, 1368.

[43] David J Heller, Pamela G Coxson, Joanne Penko, Mark J Pletcher, Lee Goldman, Michelle C Odden, Dhruv S Kazi, and Kirsten Bibbins-Domingo, *Evaluating the impact and cost-effectiveness of statin use guidelines for primary prevention of coronary heart disease and stroke*, Circulation **136** (2017), no. 12, 1087–1098.

[44] Yuehua Hu, Junting Liu, Hongyan Yao, Dapeng Yin, Jian Zhang, and Guoshuang Feng, *Different percentile regression of blood glucose among adolescents aged 12-20—united states, 1999- 2018*, China CDC Weekly **3** (2021), no. 3, 46–49.

[45] Yuli Huang, Xiaoyan Cai, Weiyi Mai, Meijun Li, and Yunzhao Hu, *Association between prediabetes and risk of cardiovascular disease and all cause mortality: systematic review and meta-analysis*, Bmj **355** (2016).

[46] Garud N Iyengar, *Robust dynamic programming*, Mathematics of Operations Research **30** (2005), no. 2, 257–280.

[47] Anuradha V Khadilkar, Nikhil Lohiya, Sejal Mistry, Shashi Chiplonkar, Vaman Khadilkar, Neha Kajale, Veena Ekbote, Smruti Vispute, Rubina Mandlik, and Hemchand et al. Prasad, *Random blood glucose concentrations and their association with body mass index in indian school children*, Indian journal of endocrinology and metabolism **23** (2019), no. 5, 529.

[48] Mee Kyoung Kim, Kyungdo Han, Yong-Moon Park, Hyuk-Sang Kwon, Gunseog Kang, Kun-Ho Yoon, and Seung-Hwan Lee, *Associations of variability in blood pressure, glucose and cholesterol concentrations, and body mass index with mortality and cardiovascular outcomes in the general population*, Circulation **138** (2018), no. 23, 2627–2637.

[49] Min Jung Ko, Ae Jeong Jo, Yun Jung Kim, Shin Hee Kang, Songhee Cho, Sang-Ho Jo, Cheol-Young Park, Sung-Cheol Yun, Woo Je Lee, and Duk-Woo Park, *Time-and dose-dependent association of statin use with risk of clinically relevant new-onset diabetes mellitus in primary prevention: A nationwide observational cohort study*, Journal of the American Heart Association **8** (2019), no. 8, e011320.

[50] Roger Koenker, *Quantile regression, volume 38 of*, Econometric Society Monographs (2005).

[51] Sameer Kumar, Muer Yang, and Paul Bekx, *Why do many diabetic patients not take the a1c test as recommended?*, IISE Transactions on Healthcare Systems Engineering **7** (2017), no. 3, 145–154.

[52] ———, *Why do many diabetic patients not take the a1c test as recommended?*, IISE Transactions on Healthcare Systems Engineering **7** (2017), no. 3, 145–154.

[53] Murat Kurt, Brian T Denton, Andrew J Schaefer, Nilay D Shah, and Steven A Smith, *The structure of optimal statin initiation policies for patients with type 2 diabetes*, IIE Transactions on Healthcare Systems Engineering **1** (2011), no. 1, 49–65.

[54] Fritz Lang, Thea Von Harbou, Alfred Abel, Gustav Fröhlich, Brigitte Helm, Gottfried Huppertz, and Universität der Künste Berlin, *Metropolis*, Lorrimer Pub., 1973.

[55] Yann Le Tallec, *Robust, risk-sensitive, and data-driven control of markov decision processes*, Ph.D. Thesis, 2007.

[56] Meng Lee, Jeffrey L Saver, Keun-Sik Hong, Sarah Song, Kuo-Hsuan Chang, and Bruce Ovbiagele, *Effect of pre-diabetes on future risk of stroke: meta-analysis*, Bmj **344** (2012).

[57] ———, *Effect of pre-diabetes on future risk of stroke: meta-analysis*, Bmj **344** (2012).

[58] Yamini S Levitzky, Michael J Pencina, Ralph B D'Agostino, James B Meigs, Joanne M Murabito, Ramachandran S Vasan, and Caroline S Fox, *Impact of impaired fasting glucose on cardiovascular disease: the framingham heart study*, Journal of the American College of Cardiology **51** (2008), no. 3, 264–270.

[59] Xiaoping Liu, Rui Liu, Xing-Ming Zhao, and Luonan Chen, *Detecting early-warning signals of type 1 diabetes and its leading biomolecular networks by dynamical network biomarkers*, BMC medical genomics **6** (2013), no. 2, 1–10.

[60] JM Lobo, BT Denton, JR Wilson, Nilay D Shah, and SA Smith, *Using claims data linked with electronic health records to monitor and improve adherence to medication*, IISE Transactions on Healthcare Systems Engineering **7** (2017), no. 4, 194–214.

[61] Jennifer E Mason, Darin A England, Brian T Denton, Steven A Smith, Murat Kurt, and Nilay D Shah, *Optimizing statin treatment decisions for diabetes patients in the presence of uncertain future adherence*, Medical Decision Making **32** (2012), no. 1, 154–166.

[62] Andy Menke, Sarah Casagrande, Linda Geiss, and Catherine C Cowie, *Prevalence of and trends in diabetes among adults in the united states, 1988-2012*, Jama **314** (2015), no. 10, 1021–1029.

[63] ———, *Prevalence of and trends in diabetes among adults in the united states, 1988-2012*, Jama **314** (2015), no. 10, 1021–1029.

[64] Dariush Mozaffarian, Emelia J Benjamin, Alan S Go, Donna K Arnett, Michael J Blaha, Mary Cushman, Sarah De Ferranti, Jean-Pierre Després, Heather J Fullerton, Virginia J Howard, and et. al., *Executive summary: heart disease and stroke statistics—2015 update: a report from the american heart association*, Circulation **131** (2015), no. 4, 434–441.

[65] Mahsa Nazari, Saeed Hashemi Nazari, Farid Zayeri, Mehrzad Gholampour Dehaki, and Alireza Akbarzadeh Baghban, *Estimating transition probability of different states of type 2 diabetes and its associated factors using markov model*, Primary care diabetes **12** (2018), no. 3, 245–253.

[66] Christopher Nielson, Theodore Lange, and Nicholas Hadjokas, *Blood glucose and coronary artery disease in nondiabetic patients*, Diabetes care **29** (2006), no. 5, 998–1001.

[67] Arnab Nilim and Laurent Ghaoui, *Robustness in markov decision problems with uncertain transition matrices*, Advances in neural information processing systems **16** (2003), 839–846.

[68] Ankur Pandya, Stephen Sy, Sylvia Cho, Milton C Weinstein, and Thomas A Gaziano, *Cost-effectiveness of 10-year risk thresholds for initiation of statin therapy for primary prevention of cardiovascular disease*, Jama **314** (2015), no. 2, 142–150.

[69] Ben Pelzer and Rob Eisinga, *Bayesian estimation of transition probabilities from repeated cross sections*, Statistica Neerlandica **56** (2002), no. 1, 23–33.

[70] Ben Pelzer, Rob Eisinga, and Philip Hans Franses, *Inferring transition probabilities from repeated cross sections*, Political Analysis (2002), 113–133.

[71] Xinyue Peng, Thatcher W Root, and Christos T Maravelias, *Optimization-based process synthesis under seasonal and daily variability: Application to concentrating solar power*, AIChE Journal **65** (2019), no. 7, e16458.

[72] Kathleen Pickrell, *Miller-keane encyclopedia and dictionary of medicine, nursing, and allied health*, Hospitals & Health Networks **77** (2003), no. 8.

[73] Michael Pignone, Stephanie Earnshaw, Jeffrey A Tice, and Mark J Pletcher, *Aspirin, statins, or both drugs for the primary prevention of coronary heart disease events in men: a cost–utility analysis*, Annals of internal medicine **144** (2006), no. 5, 326–336.

[74] András Prékopa, *Stochastic programming*, Vol. 324, Springer Science & Business Media, 2013.

[75] Martin L Puterman, *Markov decision processes*, Handbooks in operations research and management science **2** (1990), 331–434.

[76] Paul M Ridker, Aruna Pradhan, Jean G MacFadyen, Peter Libby, and Robert J Glynn, *Cardiovascular benefits and diabetes risks of statin therapy in primary prevention: an analysis from the jupiter trial*, The Lancet **380** (2012), no. 9841, 565–571.

[77] CP Robert, *Monte carlo methods. hoboken*, New Jersey: Wiley Online Library, 2004.

[78] Naveed A Sattar, Henry Ginsberg, Kausik Ray, M John Chapman, Marcello Arca, Maurizio Averna, D John Betteridge, Deepak Bhatnagar, Elena Bilianou, Rafael Carmena, and et. al., *The use of statins in people at risk of developing diabetes mellitus: evidence and guidance for clinical practice*, Atherosclerosis Supplements **15** (2014), no. 1, 1–15.

[79] Nazir Savji, Caron B Rockman, Adam H Skolnick, Yu Guo, Mark A Adelman, Thomas Riles, and Jeffrey S Berger, *Association between advanced age and vascular disease in different arterial territories: a population database of over 3.6 million subjects*, Journal of the American College of Cardiology **61** (2013), no. 16, 1736–1743.

[80] Kivity Shaye, Tirosh Amir, Segev Shlomo, and Sidi Yechezkel, *Fasting glucose levels within the high normal range predict cardiovascular outcome*, American heart journal **164** (2012), no. 1, 111–116.

[81] Gitanjali M Singh, Goodarz Danaei, Pamela M Pelizzari, John K Lin, Melanie J Cowan, Gretchen A Stevens, Farshad Farzadfar, Young-Ho Khang, Yuan Lu, and Leanne M et al. Riley, *The age associations of blood pressure, cholesterol, and glucose: analysis of health examination surveys from international populations*, Circulation **125** (2012), no. 18, 2204–2211.

[82] Allen L Soyster, *Convex programming with set-inclusive constraints and applications to inexact linear programming*, Operations research **21** (1973), no. 5, 1154–1157.

[83] Neil J Stone, Jennifer G Robinson, Alice H Lichtenstein, David C Goff Jr, Donald M Lloyd-Jones, Sidney C Smith Jr, Conrad Blum, and J Sanford Schwartz, *Treatment of blood cholesterol to reduce atherosclerotic cardiovascular disease risk in adults: synopsis of the 2013 american college of cardiology/american heart association cholesterol guideline*, Annals of internal medicine **160** (2014), no. 5, 339–343.

[84] Takehiro Sugiyama, Yusuke Tsugawa, Chi-Hong Tseng, Yasuki Kobayashi, and Martin F Shapiro, *Different time trends of caloric and fat intake between statin users and nonusers among us adults: gluttony in the time of statins?*, JAMA internal medicine **174** (2014), no. 7, 1038–1045.

[85] Rishi Sukhija, Sastry Prayaga, Mohammad Marashdeh, Zoran Bursac, Priyanka Kakar, Darpan Bansal, Rajesh Sachdeva, Sree Hari Kesan, and Jawahar L Mehta, *Effect of statins on fasting plasma glucose in diabetic and nondiabetic patients*, Journal of Investigative Medicine **57** (2009), no. 3, 495–499.

[86] Adam G Tabák, Christian Herder, Wolfgang Rathmann, Eric J Brunner, and Mika Kivimäki, *Prediabetes: a high-risk state for developing diabetes*, Lancet **379** (2012), no. 9833, 2279.

[87] Konstantinos T Tsaousis, *Blood glucose and cholesterol concentrations in a mediterranean rural population of andros island, greece*, International journal of preventive medicine **5** (2014), no. 11, 1464.

[88] Phillip Tuso, *Prediabetes and lifestyle modification: time to prevent a preventable disease*, The Permanente Journal **18** (2014), no. 3, 88.

[89] Belgin Unal, Simon Capewell, and Julia Alison Critchley, *Coronary heart disease policy models: a systematic review*, BMC public health **6** (2006), no. 1, 1–10.

[90] Jan Van de Kassteele, RT Hoogenveen, PM Engelfriet, PHM Van Baal, and HC Boshuizen, *Estimating net transition probabilities from cross-sectional data with application to risk factors in chronic disease modeling*, Statistics in medicine **31** (2012), no. 6, 533–543.

[91] Ardo Van Den Hout, *Multi-state survival models for interval-censored data*, CRC Press, 2016.

[92] Salim S Virani, Alvaro Alonso, Emelia J Benjamin, Marcio S Bittencourt, Clifton W Callaway, April P Carson, Alanna M Chamberlain, Alexander R Chang, Susan Cheng, and et al. Delling Francesca N, *Heart disease and stroke statistics—2020 update: a report from the american heart association*, Circulation **141** (2020), no. 9, e139–e596.

[93] Kang-Ling Wang, Chia-Jen Liu, Tze-Fan Chao, Su-Jung Chen, Cheng-Hsueh Wu, Chi-Ming Huang, Chun-Chin Chang, Ko-Fan Wang, Tzeng-Ji Chen, Shing-Jong Lin, and et al., *Risk of new-onset diabetes mellitus versus reduction in cardiovascular events with statin therapy*, The American journal of cardiology **113** (2014), no. 4, 631–636.

[94] Shoumeng Yan, Meng Li, Xiaoyu Ma, Shan Jiang, Mengzi Sun, Changcong Wang, Yingan Pan, Chong Sun, Yan Yao, Lina Jin, and et al., *Association of multiple mineral and vitamin b group intake with blood glucose using quantile regression analysis: Nhanes 2007–2014*, Food & nutrition research **63** (2019).

[95] Dukyong Yoon, Seung Soo Sheen, Sukhyang Lee, Yong Jun Choi, Rae Woong Park, and Hong-Seok Lim, *Statins and risk for new-onset diabetes mellitus: A real-world cohort study using a clinical research database*, Medicine **95** (2016), no. 46.

[96] Salim Yusuf, Jackie Bosch, Gilles Dagenais, Jun Zhu, Denis Xavier, Lisheng Liu, Prem Pais, Patricio López-Jaramillo, Lawrence A Leiter, Antonio Dans, and et. al.,

*Cholesterol lowering in intermediate-risk persons without cardiovascular disease*, New England Journal of Medicine **374** (2016), no. 21, 2021–2031.

[97] Yuanhui Zhang et al., *Robust optimal control for medical treatment decisions.* (2014).

[98] Xiaohui Zhuo, Ping Zhang, Henry S Kahn, and Edward W Gregg, *Cost-effectiveness of alternative thresholds of the fasting plasma glucose test to identify the target population for type 2 diabetes prevention in adults aged* $\geq$ 45 *years*, Diabetes Care **36** (2013), no. 12, 3992–3998.