ESSAYS ON THE MANAGEMENT OF APPOINTMENTS FOR CHRONIC CONDITIONS

by

Zlatana Dobrilova Nenova

B.S., Guilford College, 2011

M.A., University of Pittsburgh, 2013

Submitted to the Graduate Faculty of

the Department of Business Analytics and Operations in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2017

UNIVERSITY OF PITTSBURGH

KATZ GRADUATE SCHOOL OF BUSINESS

This dissertation was presented

by

Zlatana Dobrilova Nenova

It was defended on

June 5th 2017

and approved by

Jennifer Shang, Department of Business Analytics and Operations

Jerrold May, Department of Business Analytics and Operations

John Hotchkiss, Strategic Analyst, Chief Business Office, VA Pittsburgh Healthcare System

Luis Vargas, Department of Business Analytics and Operations

Lisa Maillart, Department of Industrial Engineering

Dissertation Director: Jennifer Shang, Department of Business Analytics and Operations

ESSAYS ON THE MANAGEMENT OF APPOINTMENTS FOR CHRONIC CONDITIONS

Zlatana Dobrilova Nenova, PhD

University of Pittsburgh, 2017

Treating chronic conditions is a fairly complex task, which requires well-timed appointments to control one's disease progression. In my dissertation I would like to optimize the monitoring strategies and better predict the demand-for-care of patients with chronic kidney disease (CKD). To do that I design a chronic disease monitoring framework which consists of forecasting, survival analysis and Markov Decision Process (MDP) models. First, I propose a forecasting model which quantifies the impact of CKD-related doctor's appointments on patient's disease progression. The model accounts for patient's comorbidities, vital signs, and important laboratory values. Second, I propose a survival analysis model, which estimates the expected life days of a patient given his or her current health status. Finally, I use the information gained from the first two models to parametrize and solve the MDP, which can suggest monitoring strategies and predict medium-term demand for CKD-patient-care in a clinic. In addition to the chronic disease monitoring framework, I examine CKD patient characteristics associated with a higher resource utilization.

Keywords: optimal appointment allocation, chronic disease, Markov Decision Process.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

I would like to thank everybody who has helped me in the process of completing my dissertation and PhD. First, I would like to thank my adviser Jennifer Shang for her personal and professional support. I will always be grateful for her help and mentorship, which I needed to graduate in four years. I want to thank Jerry May for his guidance and infectious excitement when discussing big data, which is what sparked my desire to get a PhD in the first place. I also express my gratitude to Dr. John Hotchkiss for his help in ensuring the clinical validity of all models included in my dissertation. Finally, I would like to acknowledge the other members on my committee – Luis Vargas and Lisa Maillart – for helping me develop different parts of my dissertation, challenging me and always being available for advice.

I would not have completed this journey without my family: my parents Dobril and Snejana, my husband Robert, and my parents-in-law Bob and Elaine. My parents were always there to provide me with emotional support and encouragement, my husband and his parents kept me motivated without letting me forget to live life and enjoy myself along the way. I love you all, and I dedicate this dissertation to you.

1.0 INTRODUCTION

Chronic conditions are diseases, which once developed, cannot be cured. If left untreated, they can significantly reduce patients' quality of life. Still, if properly managed chronic disease progression can be slowed down and even halted [19]. Unfortunately, studies suggest that current chronic care is often suboptimal [6, 17, 40, 45, 89, 92, 114].

In this dissertation, I present a framework of models, which aims at optimizing the treatment of patients with different chronic illnesses. The framework is applied to a single condition – chronic kidney disease (CKD). CKD is selected due to its high prevalence rate, cost of care [98], complexities [86] and the decrease in the number of new medical graduates with a nephrology subspecialty [90], which makes finding the balance between the supply and demand for CKD care particularly hard to achieve.

The models presented in this dissertation attempt to improve current treatment practices by optimizing the timing of follow-up appointments with specialists who have a direct impact on patients' disease progression. The appointment frequency alternatives we consider are based on guidelines suggested by the National Institute for Health and Care Excellence. We believe that appointment timing is important because visits should be neither too frequent nor too scarce. The first option is undesirable because it could overwhelm the patient and prevent the medical provider from seeing individuals in a greater need for care. The second alternative should be avoided because it could inhibit early progression detection, which could lead to patients' hospitalizations and death.

Models for personalized care have gained popularity in recent years. Still, most of them have focused on detecting [8, 23] rather than monitoring diseases [38]. Furthermore, current CKD appointment optimization models focus only on determining the optimal timing of dialysis initiation [61] or the placement of a fistula - the vascular access needed for dialysis

[107]. Thus, no papers have attempted to estimate the best time for a follow-up appointment for CKD patients prior to dialysis initiation. Helm [38] optimizes appointment decisions for glaucoma patients, but does not use a Markov Decision Process (MDP) model.

Given the gap in the MDP literature I selected a dissertation topic which aims at designing a general disease monitoring framework for optimizing the timing of a follow-up appointment for a chronically ill patient. To do that I developed two parameter estimation models, which feed information into an optimization (MDP) model.

The first parameter estimation model applies a case-based reasoning (CBR) method to forecast patient's disease progression given a set of health descriptors and a desired appointment frequency. In a manner similar to other CBR models, ours contains a database, which incorporates information on millions of past patient (solution) cases. When a new patient (target) arrives, he is compared to all cases in the database, and similar solution cases are retrieved and used to forecast his disease progression. The model hinges on the assumption that patients with a similar disease history will experience compatible disease progression.

The forecasting model addresses a very important aspect of any CBR, namely, how many nearest neighbors should be retrieved when target and solution cases are matched and if multiple cases are selected, whether they should be equally weighted or weighted based on their closeness to the target. These questions are answered through a two-step decision model [82].

The model contributes to the current medical CBR literature with its novel stability measure, which determines health descriptors' importance weights. The measure ensures that when looking for nearest neighbors, all selected solution cases are especially close matches to the target across features, where the target case has high variability. We created this measure based on a recommendation by Dr. Hotchkiss.

Finally, the model is the first CBR which estimates patients' disease progression based on his health status and a desired appointment frequency. This allows us to estimate the transition probability matrix of an MDP model, which determines when it is optimal for a patient to be seen next.

The second parameter estimation model applies a Cox proportional hazards model with two competing risks (death and dialysis). The model is designed to estimate the terminal

reward of the MDP. In this dissertation, we have included a more complex version of the survival analysis model. In addition to the variables needed to parametrize the MDP, the model incorporates past appointment frequency information. It also introduces a segmented relationship across all continuous predictors. The breakpoint regression results suggest that if a patient with CKD needs on average over seventeen kidney disease related appointments in a year then his doctors should begin a palliative care preparation. This is the first data-driven rule, which is based on readily available data, designed to facilitate targeted palliative care introduction.

As noted above the two parameter estimation models are used to design an optimization model, which determines when a CKD patient should be seen next. This is the first MDP model, which can design sensible CKD monitoring policies, evaluate current practices and estimate the expected aggregate demand for care at a CKD clinic. The model confirms that older and sicker patients need to follow stricter monitoring strategies. It also highlights the need for considering patients' possible future disease progression when making appointment decisions.

This is the first MDP model used to produce aggregate demand for care estimates. Thus, management can ensure that a clinic has enough medical professionals to meet patients' demand for care. Recruiting doctors with a nephrology subspecialty is especially problematic. That is why accurate and timely demand predictions are important because they provide the administrative staff with the extra time needed to hire additional nephrologists. Additionally, these estimates can signal if a clinic will have a lower than expected demand and can therefore accept new patients.

The dissertation also contains a model which uses a novel method to highlight the relationship between CKD patients' health status and their demand for care. The model suggests that the most significant drivers of hospitalizations and outpatient appointments are comorbidities, more specifically diabetes and heart failure. The results highlight the importance of developing better referral policies so that patients with such coexisting conditions can get timely access to cardiologists and endocrinologists. While more intense monitoring strategies might result in an increase in the number of outpatient appointments, they will hopefully translate into fewer hospitalizations.

We would like to thank the U.S. Department of Veterans Affairs for the data they have provided, which was used in all models. This work was made possible because of the continuing partnership between the Pittsburgh Veterans Engineering Resource Center (VERC) and the Katz Graduate School of Business.

The remainder of the dissertation is organized as follows. First, we discuss the CBR forecasting model in Chapter 2. Chapter 3 focuses on our Cox proportional hazards model. Our optimization model is presented in Chapter 4. The clustering analysis results are examined in Chapter 5. Finally, we conclude and present additional details in Appendices A through C.

2.0 REPRESENTATION AND RETRIEVAL OF MEDICAL TIME SERIES DATA FOR CASE-BASED FORECASTING

2.1 INTRODUCTION

Accurate forecasting models are important for optimizing decisions in a multifaceted environment. Humans have limited ability to process information [76]. Therefore, using data driven models could be beneficial in settings where the information a person has to receive, process and use to make decisions is too complex. Standard multivariate forecasting methods use past data to provide estimates for future items' trajectories [71]. Still, there is a need for models that provide accurate forecasts and incorporate variables which allow a decision choice evaluation.

A case-based reasoning model (CBR), which can include multivariate time series and decision choice data, is effective in settings where decisions are experience-based. CBRs [101] mimic the experience-driven decision making process of humans, and are widely used in health science research [10, 32, 84], due to their similarities to the daily procedures followed by medical professionals.

Additionally, CBRs are appealing because they can use intermittent, multivariate, and possibly correlated time series data. Research areas in need of more precise forecasting models built on such data include disease modeling [3], high-frequency financial tick data analysis [91], and natural disaster predictions [68]. While the development of such models could be challenging, it is important to ensure that they are complete, tractable, and easy to follow, which would make them attractive to both researchers and practitioners.

In this paper, we propose a CBR procedure for modeling time-series data, which could be used to accurately predict the disease progression of patients with chronic conditions. The

model is useful because it allows us to evaluate different treatment policies, which is necessary when optimizing one's regular monitoring strategy. To demonstrate the procedure, we built a detailed model for forecasting the progression of patients with chronic kidney disease (CKD). We chose CKD as the chronic condition to showcase our model, because of its prevalence, cost [98] and complexities [86].

As with any CBR model, we built a database of cases, which are snippets of patients' histories. That database is called a *case base*. The cases are used to predict the progression of a new CKD patient, referred to as the *target*. The target is compared to all cases in the case base. One or more of the most similar cases are retrieved from the case base. The retrieved cases are used to forecast the future trajectory of the target individual.

What separates our model from previously developed CBRs and improves its prediction capabilities are its novel method of representing key time series features, and its customized retrieval procedure, which improves the information extraction from the time-series-based CBRs. As noted earlier, CBRs mimic a human reasoning process. Humans examine patterns in time series data to draw inferences about its future values, so a CBR should also incorporate features that summarize such patterns. Based on discussions with our medical collaborator, we designed a process that uses standard statistical tools to facilitate information extraction. For example, our collaborator suggested that, when dealing with patients with chronic conditions, doctors should be particularly careful when observing gradual or sudden changes in health measures. To account for that, we represent and select similar cases based on the stability of the target's health status descriptors, as well as his closeness to possible solution cases. The stability measures ensure that features with greater variability are more important when deciding on possible solution cases.

Furthermore, we provide a technique to answer a question that arises in the retrieval stage of any CBR problem: how many nearest-neighbor solution cases should be selected from the case base in order to create a prediction for the target case? We propose a case-specific, customized retrieval approach. We demonstrate the approach on our CKD dataset, and produce numerical confirmation that the customized method produces results superior to those of any non-customized technique.

As noted in [10], when it comes to matching techniques "the nearest neighbor retrieval algorithm is still widely applied in medical CBR systems," meaning that, typically, a prediction for the target is produced using only its single nearest neighbor. While this method makes case-based reasoning simple because the target is predicted to be exactly like the nearest neighbor [4], we consider strategies, which use information from up to the four nearest neighbors because any "increase in the ratio of controls to cases lead to gain in power until a ratio of 4 to 1; after that point, gains in power usually become too small to be worthwhile" [29, 58]. Furthermore, we examine whether we should weigh solution cases equally or depending on their dissimilarity with the target patient.

Therefore, we develop a model that identifies the most desirable number of solution cases from a fixed set of strategic alternatives. The predictors we use to determine the preferable approach for a target are associated with its characteristics, which in our CBR model include patient's (1) comorbidities, (2) oscillations in kidney filtration capabilities, (3) lab values' extremes and (4) lab values' dissimilarities across the target and nearest neighbor cases. The model works in two steps. First, we determine if a target will be better forecasted using (1) one nearest neighbor, or (2) equally or (3) non-equally weighted multiple nearest neighbors. If a target is better classified using multiple equally (non-equally) weighted nearest neighbors, we choose among using the two, three or four equally (non-equally) weighted solution cases closest to the target. The weights used in the non-equal approach are based on the total uniformized dissimilarity between the target and each of its solution cases. While the two-step model was developed with chronic kidney disease patients in mind, it is also applicable to other CBR environments that have similar time series and/or categorical features.

Our work contributes to the case-based reasoning literature in the following ways. First, we present an approach for describing medical time series data, which (1) accommodates its nonequispaced collection nature, and (2) teases out important features' stability information. Second, we propose a multi-stage, data mining-type model for personalizing the sample size of solution cases used to forecast the progression of a new patient, and show empirical evidence of its superiority.

The remainder of the paper is organized as follows. The relevant case-based reasoning, analytical model, and chronic kidney disease literatures are discussed in section 2.2. We

describe our case-based reasoning features, retrieval and adaptation procedures in §2.3. In section 2.4, we examine the chronic kidney disease data used to showcase the model and describe the CBR case settings. In §2.5 we discuss the process used to select our two-step nearest neighbor model. We discuss our model and its numerical results in sections 2.6 and 2.7. Concluding remarks and potential directions for future research are addressed in section 2.8.

2.2 LITERATURE REVIEW

In this section, we discuss three streams of relevant literature. First, we focus on the research done in case-based reasoning. Second, we address related analytical models. Third, we discuss the relevant chronic kidney disease literature.

2.2.1 Case-Based Reasoning Literature

Case-based reasoning can be traced back to the work by Schank and Abelson [101]. CBR models consist of five-steps:(1) feature selection, (2) case retrieval, (3) adaptation, (4) validation and (5) update [5]. The retrieval step in any CBR process has a great impact on its performance [56].

Not surprisingly, there exists a non-trivial literature on different components of the case retrieval process, including the selection of neighbor sample sizes. Even though the dominant matching technique used in a health science CBR is the nearest neighbor retrieval algorithm [10, 32, 84], researchers should use more than one nearest neighbor "to improve the generalization properties of the retrieval, reduce sensitivity to noise, and obtain more accurate results (by interpolation)" [47].

Jarmulak [47] and Ahn [1] used a genetic algorithm (GA) to select an optimal population-based nearest neighbor sample size policy. Jarmulak [47] used a standard genetic algorithm with a cross-validation strategy to avoid overfitting. Ahn [1] used a GA-optimized k-nearest neighbor (k-NN) algorithm with feature weighting. Additionally, Lee and Park [62] proposed

a linear-integer mathematical programming (MP) method for optimizing the number of case neighbors. Still, the MP procedure in Lee and Park [62] could be computationally intensive, and depends on an adjustment factor $p \in [0, 0.5]$, which is not clearly defined [1].

Rather than using a population-based policy, we develop a fast and accurate method for customized sample size selection. Customization is important, because there is no reason to believe that cases are uniformly distributed in the feature space. We expect that the preferable number of solution cases for a target that is near a large number of other cases will be different when compared against a target case that is relatively isolated. When considering unusual target cases, which are significantly separated from nearly all prior cases, using a one- rather than an n-nearest-neighbors strategy might be more appropriate, because basing a prediction on multiple cases may introduce undesirable variance and inaccuracy into the forecast.

2.2.2 Relevant Analytics Models

Snippets from the case base are retrieved based on their proximity to the target case. Furthermore, cases have multiple features, so combining feature distances to determine cases' overall similarity is another important case-retrieval component. The most common approach in the literature is to use weighted sums of feature distances [118]. More recent papers discuss optimizing weights through genetic algorithms [47, 1]. However, all three methods [118, 47, 1] use population-based approaches, meaning that the same set of weights is used for all target-to-case-base comparisons.

Based on discussions with our medical collaborator, we believe that a target-specific weight mechanism is more appropriate for our medical decision-making problem. That is, we suggest that feature weights should be patient dependent. For example, a patient who has a history of stable lab results but whose blood pressure (BP) has been increasing, should be matched to patients in a way that depends primarily on the BP trend. Focusing mostly on it is sensible, because unstable results require more attention than do consistent measures. We describe our stability-dependent similarity determination procedure in §2.4. We believe this is the first solution retrieval process that takes into account the consistency of lab values.

To speed up the CBR retrieval process, researchers have applied various indexing methods, such as X-trees [11], R-trees [35], LCS-trees [96] and TV-trees [109]. Others implement cluster-based methods [9]. We use a filtering approach [77] in which patients are first filtered based on a set of static features, and then compared to the target patient.

As noted earlier, we select the number of nearest neighbors using a two-step analytical model [82].Knowing the structure of our classification tree prior to building it, we can justify the control over the number and type of splits at each of its levels. At the first step, the model separates patients that are to be predicted based on their nearest neighbor from individuals that should be predicted using multiple, unequally and equally weighted solution cases. At the second step, it separates the individuals categorized as multiple-case predictees into the three multiple nearest neighbor classes considered. Thus, we allow for a triple node-split at both levels of the decision tree. We consider five different induction algorithms, which are discussed in 2.5, to build our model. Four of the algorithms selected use decision trees. They include oblique decision trees [79]; CART [13]; support vector networks [18]; and C5 [93].

2.2.3 Chronic Kidney Disease

To develop and evaluate our model, we use data on patients with chronic kidney disease (CKD). CKD is described by the gradual decline in a patient's kidneys' filtration function. It has five stages, where the fifth one is also referred to as end-stage renal disease (ESRD). The stages are specified by the patient's glomerular filtration rate (GFR), because it is the best measure of kidney function [108]. Unfortunately, GFR lab tests are fairly expensive, so doctors typically estimate glomerular filtration rate (eGFR) to define the patient's disease stage. The estimation procedure is based on the patient's race (African-American or non-African-American), gender, age, and serum creatinine [65].

We selected CKD because of its prevalence and its treatment complexities. According to the newly issued kidney disease fact sheet, around ten percent of U.S. adults may have chronic kidney disease. In 2011 alone, over one hundred thousand individuals started treatment for end stage renal disease [26].

In an effort to improve the treatment of the disease, researchers have developed a variety of CKD models. Some focus on determining the optimal timing of dialysis initiation [60, 61] . The models consider demographic characteristics, comorbidity conditions, discretized disease markers, and clinical flags to describe an individual. The authors utilize two of the seven continuous disease markers we selected: eGFR and Albumin. They discretize eGFR into five groups and Albumin into two groups, although they neither appear to provide conceptual support, nor a citation to the medical literature, for their discretization policy.

Another recent CKD article examines when a patient should have vascular access surgery, in preparation for an expected dialysis initiation [105]. The authors focus on patients' eGFR measures, and use a linear regression model to estimate CKD progression parameters. We believe that a linear assumption may be problematic, because "in contrast to the traditional paradigm of steady GFR progression over time, many patients with CKD have a nonlinear GFR trajectory or a prolonged period of nonprogression" [67]. In addition to patients' eGFR, there are other lab values that should be considered when predicting disease progression.

The first step of our prediction model falls into the category of CKD trajectory estimation models. We allow for a non-linear disease progression in the laboratory values and vital signs selected [67, 87]. We do not focus on specific time intervals of our patients' lives, such as last years of life [78] or last two years before the initiation of long-term dialysis [87]. Our underlying trajectory estimation approach is similar to a Bayesian analysis spline regression [67]. We do not use a visual graphical analysis exploratory approach [78] or a semi-parametric mixture model [87].

2.3 CBR SETTINGS

In this section we will discuss the data transformations and summary information needed to design the case-based reasoning dataset. Following a detailed description of the case base (2.3.1 and 2.3.2) we will explain how features are compared and used to retrieve cases "similar" to the targets we would like to forecast (2.3.3). Finally, we will define our forecast building adaptation procedure (2.3.4).

We selected a CBR model because it allows us to incorporate complex patient information, including static and dynamic categorical data, and irregularly sampled time series data. The flexibility of the method's structure provides an opportunity to develop a model that is more detailed and complete than its predecessors [60, 105]. It enables us to evaluate how alternative control strategies impact a target's future development.

CBR models are not decision models in and of themselves. Still, they can be very effective in parametrizing standard decision models, such as MDPs, with state spaces describing patients' health, and with action spaces addressing possible appointment frequency alternatives. CBRs are particularly useful when estimating transition probabilities, because of their ability to incorporate MDP state and action space information.

2.3.1 Case Structure

Cases are defined over a time window of length T. The first $T-1$ time periods represent the historical records. The last month is used for forecasting. The length of the historical records' time window should be recommended by a domain expert. We use a sliding window technique to maximize the number of created cases. If a record has M $\geq$ T months of data, we can create $M-T+1$ unique snippets out of it. Each case contains a time stamp and an instance ID, which allows us to track the cases created using the moving window technique.

A case can contain irregularly sampled time series data. To address the difficulty caused by the unequally spaced time intervals, we smooth our data [3]. The goal of the smoothing process is to produce an entry for time series' periods 1 through M for each record in our dataset.

When a case contains multiple time series, we need to determine whether smoothing them separately or simultaneously is more appropriate. To answer that question, we can randomly select and withhold p percent of a patient's time series data and use it as a test set. The remaining data can be used to fit the simultaneously and separately smoothed models. Once the models are fitted, they can be used to estimate the withheld patient's data. The estimates can be compared to the actual values, and standard accuracy measures can be used to evaluate the two models [43]. In the CKD model used to showcase our CBR

tool, we concluded that smoothing time series one at a time produced more accurate results.

We use four models to independently smooth our patients' time series: thin plate regression, with and without shrinkage [21], and cubic spline regression, with and without shrinkage [95], unless a given model does not converge, in which case its settings are not considered. We set $\gamma = 1.4$ [52] in the Generalized Cross-Validation score of the fitted model. To select a model, we first optimize the sample size dependent maximum degrees of freedom (k) allowed across all four models by examining the residual pattern removed by k [119]. We can also use a QQ plot to examine models' fit through their residuals [119]. While residuals' normality is not an assumption required by a general additive model, QQ plots can tell us how well the model fits the data. Rather than plotting the residuals against standard normal quantiles, we check if the sample residuals are consistent with a population assumption of normality, using the Shapiro-Wilk test because a manual examination of thousands of QQ plots is not practical. While we do not require models' residuals to be normality distributed (p-value > 0.05), we select the model with the highest Shapiro-Wilk p-value because it should have better information extraction properties. A model is selected based only on the normality of its residuals if none of the k values are deemed appropriate.

We observed that an extended time interval between data entry points could cause unrealistically high or low smoothed values when the maximum degrees of freedom are greater than nine. To spot such abnormalities, we check whether the smoothed time series have impossible values (e.g., negative laboratory values), or if the maxima (minima) of the smoothed time series are at least 10% higher (lower) than the maxima (minima) of the original data. If one or more anomalies are observed, we reduce the value of k. If the reduction does not help we can also increase the value of γ, in an effort to smooth the spline regression even further.

The instance ID and time stamp of each case allow us to locate the smooth and raw time series data associated with a case. While the smoothed values are directly incorporated into the case definition, the raw data have to be summarized. The relevant summary statistics are the raw times series extrema (minima and maxima) and standard deviations. The raw data information complements the smoothed trajectory data, and addresses the variability in the time series.

Finally, we record the number of peaks, the number of troughs, and the two-sided Wald-Wolfowitz run test's p-value [75] for the smoothed time series in the case. The peak and though information helps us differentiate between stable and fluctuating time series that display similar percentage changes between time periods 1 and $T-1$. The null hypothesis of the non-parametric test states that each element in the transformed time series sequence is independently drawn from the same distribution. A small p-value could be associated with a steady increase or decrease in the smoothed time series [80] (Table 1).

Table 1: A Generic Example of a Time Series Case Summary

			Value
Identifier		Record ID	1
		Begin	01/01/00
		End	01/01/01
Historical Data	Smoothed Data	1	1
		...	...
		$T-1$	12
		peaks	0
		troughs	0
		WW p-value	0.002465
	Raw Data	min 1:$T-1$	0.7
		max 1:$T-1$	13
		sd 1:$T-1$	2.3
Forecasting Information	Smoothed Data	T	14
	Raw Data	min 2:T	1
		max 2:T	16
		sd 2:T	2.5

Additionally, a case may contain information in the form of categorical data (Table 2). To record such information, we save the case state in periods $T-1$ and T. The case's state in time period $T-1$(T) is used for case retrieval (target adaptation).

Table 2: A Generic Example of a Categorical Data Case Summary

		Value
Identifier	Record ID	1
	Begin	01/01/00
	End	01/01/01
Historical Data	Category 1 ($T-1$)	0
	...	...
Forecasting Information	Category 1 (T)	1
	...	...

2.3.2 Features

All features should be selected with the help of a domain expert. Cases are matched based on their categorical and time series data. Categorical data are used to stratify patients prior to predicting their future time series' trajectory behavior. Times series data are compared based on their raw and smoothed information. If we believe that time series fluctuation and extrema are important, we match cases based on the raw data summaries only. Conversely, if we are interested in the overall trajectory of a time series, we compare cases based on the smoothed data only. If we are interested in the time series fluctuation and trajectory, we should compare cases based on their smoothed and raw data.

2.3.3 Retrieval

First, solution cases are filtered based on the categorical data states. We select only case snippets whose categorical data statuses match those of the target patient. Once we select

the relevant solutions, we compare them to the target case. We scale the smoothed time series so that they have a mean of zero and a variance of one, because we are interested in the trajectory of the series, not in the level of the values themselves. The degree of separation of the points along the y-axis is not material. Once scaled, the time series are compared using a dynamic time warping (DTW) distance measure [51], because it allows for a nonlinear alignment of time series that could be similar to one another but locally out of phase. The peak, trough, and descriptive statistics measures are not scaled, because magnitudes of their raw value are important when selecting solution cases. They are compared using a Euclidean distance measure, because it is one of the most commonly used methods for comparing continuous variables [22]. We purposefully selected these dissimilarity measures because they are not disease specific.

We examine the target patient's stability across each time series used to describe a case. While the method we use mimics our medical collaborator's importance feature assignment, we believe this idea is generalizable for any chronic disease CBR. The stability measure highlights features with high variability, gradual changes and fluctuations observed in the not too distant past. This is done because such changes should be closely monitored.

Before we begin the stability measure calculation, we fit a fifth degree Taylor polynomial to all smoothed time series across all selected solution cases, to determine the positions of their first inflection points. For all time series' features and across all solution cases, we obtain (1) the raw data's standard deviation, which embodies features' variability, (2) the smoothed time series' percentage change between months one and twelve and (3) the smoothed time series' linear regression slope, which are used as features' proxies for gradual change, and (4) the absolute difference in the smoothed time series' pre- and post-inflection point linear regression slopes, which highlights fluctuations in the not too distant past. Due to their magnitude differences, the four values are first uniformized and then summed for each feature. The feature-specific total uniformized stability measures are separated into ten equally sized bins, and the bin cut-off values are used to classify the features' stability of our target patient. Stability scores are thus between 1 and 10, where the greater the score, the more relatively unstable the target's feature is, as compared with the same feature in all relevant solution cases.

Once the stability ranks (s_j) of all j =1, ..., J target case features, and the distances between the descriptors of the target and solution cases are obtained, we begin our selection procedure. First, we find a minimal positive dissimilarity distance between the target and all solution cases for all features (d_j). The minimum value helps us gauge the proximity of the closest non-identical solution features. Using the minimal dissimilarity distance and the stability ranks, we perform a window search around each target feature, which expands W times (w = 1, ..., W). We record the cases whose features meet our threshold requirement of $d_j(1+0.5w\frac{11-s_j}{10})$, for all w. To forecast a target using N nearest neighbors we set aside the first N solution cases that meet $J, ..., \lfloor J/2+1 \rfloor$ out of the J threshold requirements, where N is specified before the CBR model runs. Thus, we get $(J - \lfloor J/2+1 \rfloor+1)$ possible solution case groups. If there do not exist N cases that satisfy at least J, ..., $\lfloor J/2+1 \rfloor$ of the threshold requirements at step W, we save as many cases as do exist, and proceed to the next step of the retrieval procedure. For example, if cases are matched based on eight features, i.e. J = 8, there will be a total of four possible solution case groups, where the N solutions in each group satisfy 8, 7, 6, and 5 of the 8 threshold requirements.

Using the rule above, we identify cases that meet more than half of the feature thresholds. The solution case groups are compared across each feature, in order to select the most preferable solution group. The comparison is done in two steps. First, we check to see if two or more of the feature groups are significantly different from each other, using a Kruskal-Wallis (K-W) test. We use a non-parametric test because the normality assumption needed to perform ANOVA analysis cannot be guaranteed. If the K-W test has a p-value of no more than 0.01, we use the Nemenyi multiple comparison test [81] with $\alpha = 0.05$. We conclude that a group is preferable for a specific feature if its dissimilarity values are significantly smaller than the majority of its alternatives. Such a feature group is given a rank of two, and the others are given a rank of one. Once all feature-specific threshold groups have been compared and ranked, the ranks are summed up. The group with the greatest sum is selected. If two or more groups have the same maximum sum, the group that satisfies the greater number of threshold requirements is selected.

2.3.4 Adaptation

The information for time period T from all selected solution cases is used to forecast the target case trajectory. To obtain the non-equally weighted nearest neighbor model results, we weight the information from each case, based on its proximity to the target patient, where proximity is determined by its uniformized feature dissimilarities. The weight $w_n = \frac{\sum_j d_{j,n}}{\sum_j \sum_n d_{j,n}}$ is neighbor specific, and incorporates the information from all features, where n = 1, ..., N is the number of neighbors and j = 1, ..., J is the feature number. To get the equally weighted nearest neighbor forecasts, we set $w_n = 1/n$ for all n. Combining dissimilarity measures has been used in gene analysis [55, 66], where distances from gene expression data and prior biologically meaningful patterns are averaged to produce better results.

The weights are then used to obtain the composite percentage change of the solution cases. The composite percentage change ($P_{T,j} = \sum_n w_n p_{T,n,j}$) is a weighted average of the percentage changes observed across all N solution cases between months $T-1$ and T. To forecast the smoothed time series one month into the future, $P_{T,j}$ is (1) increased by one and (2) multiplied by the smoothed target value of feature j in month $T-1$. To forecast the raw data descriptive statistics, we use the information from all N solution cases, and update the raw value for feature j, summary measure k, for period [2:T] to $Y_{T,j,k} = \sum_n w_n Y^n_{T,j,k}$. Finally, we update the categorical data characteristics of the target for period [2:T], if at least one of the selected solution cases experiences such a change between periods $T-1$ and T.

To evaluate our model, we use a standard forecast-value-to-actual-value comparison. We hold out the actual smoothed time series data between period T and T + F, where F is the forecasting time window normally used by a domain expert. We forecast all time series values between periods T and T + F, using the CBR model and our targets' historical record information between periods 1 and $T-1$. Finally, we compare the forecasted and actual values using a standard forecast accuracy measure, such as a mean absolute percentage error (MAPE).

2.4 DATA

In this section, we discuss the data used to showcase the model described in section 2.3. Following our medical collaborator's suggestion, each case is defined over a time window of length T = 13. The time series data contains information on patients' laboratory values and vital signs. The categorical data included in each case contains patient's comorbidity and dialysis status. Additional information on patients' appointment frequencies is provided.

2.4.1 Laboratory Values and Vital Signs

Cases include four laboratory values (estimated glomerular filtration rate, albumin, phosphate, and potassium) and three vital signs (weight, systolic blood pressure, and diastolic blood pressure). Estimated glomerular filtration rate (eGFR) is used to monitor CKD progression [44] and estimate the flow rate of filtrated fluids that goes through one's kidneys. Albumin helps individuals maintain growth and repair tissues, and its decrease could indicate inadequate nutrition, proteinuria, or inflammation. Chronic kidney disease could affect the regulation of one's potassium and phosphorus levels. Excess potassium could cause hyperkalemia, and result in a hospitalization due to heart arrhythmia and sudden cardiac death. Similarly, high levels of phosphorus can have an adverse effect on a patient's heart, blood vessels and lungs [69]. Blood pressure is important for determining patient's risk of cardiovascular complications. Kidney disease could lead to water retention, which can be detected by measuring patient's weight.

Using the smoothing process described in §2.3, we transformed our intermittent time series to equispaced data. We set k = {10; 20; 30; 40}. We do not consider k > 40 in order to prevent over-fitting, which is especially important for data with high measurement errors. In cases where the abnormalities discussed in §2.3.1 were observed we set k = {8; 6; 3}. We limited k to 5 or 3 for situations in which the abnormality issues were not resolved. If we observed the same difficulty again, we restricted k = 3 and set $\gamma = \{4; 6\}$ to further smooth the spline regression line.

We compare an individual based on eight time-series-based features: (1) smoothed eGFR time series; (2) smoothed eGFR time series' peak and trough information; patients' (3) albumin; (4) potassium; (5) phosphate; (6) weight; (7) systolic blood pressure; and (8) diastolic blood pressure raw data summary statistics (Table 3). We chose the smoothed eGFR data because doctors are interested in the overall trajectory of the patient, and not in the fluctuations, which are sometimes caused by measurement errors. The error is primarily a result of the timing of the lab test. For example, if the test is run after dialysis, the eGFR results may be inflated, because the patient's blood has just been filtered through a dialysis machine. We selected the raw data information for the other features because doctors are concerned when they observe low albumin, high potassium, high phosphate, high blood pressure, and fluctuating or increasing weight.

Table 3: CBR Time Series Features

Time Series Features	Data Type
eGFR time series	(Periods 1: T – 1) Smooth data
eGFR peaks and troughs	Smooth data
Phosphate	Raw data
Potassium	Raw data
Albumin	Raw data
SBP	Raw data
DBP	Raw data
Weight	Raw data

2.4.2 Comorbidities and Dialysis Status

We monitor five comorbidities: type II diabetes, heart failure, peripheral vascular disease (PVD), cerebrovascular disease (CVD), and cirrhosis. We consider diabetes because it was the primary cause in 44% of all new kidney failure cases in 2011 [26]. We track heart

failures, peripheral and cerebrovascular disease because CKD is recognized as an independent factor for heart disease outcomes [99]. Finally, we screen for cirrhosis because it could cause inaccurate kidney filtration estimates [106].

This information is based on inpatient admission, inpatient discharge, and outpatient appointment ICD9 and ICD10 codes. To categorize all patients, we (1) filter all ICD9 and ICD10 codes into groups based on the five comorbidities and dialysis status, and (2) use the date of the first recording of a complication as an estimate of its date of onset.

We use this categorical data to filter out our CBR solutions (Table 4). First, we match target patients to solution cases with identical comorbidities and dialysis status, because disease complexities are expected to have an effect on patients' trajectories. Second, all selected cases are separated into three equally sized groups, based on the percentage change in their smoothed eGFR values between months one and twelve. That is done in order to find the cut-off values of the groups, which can be labeled as (1) increasing/stable, (2) with progressive, or (3) with accelerated/catastrophic loss of eGFR [87], and filter the solution cases that are in the same eGFR percentage change category as the target case.

Table 4: CBR Categorical Data Features

Categorical Variables	Values
Diabetes	$\{0, 1\}$
Heart Failure	$\{0, 1\}$
PVD/CVD	$\{0, 1\}$
Cirrhosis	$\{0, 1\}$
Dialysis	$\{0, 1\}$

2.4.3 Appointment Information

We have recorded patient appointments with medical providers, who can have a direct impact on patients' kidney disease progression: (1) General Internal Medicine, (2) Cardiology, (3) Diabetes, (4) Hematology, (5) Pulmonary/Chest, (6) Renal/Nephrology (except dialysis), (7) Primary Care/Medicine, and (8) Hepatology. That information is not used for determining case similarity in this paper. It is used to estimate the impact of an action choice in an MDP transition probability, because different appointment patterns generate different future patient trajectories.

Patient appointment information is used to determine the number of days between the last appointment before, and the first appointment after, the case's twelfth month. The number of days is used to partition patients into four groups: those with appointments less than one and a half months apart, those with appointments between one and a half and three months apart, those with appointments between three and six months apart, and those with appointments more than six months apart. The four alternatives represent the action states in our MDP, and are based on a policy suggested by the National Institute for Health and Care Excellence, which recommends that patients in Stage 1 and Stage 2 should be seen once every twelve months, patients in Stage 3 - once every six months, patients in Stage 4 - once every three months, and patients in Stage 5 - once every one and a half months [72].

2.5 TWO STEP MODEL SETTINGS

We developed a model, which selected the optimal number of nearest neighbors based on how the nearest neighbor policies performed on various target patients. Depending on our model building data set's size, we could either use all of it or select a representative sample based on criteria identified by a domain expert.

For our model we used a representative sample based on the magnitude and change of targets' eGFR. We built the model on cases which were of primary interest to us and had an average smoothed eGFR $\in$ [10, 60], because such patients had CKD Stages 3 - 5 and

with a few exceptions were not on dialysis. We calculated the percentage change in the smoothed eGFR values within the time window across which the target cases were defined. The changes were ordered, and a stratified sample of 2000 patients was obtained. Finally, we included the patients with the highest and lowest percentage change in eGFR. Once the 2002 patients were selected, we used the same stratification technique to select 500 patients out of the them. The 500 patients were used as a validation set. The remaining 1502 were used as a training set.

Each of the target patients was forecasted one (F = 1) month into the future. This time horizon allowed for a less subjective accuracy measures selection. Additionally, the iterative nature of our forecasting approach requires good short-term predictions, because inaccurate month 1 forecasts could have a multiplicative effect on the inaccuracies in the forecasts that would follow.

We recorded seven different 1-month forecasts for each target case using the procedure described in §2.3. The policies used N equal to (1) 1, (2) 2 equally weighted,(3) 2 unequally weighted, (4) 3 equally weighted, (5) 3 unequally weighted, (6) 4 equally weighted, or (7) 4 unequally weighted nearest neighbors to forecast all target cases - §2.3.4.

Table 5: Nearest Neighbor's Optimal Frequencies

	1-NN	2e-NN	2n-NN	3e-NN	3n-NN	4e-NN	4n-NN	Grand Total
Training Set	512	151	133	124	145	179	258	1502
Test Set	167	45	54	38	55	38	103	500

Once each target's forecasts were recorded, they were compared to the actual smoothed time series of the target patient. We calculated the time series' absolute percentage error, rounded off after the tenth decimal point, across the seven forecasted months when using (1) 1 (1-NN), (2) 2 equally weighted (2e-NN), (3) 2 non-equally weighted (2n-NN), (4) 3 equally weighted (3e-NN), (5) 3 non-equally weighted (3n-NN), (6) 4 equally weighted (4e-NN), or (7) 4 non-equally weighted (4n-NN) nearest neighbors. Since there were multiple time-series-based features, we focused on the one which was of primary importance for the

model user. In this model forecasting accuracy was evaluated in terms of patients smoothed eGFR values, because it is the best measure of kidney function [108].

We did not consider strategies that included information from more than four nearest neighbors because any "increase in the ratio of controls to cases lead to gain in power until a ratio of 4 to 1; after that point, gains in power usually become too small to be worthwhile" [29, 58]. That conclusion might explain the popularity of matching ratios between 1:1 and 1:4 [85].

We deemed a forecasting approach as preferable if it had the smallest primary feature absolute percentage error, which is a measurement commonly used when examining the accuracy of different forecasting methods. We recorded the preferable forecasting approaches for each target case (Table 5).

We used two hundred and forty-five predictors to determine the preferable forecasting technique for each patient. All predictors were based on the closeness of the target and solution cases when one through four nearest neighbors were used, and previously discussed case-base features, which fell into four categories: (1) smoothed and (2) raw laboratory values, (3) appointment history, and (4) comorbidities. The predictors based on smoothed lab results included (1) the change and (2) the squared change between months 1 and 12, (3) the number of peaks, (4) the number of troughs, (5) the average, (6) the squared average, (7) the standard deviation, (8) the variance, (9) the slope, and (10) the squared slope of the smoothed eGFR. The predictors based on raw lab results included (1) the minima, (2) the squared minima, (3) the maxima, (4) the squared maxima, (5) the standard deviations, and (6) the variances of patient's raw eGFR, albumin, phosphate, potassium, systolic and diastolic blood pressure, and weight.

The appointment-based predictors included (1) the number and (2) the squared number of relevant outpatient appointments attended, (3) the number and (4) the squared number of days the target patient was hospitalized in the twelve-month time window across which the target was defined.

The closeness measures included (1) the smoothed eGFR time series', (2) the smoothed eGFR time series' peaks and troughs and (3) the albumin, (4) the potassium, (5) the phosphate, (6) the weight, (7) the systolic and (8) the diastolic blood pressure averaged dissimilarities and squared averaged dissimilarities when a patient is forecasted using one, two, three, and four nearest neighbors for all features used in the retrieval procedure described in §2.3.3. Additionally, we used the difference and squared difference between the dissimilarities across the eight features listed above when the dissimilarity of the nearest neighbors was compared to the averaged dissimilarity of the two/three/four nearest neighbors, when the averaged dissimilarity of the two nearest neighbors was compared to the averaged dissimilarity of the three/four nearest neighbors, and finally when the averaged dissimilarity of the three nearest neighbors was compared to the one of the four nearest neighbors. The difference metrics were standardized, so their estimated coefficients would measure their relative importance.

The five binary comorbidity characteristics included whether the target patient had been diagnosed with (1) diabetes, (2) heart failure, (3) PVD/CVD, or (4) cirrhosis, or is on (5) dialysis. Based on the five binary characteristics, patients fell into one of thirty-two classes ($2^5 = 32$). Twenty-five of the thirty-two classes had nonzero frequencies in the training set and were therefore used as possible predictors. The classes varied from patients without comorbidities and not on dialysis (Class 1) to individuals on dialysis with all four comorbidities (Class 32). The last comorbidity based variable considered was the number of comorbidities each patient had.

The methods we selected to establish a relationship between the preferable forecasting method for a target case and its descriptors in both steps of the model included logistic regression, discriminant analysis, C5.0, CART, and Support Vector Machines (SVM). We used a CART model with a Gini impurity measure for categorical targets, a discriminant analysis model with step-wise parameter selection, and logistic regression models with either forward or backward conditional parameter selection. We considered SVM models with two and three-degree polynomial and RBF kernel types. To account for possible over-fitting, we used regularization parameters C = 1, ..., 10 for each kernel type. We fitted five C5.0 models with different minimum numbers of records per child branch settings (M = 2, 5, 10, 15, 20).

The model we selected had two steps. First, we classified patients as having a preference

for the 1-NN, equally (2e-, 3e-, 4e-NN) and non-equally (2n-, 3n-, 4n-NN) weighted multiple-NN forecasting approaches. The second step required two different models, one model for cases that were predicted as having equally weighted nearest neighbor preferences and a second model for patients that were classified as having non-equally weighted nearest neighbor preferences. Each of the second step models had three classification alternatives, namely use (1) two, (2) three or (3) four nearest neighbors.

2.6 TWO STEP MODEL DESCRIPTION

We developed a model that selects an optimal nearest neighbor strategy for a CBR target case, based on its characteristics. The model uses a two-step process to select among seven alternatives. We imposed that structure on our model because it (1) produced good classification results and (2) was intuitively plausible. We did not use a single-step model because it proved suboptimal, especially for those nearest neighbor groups that are underrepresented in the training and test sets. Similarly, using methods with more than two classification steps resulted in a large decrease in the correctly predicted cases, especially across targets that were best forecasted using only the single nearest neighbor.

The classification tree structure provides some insights about the characteristics associated with the seven nearest neighbor forecasting approaches. Complete details of the model are provided in Appendix A. In its first step, the model classifies each target case into one of three categories: best forecast using (1) a single, (2) multiple equally weighted, or (3) multiple unequally weighted nearest neighbors.

The first model achieves separation by using four types of variables: (1) dissimilarities -, (2) comorbidities-, (3) filtering-features-, and (4) raw-summary-statistics-based (Table 25a-25b). The dissimilarity predictors contain information on all matching features used in the CBR retrieval process. Their inclusion confirms the significant contribution of all time series features used to gauge the closeness of target and solution cases. The comorbidity group contrasts patients with no comorbidities with those who have at least one coexisting chronic condition. The filtering feature is a function of the change in patient's smoothed eGFR over

the previous twelve months. A function of this predictor is used to select relevant solutions cases, and is discussed in section 2.3.3. The raw summary statistics highlight the maximum phosphate reading for a patient in the previous year. An elevated phosphate level may cause an adverse cardiac event.

The relative probability of being best forecasted using only one nearest neighbor is greater if the standardized averaged dissimilarity difference in the eGFR peak metric between using one and two nearest neighbors increases, and all other predictors are held constant. That is also the case when the standardized and squared albumin (systolic blood pressure) difference between two(three) and four nearest neighbors increases. Furthermore, patients with large differences in their standardized average potassium dissimilarity readings when using two and three nearest neighbors have greater odds of being better forecasted using only one nearest neighbor. The opposite holds if the squared eGFR change (diastolic blood pressure dissimilarity) increases and all other predictors are held constant.

Therefore, we conclude that target cases with a preference for the 1-NN forecasting policy have a smaller number of similar solution cases in their (1) eGFR peak and trough, (2) albumin, (3) potassium, and (4) systolic blood pressure feature spaces. Target patients with a preference for multiple nearest neighbor forecasting policies experience large changes in their eGFR and diastolic blood pressure dissimilarity readings in the previous twelve months. The patients with a preference for non-equally weighted multiple nearest neighbors often have co-existing chronic conditions and less extreme phosphate readings, and experience significantly greater standardized squared difference in their solution cases' albumin dissimilarity values when one (three) and three (four) nearest neighbors are compared. Target cases with a preference for equally weighted solution case forecasts have significantly higher standardized diastolic blood pressure (weight) dissimilarity differences (differences and squared differences) when comparing one and three (four) nearest neighbors.

The portion of the model that separates cases categorized as being forecasted using non-equally weighted multiple nearest neighbors, primarily involves dissimilarity-based predictors. The other independent variables are raw-summary-statistics and hospitalization based (Table 27). The model suggests that patients for whom only two nearest neighbors should be used are those with larger standardized averaged phosphate (eGFR smoothed time

series) dissimilarity differences when using three (four) rather than two nearest neighbors. Patients with a preference for the four nearest neighbor option typically have a very small standardized difference between their eGFR peak and trough values when using two and four nearest neighbors. The probability of having more accurate forecasts with the 3-NN policy significantly increases if there exists a greater standardized difference between solution cases' smoothed eGFR time series values when using three and four nearest neighbors.

The portion of the model that separates cases that are best forecasted using equally weighted multiple nearest neighbor cases involves eight dissimilarity and two smoothed time series based variables (Table 29). According to the model, patients with a preference for the two nearest neighbor forecasting policy are more overweight, have a more gradual decline in their eGFR, as represented by the slope of the smoothed eGFR time series, and have a relatively smaller (larger) standardized difference (squared difference) between systolic blood pressure (phosphate) dissimilarity measures when one (three) and two (four) nearest neighbors are compared. Targets with a preference for the four nearest neighbor forecasting policy have larger (smaller) standardized differences (squared differences) between their eGFR peak and trough (weight) dissimilarity measures when one and four (three) nearest neighbors are contrasted. Finally, patients with a preference for the 3-NN policy had larger (smaller) standardized differences between their solution cases' smooth eGFR time series (systolic blood pressure) closeness measures when one and two (three) nearest neighbors are compared.

The structure of both the first and the second step models support the conclusion that target patients in spatial regions that are sparsely populated are better forecasted with fewer neighbors than would be target patients in spatial regions that are more densely populated. In other words, for patients who are less typical, the increase in forecast accuracy that might result from using more neighbors is more than offset by the loss in accuracy that results from using neighbors that are not sufficiently similar to the target case. Still, the model ensures that targets within more densely populated feature space regions are forecasted using a larger sample of solutions, which reduces the sensitivity to noise, and helps obtain more accurate forecasts. Additionally, the first step of the model suggests that patients with large changes in their twelve-month time horizon eGFR readings should require the use of multiple solution cases. This highlights the importance of relying on rich CBRs with diverse

patient populations, which could accurately forecast both stable and deteriorating patients' disease progressions.

2.7 MODEL EVALUATION

We design the two step customized strategy for selecting relevant solution cases because using the same policy for predicting all patients' progressions could result in very inaccurate forecasts for some individuals. While we are interested in having good correct classification results, we also want to make sure that when a mistake a made it does not have a severe effect on our forecasting results.

Table 6: Two Step Model Test Set Confusion Matrix

	4n-NN	4e-NN	3n-NN	3e-NN	2n-NN	2e-NN	1-NN	N/A	Total
4n-NN	47	2	9	10	0	4	31	0	103
4e-NN	9	6	3	4	3	3	10	0	38
3n-NN	21	2	7	6	6	2	11	0	55
3e-NN	6	3	2	6	2	4	15	0	38
2n-NN	17	0	4	7	6	3	17	0	54
2e-NN	10	2	0	7	0	8	18	0	45
1-NN	41	3	5	20	4	8	85	1	167
Total	151	18	30	60	21	32	187	1	500

To evaluate the quality of our model we compare it against the seven pure strategies we have considered: (1) 1, (2) 2 equally weighted (2e-NN), (3) 2 non-equally weighted (2n-NN), (4) 3 equally weighted (3e-NN), (5) 3 non-equally weighted (3n-NN), (6) 4 equally weighted (4e-NN), or (7) 4 non-equally weighted (4n-NN) nearest neighbors.

One way of comparing our two step model to the pure strategies is by examining the number of correct predictions achieved using each method. Based on the test set a pure 1-NN/2e-

NN/2n-NN/3e-NN/3n-NN/4e-NN/4n-NN method will have a total of 167/45/54/38/55/38/103 correctly classified cases. The two-step model classifies correctly a total of 165 patients (Table 6). Therefore, it performs a lot better than the multiple nearest neighbors approaches and is comparable to the 1-NN method.

While the total number of correct classifications is important, we believe that the forecasting error accumulated when misclassifying a case should also be examined. To do that we calculated the difference between each of the eight nearest neighbors forecasting options and the optimal policy's accuracy measures. If the difference equaled zero that meant that the corresponding nearest neighbor policy was optimal. Similarly, if the difference measure was greater than zero that meant that the nearest neighbor policy selected a sub-optimal number of solution cases to forecast the patient with, which resulted in a less accurate forecast.

Table 7: Total Misclassification Error Test Set

Policy Total Misclassification	Error
2Step Model	1.476309929
Pure 1NN Strategy	2.303174733
Pure 2e-NN Strategy	1.805693226
Pure 2n-NN Strategy	1.805965463
Pure 3e-NN Strategy	1.641142481
Pure 3n-NN Strategy	1.636745673
Pure 4e-NN Strategy	1.531448847
Pure 4n-NN Strategy	1.527036367

We filtered out all patients which were correctly classified across each of the eight policies. We had a total of 335/333/455/446/462/445/462/397 misclassified cases using the two-step/pure 1-NN/pure 2e-NN/ pure 2n-NN/ pure 3e-NN/ pure 3n-NN/ pure 4e-NN/ pure 4n-NN policy. We compared the total forecasting inaccuracies accumulated as a result of using all eight policies (Table 7). Based on the results we concluded that the strategy that performed the worst across its misclassified cases was the 1-NN and the strategy that

performed the best was the two-step customized policy. The high total misclassification error associated with the 1-NN policy was even more troublesome considering that the number of misclassified cases added up to get this total value was smaller than that of any other strategy (333 cases). Additionally, we observed that the absolute difference between the total misclassification error across the equally and non-equally weighted solution cases increased in the neighbors' sample size.

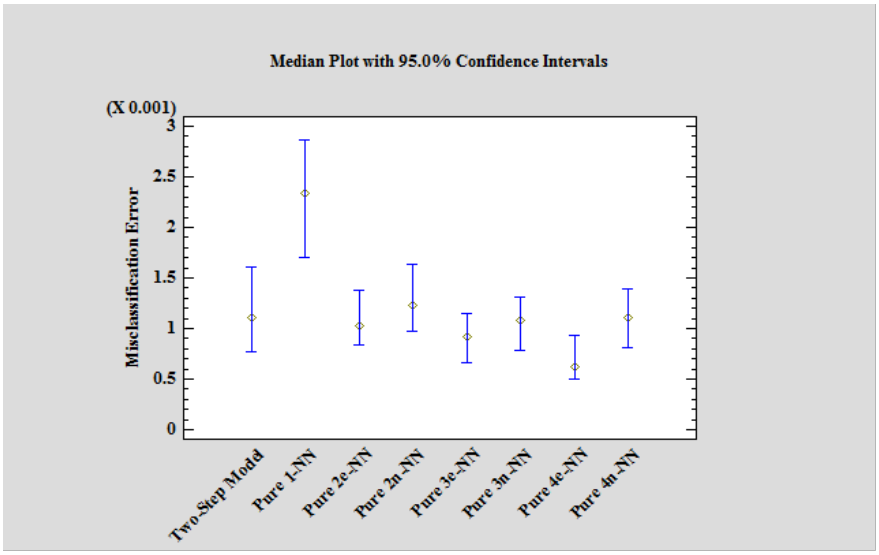


Figure 1: Misclassification Error

Furthermore, we used a Kruskal-Wallis test to examine the misclassification errors across the eight policies. We concluded that there was a significant difference between the policies (p-value < 0.001) and the 1-NN pure strategy had a significantly poorer performance across its misclassified cases than any of the other seven policies (Fig. 1). This meant that even though the 1-NN pure strategy had the highest correct classification rate, when it erred it erred by a lot. This is a cause for concern especially since the 1-NN approach is still commonly used in the CBR literature.

Based on our results we could conclude that the two-step model was preferable when compared against any of the seven pure strategies. First, it had a significantly lower misclassification error when compared to the 1-NN pure strategy and it had a compatible misclassification error performance when measured against the other pure policies. Second, it had

the lowest total misclassification error among all eight methods. Third, it had a competitive correct classification when compared against the 1-NN pure strategy and significantly better prediction capabilities than any of the multiple-NN models, i.e. it predicted correctly 60.19% more cases than the best multiple-NN model (4n-NN).

2.8 CONCLUSION

Patients with chronic diseases may be very care-demanding. Informed scheduling of such individuals requires accurate models for disease progression, so that limited appointment time slots are optimally allocated. In this paper, we presented an experience-based model that could serve as a critical component of a decision model to help medical providers in managing their patients. The model is set up to accommodate any chronic condition. We test it on patients with Chronic Kidney Disease because of their prevalence and complexities.

We focus on an important component of improving the accuracy of any CBR model - selecting the preferable number of nearest neighbors used when forecasting the trajectory of a target case. We are the first to design a two-step empirical model, which specifies the number of solution cases needed to create accurate forecasts. This allows us to determine the patients who can be predicted using one or multiple solution cases. Our selection mechanism can be easily updated for different CBR environments.

Our empirical results demonstrate that a model-based customization of the nearest neighbor determination problem is superior to any pure strategy. The preferable number of nearest neighbors appears to be related to the density of cases in the immediate vicinity of a target patient. Relatively isolated cases, meaning patients who are relatively unlike those in the provider's past experience, are better matched with a small number of neighbors. Those who are more typical are better forecasted with a greater number of past cases.

Our findings suggest several avenues of future research. First, we would like to examine different feature weighting techniques used when selecting CBR solution cases. We would like to examine the impact of forecasting patients two and three months at a time on the accuracy of the produced results, which would speed-up the long-term projections of a pa-

tient's trajectory. Finally, we believe that it is important to further examine the costs of misclassifying cases, and to incorporate that information into our analytical model.

3.0 WHEN PATIENT'S APPOINTMENT UTILIZATION SIGNALS THE NEED FOR A TARGETED-PALLIATIVE-CARE PREPARATION.

3.1 INTRODUCTION

Chronic kidney disease (CKD) is associated with a gradual or sudden loss of kidney function. According to the CDC, each year it kills more people than breast and colorectal cancer [121]. Not surprisingly, kidney disease is among the ten leading causes of death [27]. Its financial burden is evident when examining Medicare spending records. According to the United States Renal Data System [97], care for CKD patients over the age of 65 exceeded 50 billion dollars in 2013, which was approximately 20% of the Medicare budget for this age group.

The high demand for expensive care raises questions on how to better allocate patient appointments so that individuals are seen by medical professionals neither too often nor too rarely. As with any disease, there are official guidelines [64], which are designed to help clinicians in making treatment decisions. Additionally, researchers have developed "accurate, easy to implement, and highly generalizable" models for predicting CKD progression [112]. Still, nobody has examined how different consultation frequencies could predict patients' end-of-life trajectories.

The goal of our study is to use big data and quantify appointments' effect on one's kidney disease progression. Furthermore, we determine the range of annual consultation frequencies which signal that a patient's health is failing and a targeted-palliative-care preparation should be initiated.

3.2 METHODS

3.2.1 Study Population

Our patient cohort is extracted from the Corporate Data Warehouse of the U.S. Department of Veterans Affairs. The data set includes information on 44,751 CKD patients, treated at eleven Veteran Administration (VA) facilities located in California, Florida, Massachusetts, Nevada, New York, Pennsylvania and Texas, between January 1, 2009 and February 21, 2016. All patients had CKD stages 3, 4 or 5. Patients' outcomes consist of whether individuals went on dialysis, died or were alive and did not experience kidney failure by February 21, 2016.

The work is supported through master contract numbers VA244-13-C-0581 and VA240-14-D-0038 between the U.S. Department of Veterans Affairs and the University of Pittsburgh. Informed consent was not required due to the use of de-identified, administrative data and our inability to obtain such documentation from all cohort participants.

3.2.2 Variables

Independent Variable

The events of interest are death before dialysis and kidney failure defined by dialysis initiation. To estimate dialysis initiation, we use the first dialysis record in each patient's history. We examine the competing risks of death and dialysis onset because of their palliative care decision making implications. We are interested in both outcomes because we would like to examine the CKD-appointments' effect on patients near the end of their lives.

Dependent Variables

Similar to previous studies [112], we select predictors based on their face validity (Table 8). We record whether a patient had diabetes, heart failure, cirrhosis and peripheral or cardiovascular disease (PVD/CVD) at the end of the study. We record patients' eGFR and age at baseline [112]. To estimate the appointment effect we use patients' average number of annual CKD-related consultations during the study time period.

Comorbidities are not grouped because of their small correlations. Across patients who

died, the observed sample correlations fall between 0.02348099 (PVD/ CVD and cirrhosis) and 0.24516652 (PVD/CVD and heart failure). The range of comorbidity correlations for patients who went on dialysis is between 0.006019226 (PVD/CVD and cirrhosis) and 0.23667980 (PVD/CVD and heart failure). As noted, the pairs with the least and greatest correlation values are the same in both subgroups. This observation stays the same for patients who were alive and not on dialysis at the end of the study time period.

Some of the of the dependent variables are transformed to ensure that the Cox proportional hazard assumption is met. Due to the closed nature of the VA Healthcare System the dataset contains no missing data. Therefore, no imputation techniques are needed.

Table 8: Survival Analysis Predictors

Predictors	Possible Values
Diabetes	$\{0, 1\}$
Heart Failure	$\{0, 1\}$
Cirrhosis	$\{0, 1\}$
PVD/CVD	$\{0, 1\}$
Age	[22, 100]
eGFR at baseline	[0, 60]
Average number of annual appointments	[0, 88]

3.2.3 Statistical Analysis

Model Selection

We develop a Cox proportional hazards model with a fixed number of break-points in the appointment frequencies, age and eGFR predictors. The segmented relationships are used to highlight the changes in the effect of medical consultations, age and eGFR across different levels of the three variables. We build a multi-state survival analysis model, where the two terminal states are (1) dialysis and (2) death before dialysis because we cannot assume that

the two competing end points are independent.

Model Assumptions

We check if the proportional hazard, functional form and outlier [28] assumptions of the model are satisfied. To meet the proportional hazard assumption at $\alpha = 0.01$, we use predictors' stratification, interaction terms and variable transformations. We introduce segmented relationships across continuous predictor variables where such segmentation has a significantly positive contribution to the goodness-of-fit of our model.

Model Comparison

We use a standard information criterion measure (Akaike Information Criterion i.e. AIC [2]) for model selection. The metric determines models' appropriateness by examining the trade-off between their goodness-of-fit and complexity (number of predictors).

Table 9: Cohort Characteristics

Characteristics		VA Cohort (n=44,751)
Comorbidities, No. (%)	Diabetes	13,571 (30.326)
	Heart Failure	4,888 (10.923)
	PVD/CVD	19,571 (43.733)
	Cirrhosis	893 (1.995)
Age, mean (s.d.; No.)	< 65	58.72(6.15; 13,689)
	$65 - 85$	76.15(5.51; 27,086)
	> 85	88.15(2.19; 3,986)
eGFR	Baseline, mean (s.d.)	43.96 (12.10)
	$30 - 60$, No. (%)	38,816 (86.738)
	$15 - 30$, No. (%)	4,270 (9.542)
	$0 - 15$, No. (%)	1,665 (3.721)
Annual Appointments	Mean (s.d.)	6.493 (6.397)
	Min	0
	Max	88

3.3 RESULTS

3.3.1 Cohort Description

There were 3,147 (7,902) patients who experienced kidney failure (died prior to kidney failure). Most patients had a baseline eGFR between 30 and 60 mL/min/1.73m^2 s and were between 65 and 85, where the youngest (oldest) individual was 22 (100) years old. The most prevalent comorbidity was PVD/CVD followed by diabetes, heart failure and cirrhosis. The average number of annual appointments was a little over six, where on average some patients attended less than one and others up to eighty-eight appointments every year (Table 9). Appointments were calculated so that each unique consultation was counted separately. Thus, if a patient saw a cardiologist and a nephrologist in the same day this counted as two visits. We excluded patients with more than 88 annual appointments who represented 0.262% of the original dataset because they were considered outliers.

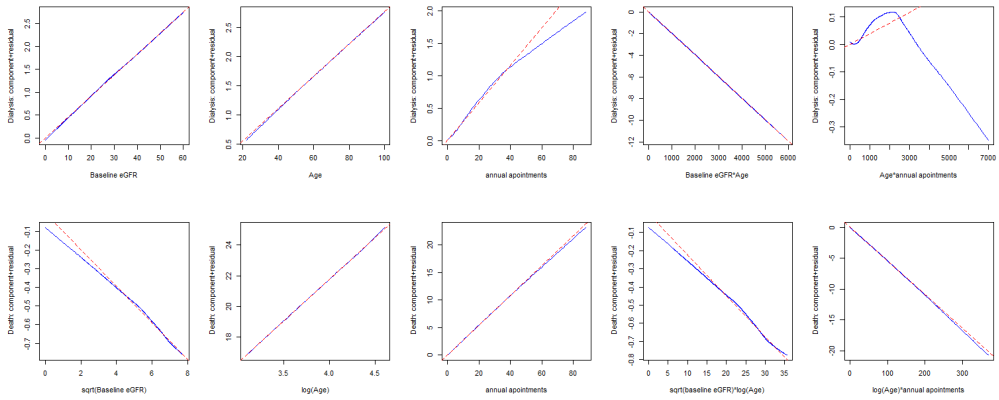


Figure 2: Linearity Assumption

3.3.2 Prediction Model

Our competing risk survival model was designed in three steps. First, we checked our model's assumptions. The first one required that subjects' censoring was not related to the probability of an event occurring. This design assumption was met, because we had no event information prior to obtaining our dataset. The second assumption pertained to the

proportional hazards of the model [80] and was met for both terminal states since the smallest kidney failure (death) proportional hazard p-value was 0.156 (0.0115). The influential points assumption was met sufficiently well. Finally, the linearity assumption [28] was substantially violated by the annual appointments predictors (Fig. 2), which highlighted the importance of using segmented Cox proportional hazard models.

Table 10: Breakpoint Predictor Significance

Terminal Event	Predictors	Breakpoint	Coefficient	$e^{coefficient}$	p-value
Death	Baseline eGFR^2	Before	-0.300611	0.740366	0.381345
		After	-0.635252	0.529802	0.000449
	log(Age)	Before	3.545338	34.651382	$3.65e^{-10}$
		After	7.228908	1378.7161	$< 2e^{-16}$
	Appointments	Before	0.319074	1.375853	$3.07e^{-10}$
		After	0.271964	1.3125397	$< 2e^{-16}$
Kidney Failure	Baseline eGFR	Before	0.1536677	1.1661033	$6.37e^{-11}$
		After	0.037726	1.0384467	$4.20e^{-7}$
	Age	Before	0.0369418	1.0376327	$1.60e^{-6}$
		After	0.0219917	1.0222353	0.036451
	Appointments	Before	0.0540712	1.0555598	0.000119
		After	-0.0139365	0.9861602	$< 2e^{-16}$

Once the optimal Cox proportional hazard model was selected, we performed a segmented regression analysis one at a time across each of the three continuous predictors: (1) age, (2) average number of annual appointments and (3) eGFR at baseline. We used the optimal breakpoints selected by the three segmented regressions to initialize the selection of our final model where all continuous predictors were segmented. The need for segmentation was significant at $\alpha = 0.05$ across all but one predictor variable regardless of whether the terminal event was death or kidney failure (Table 10).

3.4 COMMENTS

We estimate the annual appointments' threshold above which patient's demand for care signals failing health and requires palliative-care preparation. Our model uses patient information, which is readily available.

Table 11: Breakpoint Coefficient Information

Terminal Event	Predictors	Estimate	Standard Error
Death	Baseline $eGFR^{0.5}$	4.350420	0.10986086
	log(Age)	4.343341	0.04061622
	Annual Appointments	17.081968	0.06294703
Kidney Failure	Baseline eGFR	10.24653	0.08964794
	Age	56.97425	0.07003546
	Annual Appointments	18.38747	0.07414720

We conclude that increasing annual appointments up to a threshold indicates that a patient will live long enough to experience kidney failure before dying. At the same time, very frequent appointments signal that an individual will die before his/her kidneys fail. We observe that there is a significant decline (p-value $< 2*2e^{-16}$) in the probability of experiencing a kidney failure event when a patient is seen more than 17 times every year (Fig. 3 & Table 11). At the same time, our model suggests that the risk of dying grows as the number of annual appointments goes up and while the rate of increase is slightly smaller for patients seen more than 18 times per year (Fig. 4 & Table 11), it is still significantly greater than zero (p-value $< 2*2e^{-16}$). Therefore, the increase in patients' appointment frequencies signals deteriorating health and becomes an imminent death indicator when a patient requires over sixteen annual appointments.

Similar to previous studies [112, 31, 49], we conclude that the risk of kidney failure increases as patients' eGFR values go down. Additionally, our model suggests that the rate of increase in the kidney failure risk drops significantly (p-value $= 4.20e^{-07}$) if a patient has

eGFR above 10 (Fig. 3 & Table 11). Furthermore, the risk of death before dialysis decreases as patients' eGFR results increase (Fig. 4), where it is significantly smaller (p-value = 0.000449) for patients with eGFR approximately greater than 19 mL/min/1.73^2 s (Table 11). This is in agreement with previous findings, which concluded that "among patients of all ages, rates of both death and ESRD were inversely related to eGFR at baseline"[88].

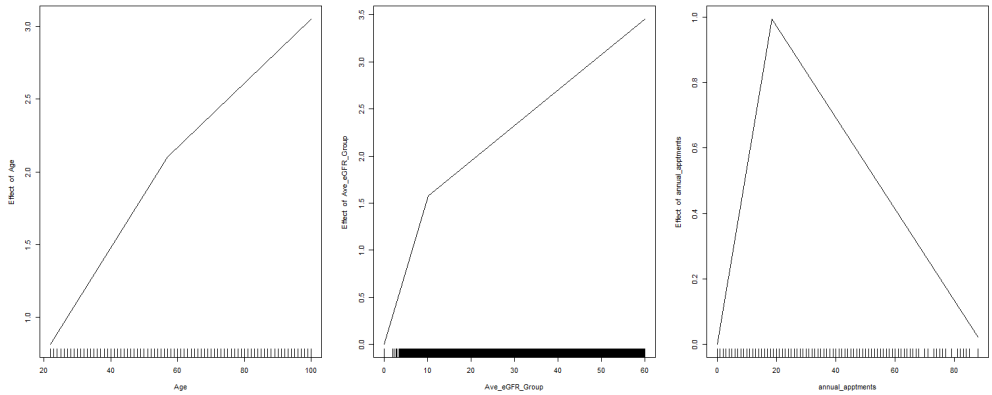


Figure 3: Breakpoint Effect on Kidney Failure Outcome

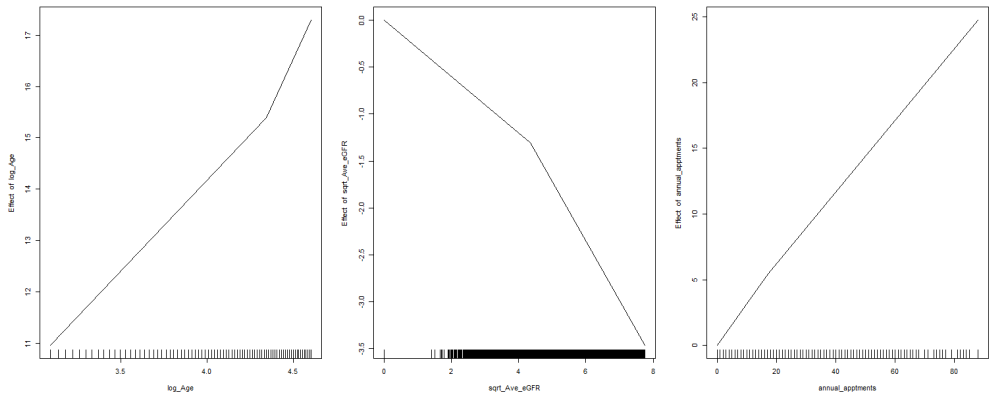


Figure 4: Breakpoint Effect on Death Outcome

Finally, we observe that the risk of death increase with age. Our results are consistent with previous research that "older patients had higher rates of death" than younger patients [88]. Furthermore, the risk of kidney failure (dying before dialysis initiation) rises significantly (p-values < 0.05) for patients above the age of 56 (76) (Table 11).

To the best of our knowledge, we are the first to quantify the appointment frequencies' effect on individual's CKD progression using data from the largest integrated healthcare system in the US. Our competing risk Cox proportional hazard model with segmented relationships suggests that consultations' data dichotomizes patients into normal and frequent users. This grouping can flag patients who will most likely die before going on dialysis. Thus, it can help provide better nursing home-based end-of-life care when behavioral, physical, social and participant-explained indicators are only beginning to manifest themselves [116]. Furthermore, the appointment frequency threshold has important implications for clinicians because it can facilitate the prioritization of CKD patients for formal palliative-care-focused appointments, which could help them make better-informed end-of-life treatment decisions [94].

What makes our conclusions more believable is the large and geographically diverse population we use. Our model's main result can be readily integrated into the palliative care decision making process of patients with CKD. Furthermore, following the approach taken in this paper we can estimate consultation frequencies' effect for different chronic conditions, as long as patients' appointment data is available. Our results are reinforced because "disparities in referral patterns and CKD care" are "less likely to be present" [112] across our patient population.

Still, using data extracted from the Corporate Data Warehouse of the U.S. Department of Veterans Affairs has its disadvantages. The major limitations are associated with its limited gender, age and race diversity. Due to the nature of the veterans' population our dataset contains a smaller percentage of females and minority patients than a representative sample of CKD patients. Therefore, our results cannot be generalized across ethnic and gender groups underrepresented in our analysis.

In conclusion, we are the first to develop an appointment-frequency-based indicator of targeted-palliative-care preparation. This objectively defined feature can be used to enrich and supplement current subjective and more loosely defined characteristics [116]. To ensure the appropriateness of our appointment frequency threshold it is advisable to perform the same analysis on different patient cohorts.

4.0 OPTIMAL STRATEGIES FOR MONITORING CHRONIC KIDNEY DISEASE AND PREDICTING PATIENTS' DEMAND-FOR-CARE.

"Kidney disease doesn't get the attention, funding or concern associated with cancers of the breast or prostate. But it actually kills more Americans - 90,000 a year - than both malignancies combined." [14]

4.1 INTRODUCTION

Chronic kidney disease (CKD) is defined by the slow decline in kidneys' filtration abilities. It has five stages, where the final one, called end stage renal disease (ESRD), is reached when a patient experiences a complete kidney failure. CKD is the ninth leading cause of death in the US [27]. It is estimated that between 11 and 16.8% of the US population have CKD [25, 16]. The disease is largely asymptomatic in its earlier stages (stages 1 through 3), which could explain why only about 10% of such patients are diagnosed [86]. NIDDK reported that there were 117,162 new ESRD cases in 2013, and although the incidence rate has recently stabilized, the number of prevalent cases continues to rise by about 21,000 every year [98].

Data from 2013 suggests that approximately 20% of all Medicare spending went towards treating kidney disease patients [98]. This is because CKD has a disease multiplier effect, i.e. approximately 55% of all CKD patients have diabetes and self-reported cardiovascular disease (CVD) [86]. Therefore, kidney patients have higher hospitalization and 30-days-readmission rates and pose a heavy financial burden on the healthcare system.

Like other chronic diseases, CKD cannot be reversed. Still, its progression can be slowed down and even halted if proper care is administered [19]. Unfortunately, many studies

suggest that kidney disease management has been suboptimal [6, 17, 40, 45, 89, 92, 114].

Standard care involves nephrologist appointments, where patient's kidney function and comorbidities' development are assessed. The laboratory value used to check how well one's kidneys are working is called Glomerular Filtration Rate (GFR). Due to its high cost, GFR is estimated (eGFR) using patient's age, race, gender and blood creatinine information. The noisy eGFR measures are accumulated over time and used to determine individual's CKD stage and rate of disease progression, which guide treatment changes aimed at slowing down patient's kidney failure. Individuals also need to be examined by cardiologists as chronic kidney disease affects cardiovascular disease outcomes [99], by endocrinologists because of the high risk of kidney failure among diabetic patients [26], etc.

To prevent further health deterioration, individuals need to see various medical specialists during well-timed appointments. Visits should be scheduled so they are neither too frequent and overwhelming to the patient and the healthcare system, nor too scarce and incapable of detecting progression early enough to prevent hospitalizations and early death.

Current guidelines designed by the National Institute for Health and Care Excellence (NICE) suggest that patients in stages (1 & 2), 3, 4 and 5 should be seen once every 12, 6, 3 and 1.5 months, respectively. Like any population-based policy, it is suboptimal because of its inability to incorporate the effect different coexisting conditions have on patient's risk of disease progression. Furthermore, current guidelines do not consider patient-specific disutilities associated with appointment visits which are a function of patient's mobility [117]. Mobility is particularly important for our patients (US Veterans) because according to the U.S. Census Bureau's American Community Survey approximately 24.1% of our population of interest lives in rural areas. Furthermore, the survey found that the median age of the rural Veterans was 15 years higher than that of urban Veterans, and this discrepancy increased in the level of rurality. Due to their age and health complications many rural Veterans rely on a centralized transportation system (Veterans Transportation Program) to take them to the regional medical center early in the morning and return home later in the afternoon regardless of their appointment time, which is often inconvenient and results in long wait times.

4.1.1 Chronic Disease Treatment Model

Despite the great need for personalized care models, disease detection research [8, 23] is much more prevalent than the work on optimizing disease monitoring [38]. What makes a monitoring problem challenging is the disease deteriorating nature, which demands continuous treatment adjustments made during a series of optimally spaced appointments.

We develop a research framework for optimizing the appointment frequencies of chronically ill patients. Our approach is applied to individuals with chronic kidney disease but can be easily adapted for patients with diabetes, heart failure, etc. The information needed to implement our model includes (1) knowledge of the features characterizing the chronic disease and (2) data on the population of interest. Our approach involves the development of three models. The first is a case-based reasoning (CBR) model, which quantifies the impact of appointment frequencies on patients' disease progressions. The second is a Cox proportional hazards model [54], which predicts patients' life expectancy given their age and health state. Finally, we use a finite horizon Markov Decision Process (MDP), which is parametrized using outputs from the first two models, to optimize patients' monitoring strategies by maximizing their expected quality-adjusted life days (QALDs). Our CBR model estimates the transition probability matrices and our survival analysis model helps determine the terminal rewards used in the MDP. Details on our framework can be found in §4.4.1.

We choose a finite horizon MDP because it assists doctors in making evidence-based and data-driven follow-up appointment decisions. It is also useful for gauging the demand for non-emergency care at a clinic or a network of clinics treating patients with a specific chronic condition. The demand for future regular appointment visits can be quite accurately predicted by aggregating the optimal appointment frequencies across all clinics' patients over a pre-specified time horizon. Such information is useful when deciding if a clinic needs to recruit more clinicians or patients.

To the best of our knowledge, this is the first model-based approach to optimize followup chronic disease appointment frequencies to help both clinicians and management better prepare for the demand of non-emergency visits. Previous work on personalizing chronic-disease care [38] suggests when to see a patient next, which is useful when making short-term

scheduling decisions. Our model provides medium-term (approximately 2 years) demand-for-care estimates, which are important for capacity adjustments as it takes time to hire highly qualified personnel.

Our model uses the appointment frequencies suggested by NICE as the basis for action space alternatives. The equispaced appointment schedules recommended by NICE are special cases of our MDP policies. To better understand the extreme NICE policies, we pinpoint the conditions under which seeing a patient as frequently and as rarely as possible is selected by our MDP, this signals when the current extreme guidelines are optimal.

Besides contributing to the MDP and chronic disease modeling literature, our optimal monitoring strategies can be easily applied to CKD patients with different comorbidities and from various age groups. We compare our policies to the NICE guidelines to highlight the importance of stratifying patients across more than just their kidney disease stages when optimizing their appointments' timing.

All models in our research are parametrized using a diverse patient cohort treated at the U.S. Department of Veterans Affairs hospitals. The data set included information on 68,514 patients with CKD stages 3 through 5, treated at eleven geographically diverse Veterans Administration (VA) facilities between January 1, 2009 and February 21, 2016.

The remainder of the paper is organized as follows. Section 4.2 reviews the literature. We propose the MDP model for chronic kidney disease management in section 4.3. In section 4.4, we discuss estimates used to parametrize the model. Empirical results and structural properties are summarized in section 4.5. Demand planning results are provided in section 4.6, followed by our conclusion in section 4.7.

4.2 LITERATURE REVIEW

We discuss three streams of literature most relevant to our research. They are (1) Markov decision processes, (2) disease management modes for optimizing treatment regimens, and (3) healthcare demand planning.

4.2.1 Markov Decision Process Models

MDPs are well equipped to address decisions that are sequential and uncertain [100]. Over the last ten years the growth in incidence rate and treatment cost of chronic diseases and cancer have prompted research interest in their treatment optimization. What makes MDP models particularly attractive for healthcare researchers are their ability to incorporate different goals. MDPs can be used to examine a problem from both the patient [8] and the population [74] perspectives. Additionally, they can be used to increase patients' length of life [73, 4], reduce cost [63], or design policies based on their cost and effect implications [37].

Table 12a: Disease Management Methods

Reference	Solution Methodology	Perspective
Alagoz et al. [4]	Discrete-time, infinite- horizon, discounted MDP	patient
Ayer et al. [8]	Discrete-time, finite-horizon POMDP	patient
Erenay et al. [23]	Discrete-time, finite-horizon POMDP	patient
Helm et al. [38]	Linear Gaussian system modeling	patient
Lee et al. [61]	MDP	population
Skandari et al. [107]	Dynamic Programming	patient
Our Research	Discrete-time, finite-horizon MDP	patient

Table 12b: Disease Management Methods

Authors	Chronic Disease	Appt. Frequency Impact	Demand-for-Care
Helm et al. [38]	√		
Our Research	√	√	√

CKD researchers have used dynamic programming to improve patients' treatment by optimizing the timing of vascular access needed for hemodialysis [107] and dialysis initia-

tion [61]. However, no research has hitherto developed a MDP-based model to address the management of an ongoing chronic disease. We employ a MDP model because it does not dichotomize patients into progressors and non-progressors [38]. Such a categorization of time series trajectories requires a judgment call that might be subjective.

4.2.2 Disease Management

Our research focuses on developing models to evaluate appointment schedules for chronically ill individuals. We reference methods discussed by [107] and others and contrast them in Tables 12a – 12b. Depending on the disease in question, authors frequently select regular or partially observable (PO) MDP models. Additionally, they apply simulation/forecasting techniques to parametrize their optimization models.

Our method differs in the generalizability of its components (Chapter 2-3). We deploy a CBR model [4] to parametrize our optimization model. We chose a CBR to forecast patient trajectories because it can accommodate complex data and estimate monitoring strategies' effects on one's disease progression. Still, we minimize the reliance on subjective inputs from clinicians and employ an analytic approach to identify a cohort of nearest neighbors used to forecast the disease progression of a new patient (see Chapter 2). This allows us to quickly replicate our model across different chronic conditions once we know their descriptive characteristics.

4.2.3 Demand Planning

Predicting demand for care is not new [20]. What separates our work from previous models is the novel approach used to estimate the medium-term demand for chronic disease monitoring appointments. Accurate demand forecasts are crucial when making staff recruitment decisions. This is particularly important in medical specialties with declining graduation rates (e.g. nephrology) and increasing patient populations [42, 90].

In their review of the literature on capacity and demand management, Jack & Powers [46] discuss aspects of the healthcare industry which make it challenging to model. The authors highlight benefits of deploying demand management strategies in hospitals and clinics. They

suggest that more research should be done on estimating the demand for care given an increase in healthcare access. While this is not the main focus of our research, we believe that it can be used to address such concerns if we can estimate the health status of all new patients who gain access to the healthcare system.

4.3 MODEL SETTINGS

To maximize patient's total quality-adjusted life days (QALDs), we choose a finite horizon discrete state MDP model. The problem is set from the patient's perspective; therefore, we do not consider cost-of-care. This assumption is realistic for our patient population - US Veterans - due to their military service-based healthcare benefits. We also assume that all individuals are risk neutral.

A patient's current state determines our model's recommendations on seeing a doctor in 1.5, 3, 6 or 12 months. The options are based on appointment frequency guidelines proposed by NICE. While the guidelines suggest that patients should see a doctor at equal time intervals determined by their CKD stage, our model recommends monitoring strategies based on one's overall state, such that the time between appointments does not have to be constant.

All patients are characterized by their age, CKD stage and comorbidities. We focus on patients between the ages of 60 and 90 who have moderate to severe chronic kidney disease (Stages 3 - 5). We select this subgroup because it demands a lot of care and as noted earlier lower stages are often not diagnosed [86]. We also record if patients have diabetes, vascular disease (peripheral or cardiovascular) and heart failure, because these comorbidities are common among CKD patients. We assume that comorbidities cannot be cured (Fig. 5a), patients age, and kidneys' filtration abilities cannot improve, which causes CKD progression (Fig. 5b).

We assume that patient's eGFR is obtained at every appointment and used to determine one's CKD Stage. Patient's health determines the type of nephrologist clinic he or she attends. Prior to dialysis a patient visits a standard nephrology clinic; once dialyzed the

patient transitions into a dialysis clinic. In this model, we focus on the care provided prior to dialysis because most patients die before dialysis is initiated. Therefore, a patient stays in the system if no dialysis initiation or death is observed. Since dialysis and death are terminal states the transition between them will not be considered in our model (Fig. 5b).

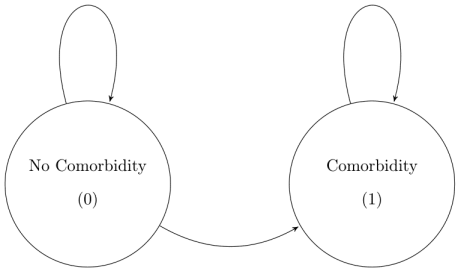
Table 13: State Space

State	Settings
Kidney	$\{3, 4, 5, 6, 7\}$
Diabetes	$\{0, 1\}$
Heart Failure	$\{0, 1\}$
Vascular Disease	$\{0, 1\}$
Age	$\{60, 61, 62, \ldots, 89, 90\}$

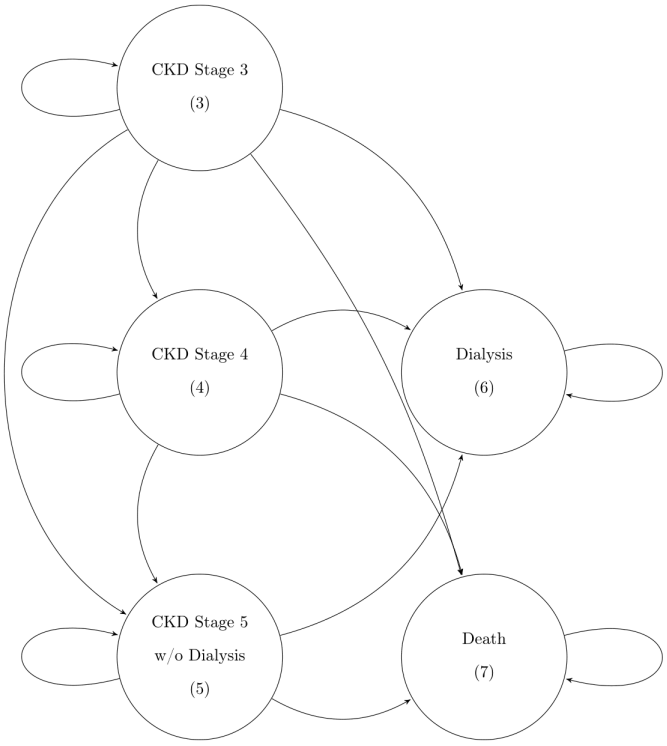
The notations necessary for developing our models are discussed next.

- n = 1, 2, ..., N, N $< \infty$. Each decision epoch equals 1.5 months, which corresponds to the smallest time interval between appointments suggested by NICE. N is selected by the model user and is the planning horizon for which scheduling appointments is intended. For example, if we need to optimize appointment frequencies over a ten-year period N = 80 = 10∗8 because there are eight 1.5-month periods in a single year. In our numerical examples (section 4.5) we set N = 16 = 2 years. We chose this value because (1) we find that our optimal policies are stable across all actions when the time horizon exceeds one year and (2) it is clearer and easier to describe the numerical results when the time horizon is relatively short.
- $S = \{S_D \otimes S_H \otimes S_V \otimes S_K \otimes S_{Age}\}$. The health state of a patient s = (d, h, v, k, age) $\in$ S specifies the diabetes (D), heart failure (H), vascular disease (V), kidney disease (K) and age (Age) group an individual belongs to (Table 13). All comorbidities are either present (1) or absent (0). The kidney state consists of five stages: 3 represents CKD Stage 3; 4 – CKD Stage 4; 5 – CKD Stage 5 without dialysis; 6 – dialysis; and 7 – death. States 6

and 7 are both terminal as patients will leave our system if they progress to these states (attend a dialysis clinic or pass away). We consider patients between the ages of 60 and 90.



(a) Comorbidity States



(b) CKD & Terminal States

Figure 5: State Space

- a_i – the action i, where $a_i \in$ A $= \{1.5; 3; 6; 12\}$ means seeing a doctor in 1.5/3/6/12 months; $\{a_1; a_2; a_4; a_8\} = \{1.5 * 1; 1.5 * 2; 1.5 * 4; 1.5 * 8\} = \{1.5; 3; 6; 12\}$.
- $[P_i]_s = P(j|s, a_i)$ – the transition probability across all states (j) given that action a_i is chosen and a patient is currently in state $s \in S = \{S_D \otimes S_H \otimes S_V \otimes S_K \otimes S_{Age}\}$. Since, we assume that comorbidities cannot be cured and CKD cannot be reversed, the probabilities associated with curing one or more comorbidities or transitioning into a lower kidney disease stage are set to zero.
- $p_{i_{(s,m)}}$ - the probability of transitioning from state s to state m given that action i is chosen; i.e. the m^{th} entry of $[P_i]_s$.
- $p_{k,i}$ - the probability of not going on dialysis or dying ($k \neq 6, 7$) before the next scheduled appointment given that the patient is currently in kidney state k and action i is chosen (Fig. 6).
- $p^{'}_{k,i}$ - the probability of initiating dialysis before the next scheduled appointment given that the patient is currently in kidney state k and action i is chosen (Fig. 6).
- $p_{6,i}$ - the probability of remaining on dialysis (k = 6) for a time interval of length 1.5∗i.

Figure 6: $p_{k,i}$ and $p^{'}_{k,i}$

- $disu_s$ – the disutility associated with being in state s = (d, h, v, k, age) ∈ S, where $disu_s = disu_d + disu_h + disu_v + disu_k + disu_{age} + disu_{\#(1+d+h+v)\ coexisting\ comorbidities}$. For more information refer to §4.4.2.

- $disu_{s'}$ - the disutility associated with being in state and $s^{'} = (d, h, v, k^{'}, age^{'}) \in$ S, where the only difference between $s^{'}$ and s is that the patient is on dialysis ($k^{'} = 6$) and she or he is in a different age state ($age^{'}$), i.e. $s^{'} = (d, h, v, 6, age^{'})$. For more information refer to equation 4.1.
- $disu_{appt}$ – the disutility associated with attending an appointment, which is especially important for individuals who live in remote areas and travel long distances to meet with a CKD expert.
- $[r_i]_s = r(s, s^{'}, a_i, disu_{appt})$ – an immediate reward, which represents the expected QALDs accrued when a patient is in health state $s \in S$ and action a_i is taken, which is associated with the disutility of attending an appointment in 1.5*i months from today ($disu_{appt}$) and the possible transition into state $s^{'} = (d, h, v, k^{'}, age^{'}) = (d, h, v, 6, age^{'}) \in S$. It depends on the time a patient will spend away from his doctor before attending a follow-up appointment. It incorporates health state and appointment attendance disutility measures. Furthermore, it reflects the probability that an individual could die or go on dialysis before a scheduled appointment occurs.
- $[R_N]_s = R_N(s, s^{'}) = R_N(d, h, v, k, age, d, h, v, k^{'}, age^{'}) = R_N(d, h, v, k, age, k^{'}, age^{'})$ – the terminal reward, which represents the total expected QALDs for a patient in health state s = (d, h, v, k, age) $\in$ S in period N, who could go on dialysis before dying ($s^{'} = (d, h, v, k^{'}, age^{'}) = (d, h, v, 6, age^{'}) \in S$).
- $LE_{B.D.}(s)$ – the life expectancy of a patient in state s = (d, h, v, k, age) $\in$ S before entering dialysis.
- $LE_{A.D.}(s^{'})$ – the life expectancy after dialysis initiation for a patient who started dialysis in state $s^{'} == (d, h, v, k^{'}, age^{'}) = (d, h, v, 6, age^{'}) \in S$.

Note that in all notations above if s and a_i are known then $s^{'}$ is also known since $s = (d, h, v, k, age)$ and $s^{'} = (d, h, v, 6, age^{'})$, where $age^{'} = age + 1.5i$ in $[r_i]_s$; or $age^{'}$ is value obtained from our Cox proportional hazard model in $[R_N]_s$. Therefore, we can simplify our notations:

- $[r_i]_s = r(s, s^{'}, a_i, disu_{appt}) = r(s, a_i, disu_{appt})$
- $[R_N]_s = R_N(s, s^{'}) = R_N(s)$

Our model accounts for the possibility that while waiting for an appointment a patient may experience an observable complication, which could result in death or dialysis initiation. If no deterioration is detected a patient will not attend any emergency appointments and will wait until the next regularly scheduled visit. Therefore, we use a "No ER" notation in Figure 7 to mark the alternative associated with following the monitoring strategy suggested by our model. Most MDP models assume that the decisions are either act now or wait for another time-period. Our action settings are different because medical professionals would need to set follow-up appointment in advance due to the high demand for specialized care, which makes the wait-and-set rule harder to enforce.

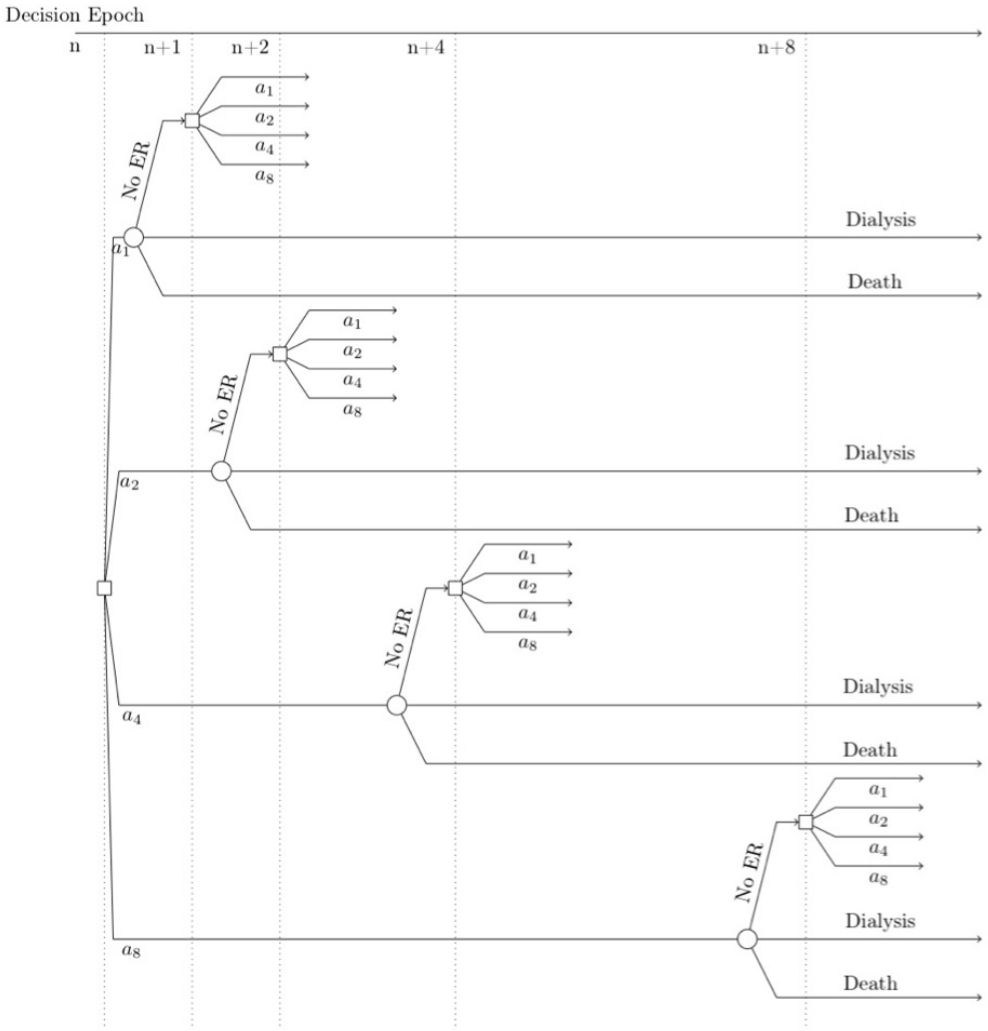


Figure 7: Decision Process

4.3.1 Optimality Equation

In our model $[v_n^*]_s = v_n^*(s, disu_{appt})$ represents the maximum total expected quality-adjusted life days in period n (n = 1, 2, ..., N - 1) when the patient is in state s and his or her appointment disutility value is $disu_{appt}$:

$$v_n^*(s, disu_{appt}) = max_i\{r(s, a_i, disu_{appt}) + \sum_j P(j|s, a_i)v_{n+i}^*(j)\}, i = 1, 2, 4, 8 \tag{4.1}$$

where

$$v_N^*(s, disu_{appt}) = R_N(s) = \begin{cases} (1 - disu_s)LE_{B.D.}(s) + (1 - disu_{s'})LE_{A.D.}(s') & \text{if } k < 6 \\ (1 - disu_s)LE_{A.D.}(s) & \text{if } k = 6 \\ 0 & \text{if } k = 7 \end{cases}$$

and

$$r(s, a_i, disu_{appt}) = \begin{cases} p_{k,i}\frac{365}{8/i}(1 - disu_s) + p'_{k,i}\frac{365}{8/i}(1 - disu_{s'}) - disu_{appt} & \text{if } k < 6 \\ p_{6,i}\frac{365}{8/i}(1 - disu_s) & \text{if } k = 6 \\ 0 & \text{if } k = 7 \end{cases}$$

We can reduce the formula to $v_n^*(s, disu_{appt}) = max_{i=1,2,4,8}[r_i + P_i v_{n+i}^*]_s$. Therefore, $v_n^i(s, disu_{appt}) = [r_i + P_i v_{n+i}^*]_s$ is the total expected quality-adjusted life days in period n if a patient is in state s, with an appointment disutility equal to $disu_{appt}$ and action i is chosen. It should be noted that actions are dependent on the length of the time horizon, i.e. action a_i is a possible alternative for all n $\in$ [1, N - i].

The terminal reward $R_N(s)$ is obtained using survival analysis length of life estimates, which are adjusted with disutility measures reported in medical journals. For a patient in a nonterminal kidney state ($k < 6$), we estimate the highest QALDs right before reaching dialysis or death by multiplying the life expectancy prior to a terminal state $LE_{B.D.}(s)$ and the utility associated with the patient's current health state $(1 - disu_s)$. This represents the highest QALDs because we assume that our patient's next stage of progress will be dialysis or

death. Similarly, we estimate the highest QALDs between dialysis initiation and death, if the patient is predicted to transition into dialysis prior to passing away, as $(1-disu_{s'})LE_{A.D.}(s')$ where the only difference in s' and s is observed in patient's kidney and age states. When added up, the two measures represent the maximum total expected QALDs for a patient in state s. Similarly, we obtain $R_N(s)$ when k=6 and we set $R_N(s)$=0 for patients who have passed away.

The immediate reward $r(s, a_i, disu_{appt})$ calculations account for the possibility of a patient dying or going on dialysis prior to his or her follow-up appointment. The reward is action dependent, i.e. if action i is chosen the patient will be seen in approximately $\frac{365}{8/i}$ days. We multiply the days between appointments $(\frac{365}{8/i})$ by the patient's health state dependent utilities $(1 - disu_s)$ to obtain the quality adjusted days a patient will gain given that he or she is in state s and action i is chosen. The same value is multiplied by $(1-disu_{s'})$ to obtain the quality adjusted days a patient will gain if he or she is in state s' and action i is chosen, where s' is the same as s except for the patient's (1) kidney state, where k is updated to 6 (i.e. the patient is on dialysis), and (2) age, which is updated to $age' = age + 1.5i$. The two values are multiplied by the probability of not transitioning into a terminal state and the probability of transitioning into a dialysis state, given the patient was initially in state s and action i was chosen, respectively. The formula represents the expected quality adjusted days gained for a patient in state s who follows policy i. We subtract $disu_{appt}$ from the reward because that is the time that the patient will lose due to attending a follow-up appointment. The same logic is used to obtain $r(s, a_i, disu_{appt})$ when k = 6. We do not include $disu_{appt}$ in $r(s, a_i, disu_{appt})$ when k = 6 because once on dialysis the patient will leave the system and will stop accumulating the appointment disutilities associated with k < 6.

4.3.2 Structural Properties of the MDP

As noted earlier the NICE policy is a special case of the monitoring strategies proposed based on the MDP. Because the NICE guideline is currently used by doctors, we would like to understand the conditions under which they will be selected as part of our optimization model. Below, we will present the NICE results associated with the extreme cases, i.e. see

a patient as frequently as possible (every 1.5 months) and as rarely as possible (once per year). The proofs accompanying the lemmas and propositions are provided in Appendix B. For our proofs to work, we need to introduce a well-known inequality proven by Karamata [50]:

Lemma 1 (Karamata, 1932)**.** *If (1)* $[P_i]_{s,s} \geq [P_i]_{s,s+1} \geq ... \geq [P_i]_{s,s^*}$ *and* $[P_{i'}]_{s,s} \geq [P_{i'}]_{s,s+1} \geq ... \geq [P_{i'}]_{s,s^*}$ *; (2)* $\{[P_{i'}]_s\} \succeq \{[P_i]_s\}$*, i.e.* $\sum_{j=s}^{l}[P_{i'}]_{s,j} \geq \sum_{j=s}^{l}[P_i]_{s,j}$ *for all* $l \in [s, s^*]$ *with an equality when* $l = s^*$*; and (3) f is* ***convex****, then* $\sum_{j=s}^{s^*} f([P_{i'}]_{s,j}) \geq \sum_{j=s}^{s^*} f([P_i]_{s,j})$ *(for f -* ***concave*** *the reverse inequality holds).*

For Lemma 1 to be valid both $[P_i]_{s,j}$ and $[P_{i'}]_{s,j}$ should be non-increasing in j. This is reasonable because a slight disease progression is more likely than a significant health deterioration, especially if a patient is monitored. The second assumption suggests that action $a_{i'}$ is associated with a lower probability of progression from state s to any more care demanding state ($l \neq s^*$) when compared to action a_i, which makes action $i^{'}$ more attractive. $\sum_{j=s}^{s^*}[P_{i'}]_{s,j} = \sum_{j=s}^{s^*}[P_i]_{s,j} = 1$ because (1) none of the comorbidities we consider can be cured, (2) CKD cannot be reversed and (3) patients get older with time. The convexity/concavity assumption will be discussed below.

Lemma 2. *If (1)* $[(P_1^{n+1} - P_1^n)v_N^*]_s - P_1^{n+1}(s,s)[r_1^{\circ(n+1)}]_s$ *is non-decreasing in n, (2)* $[r_1 + (P_1 - I)v_N^*]_s \geq 0$ *, (3)* $[r_1^{\circ(n-1)}]_s(P_1^n(s,s)[r_1]_s - P_1^{n-1}(s,s)) \geq max\{[r_1]_s, \frac{[r_2]_s}{2}, \frac{[r_4]_s}{4}, \frac{[r_8]_s}{8}\}2*(n-1)$ *for* $n \in$ *[1, N] & (4)* $[P_1^j]_s$ *first order stochastically dominates* $[P_j]_s$*, then* $[v_{n-1}^* - v_n^*]_s \geq 0$ *& non-increasing in* $n \in$ *[1, N].*

To ensure that the numerical series are non-decreasing and the third assumption is met the patient should be sick and his immediate reward $[r_1]_s$ has to be relatively small but still greater than 1 (i.e. $P_1(s,s)[r_1]_s$ approximately equals one). The second assumption guarantees that $[v_{n-1}^* - v_n^*]_s \geq 0\ \forall n$. The right hand side in the third assumption could be viewed as the highest QALDs a person would have over n - 1 time periods given his current health state. The last assumption suggests that $E_{1^j}[v_{n+1}^*]_s \geq E_i[v_{n+1}^*]_s$, where $E_{1^j}[v_{n+1}^*]_s = [P_1^j v_{n+1}^*]_s$ and i = 2, 4, 8.

Lemma 3. *If* $P_{i'}$ *for* $i^{'} = \{1, 2, 4, 8\}$ *is convex and Lemma 2 holds, then* $[P_i v_{N-q+i}^* - P_1 v_{N-q+1}^*]_s \geq [P_i v_{N-(q+1)+i}^* - P_1 v_{N-(q+1)+1}^*]_s$*, where* $i \in \{2, 4, 8\}$.

For Lemma 3 to hold we assume that the difference between the value function in periods t and t - 1 do not increase as t approaches N, which occurs under the assumptions in Lemma 2. If P is convex and it is multiplied by a positive number the new function is still convex. The convexity assumption is in line with the logic that the function has a limited downside, i.e. $min(f([P_{i'}]_s)) = 0$ when $[P_{i'}]_{s,j} = 0, i^{'} = 1, 2, 4, 8$ ($[P_{i'}]_{s,j} \in [0, 1]$).

Proposition 1. *If Lemma 2 and* $[r_1 - r_i]_s \geq [P_i v_N^* - P_1 v_{N-i+1}^*]_s, i \in \{2, 4, 8\}$ *hold, then for a patient in health state s the constant policy* a_1 *is optimal .*

Proposition 1 highlights that for the most frequent appointment policy to be constantly optimal a patient needs to be very sick. The assumption $[r_1 - r_i]_s \geq [P_i v_N^* - P_1 v_{N-i+1}^*]_s, i \in \{2, 4, 8\}$ ensures that the finite horizon will not prevent policy a_1 from being optimal across all N - 1 decision points. This is in line with the guidelines given by NICE, which suggest that very sick patients should be seen as frequently as possible by their medical providers.

Lemma 4. *If (1)* $[(P_1 - I)P_1^n(v_N^* - r_8^{\circ(n+1)})]_s$ *is non-increasing in n, (2)* $[r_1 + (P_1 - I)v_N^*]_s \leq 0$, *(3)* $[(I - P_1)(P_1^n r_8^{\circ(n+1)} - P_1^{n-1} r_8^{\circ(n)})]_s \geq 2n * max\{[r_1]_s, \frac{[r_2]_s}{2}, \frac{[r_4]_s}{4}, \frac{[r_8]_s}{8}\}$ *for* $n \in$ *[1, N] & (4)* $[P_1^j]_s$ *first order stochastically dominates* $[P_j]_s$, *then* $[v_{t-1}^* - v_t^*]_s \geq 0$ *& non-decreasing in* $t \in$ *[1, N].*

To ensure that the numerical series are non-increasing the patient should be healthier, i.e. $r_8(s)$ has to be relatively large. The additional assumptions are in line with the logic behind the requirements listed in lemma 2.

Lemma 5. *If* $P_{i'}$ *for* $i^{'} = \{1, 2, 4, 8\}$ *is concave and Lemma 4 holds, then* $[P_i v_{N-q+i}^* - P_8 v_{N-q+8}^*]_s \geq [P_i v_{N-(q+1)+i}^* - P_8 v_{N-(q+1)+8}^*]_s$, *where* $i \in \{1, 2, 4\}$.

Under lemma 4 $[v_{t-1}^* - v_t^*]_s$ is non-decreasing as t approaches N. As noted above the settings hold if the person under consideration is relatively healthy. P has to be concave because when multiplied by a negative number, a concave function is transformed into a convex one.

Proposition 2. *If Lemma 5 and* $[r_8 - r_i]_s \geq [P_i v_{N-i+8}^* - P_8 v_N^*]_s, i \in \{1, 2, 4\}$ *hold, then for a patient in health state s the constant policy* a_8 *is optimal.*

Proposition 2 suggests that infrequent appointments are appropriate for healthier individuals with more stable disease progression. Thus, we conclude that our findings are in line with the current guidelines, which suggest that patients should be seen once a year only if they are reasonably healthy.

4.4 MODEL INPUTS

In the following section, we describe our research framework. Subsequently, we discuss the source of parameters used in the MDP. The parameters are derived through historical data, not estimated by subjective judgments.

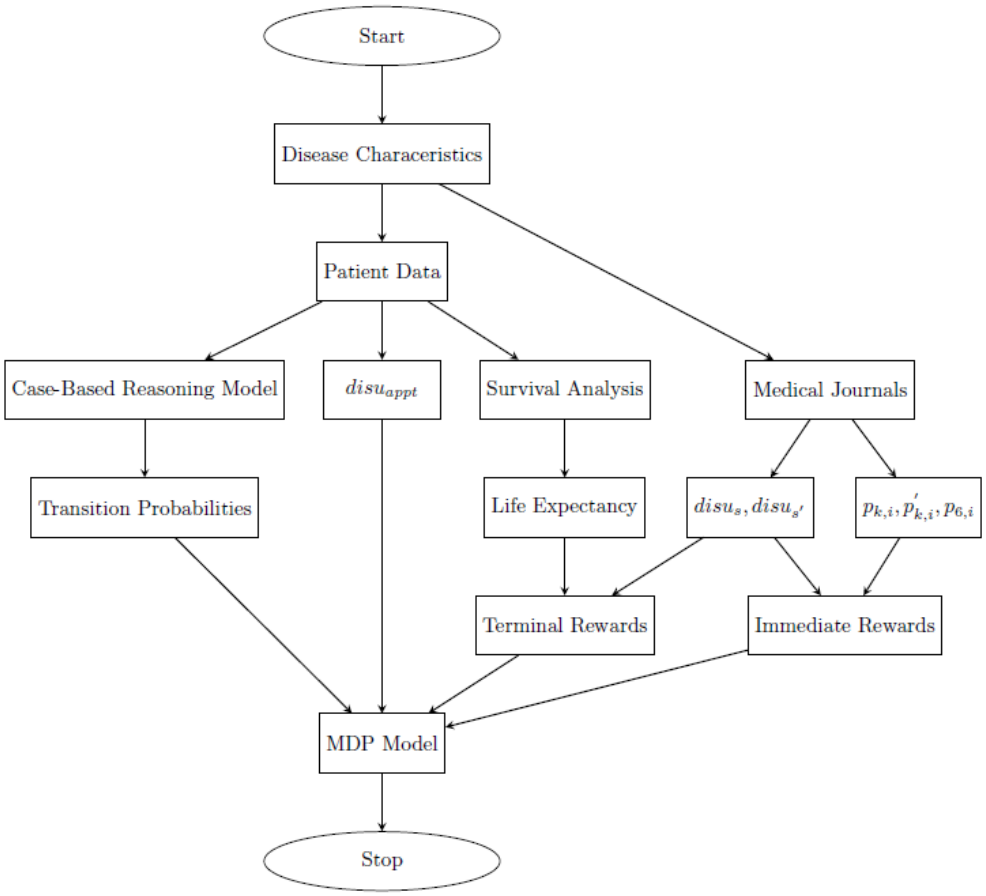


Figure 8: MDP Model Framework

4.4.1 Modeling Framework

We optimize the frequencies of chronic disease monitoring appointments over a prespecified time horizon (Figure 8). Our approach is used on a population with chronic kidney disease. Still, it can be easily applied to improve the care of patients with other chronic conditions. The framework consists of three models: two parameter estimation models (CBR, Survival Analysis) and one optimization model (MDP). It also requires summarizing research findings reported in the clinical literature (see Table 14), which pertain to the disutility of having certain comorbidities and transitioning into a terminal state before a follow-up appointment.

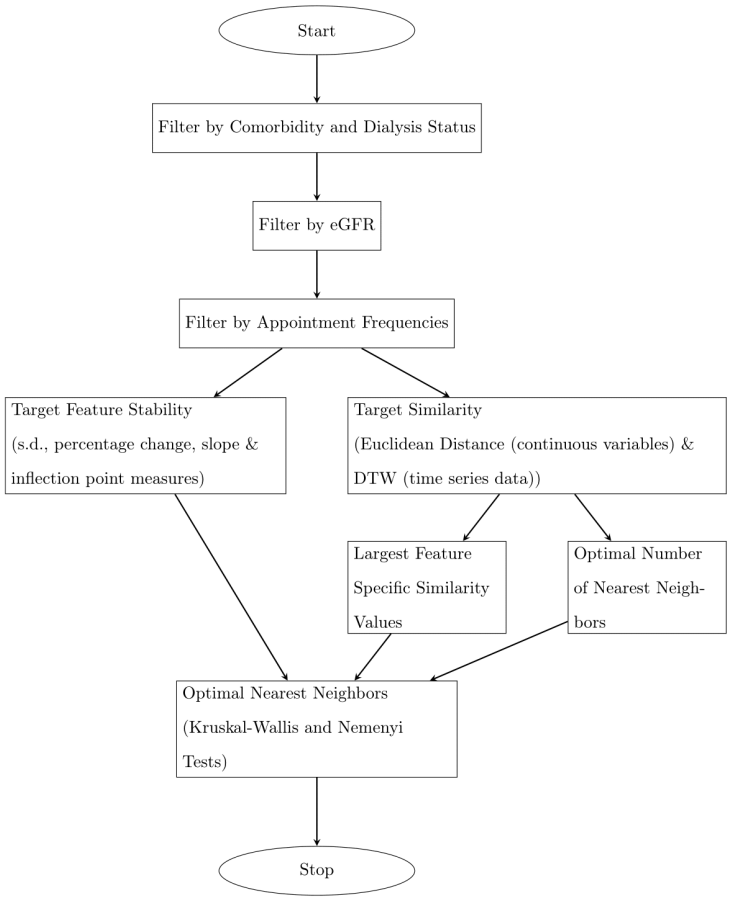


Figure 9: CBR Model

4.4.1.1 Case-based Reasoning Model

The first framework component is a case-based reasoning forecasting model. We choose this method because it can use intermittent, multivariate, and possibly correlated time series data. Furthermore, it allows for the development of a comprehensive model, which is easy to understand and implement by researchers and practitioners. To design a CBR model we create a database with 215,821 diverse patient cases. Each case contains relevant information on patient's disease progression. If we need to predict the disease progression of a new patient (target case), we compare him to the individuals (solution cases) saved in our database. We pull only similar and therefore relevant cases, which are then used to project the new individual's condition. The model hinges on the sensible assumption that if patients have similar disease histories they will experience comparable disease progressions.

Alagoz [3] have developed a disease specific CBR. In contrast, our model applies an analytic approach (see Fig. 9), which is sophisticated but easy to implement (Chapter 2), and avoids the use of disease specific selection measures because they will prevent us from directly applying our model to a new disease CBR database. It contributes to the CBR literature by developing a statistical method for extracting feature stability information. This is particularly relevant for medical CBRs, because while examining multiple patient features doctors tend to focus more on those exhibiting atypical variability. Furthermore, we propose a case-specific retrieval process, which optimizes the number of nearest neighbors used by a target. Our adaptive rule is important, because cases are not uniformly distributed in the feature space. Therefore, a patient with frequently observed characteristics will have a larger cohort of solution cases than a patient with unique features who is very different from cases in the database.

We use the CKD case-based reasoning database to forecast patients' progression given that they attended appointments within 1.5, 3, 6 and 12 months of their last visit. The forecasted results are aggregated and used to estimate our MDP transition probability matrix. We have stratified our transitional probability matrix across all states in S except for age. We have done that due to the limited number of patients in some age groups in S.

4.4.1.2 Survival Analysis Model

The second model used to parametrize our MDP is a standard Cox proportional hazard model, which can be easily replicated across different chronic conditions. It is designed to estimate the life expectancy of different patient groups before reaching a terminal state. Our MDP has two such states: death before dialysis and dialysis initiation. Therefore, we used a Cox proportional hazard model with competing risks. Furthermore, we estimate the life expectancy of dialysis patients, conditional on their age and coexisting comorbidities.

4.4.1.3 Appointment attendance disutility

The final dataset used to parametrize the MDP estimates the disutility of attending appointments ($disu_{appt}$) due to the time required to see the doctor and return from his office. We can only approximate this measure, by identifying the distance between the patient's and the clinic's ZIP codes. If such information is not available to researchers, they can analyze how sensitive their MDP results are to any $disu_{appt}$ changes. Sensitivity analysis is relatively easy because $disu_{appt} \in (0, 1]$, where 1 (0.5) corresponds to spending a whole (half a) day traveling to and from a clinic and vising a doctor. To address the worst-case scenario, we let $disu_{appt} = 1$.

4.4.2 Input Parameter Source

To finalize our model, we should fully estimate the immediate reward function ($r_i(s)$) and patients' total expected quality-adjusted life days in period N ($R_N(s)$). Therefore, we need to estimate some disutilities ($disu_s, disu_{s'}$) and probabilities of transitioning into terminal states ($p_{k,i}, p'_{k,i}, p_{6,i}$) (Table 14).

Our first set of disutilities is associated with patients' chronic kidney disease stages. Gorodetskaya [34] quantifies the quality of life among CKD patients for all disease stages, including CKD Stage 5 with and without dialysis. The authors administer three questionnaires to a diverse patient group, where individuals with advanced kidney disease (CKD Stages 4 and 5) are asked to respond to the questionnaires on multiple occasions. The second set of disutilities quantifies the negative impact of age and one or more comorbidities on patient's quality of life [110, 111]. The authors provide a detailed list of ICD9 and

CCC Code-based utility measures. All comorbidity estimates are obtained using a US-based national survey of noninstitutionalized individuals.

Using the parameter estimates from Table 14 and the disutility function listed in §4.3 ($disu_s = disu_d + disu_h + disu_v + disu_k + disu_{age} + disu_{\#(1+d+h+v)\ coexisting\ comorbidities}$) we can obtain the disutilities for all patient groups. For example, the disutility for a 65 year old patient with CKD Stage 3 and no comorbidites is $disu_s = 0+0+0+0.084+0.099+0 = 0.183$. Similarly, the disutility for a 65 years old patient with CKD Stage 3, diabetes and heart failure is $disu_s = 0.0351 + 0.0635 + 0 + 0.084 + 0.099 + 0.0876 = 0.3692$.

Table 14: Parameter Estimates

Parameter		Reference
Disutility	CKD	Gorodetskaya et al. (2005) [34]
	Comorbidities and Age	Sullivan & Ghushchyan (2006) [110]
		Sullivan et al., (2005) [111]
Time Until Death	Patients not on Dialysis	Gansevoort et al., (2013) [30]
		Turin et al. (2012) [115]
	Patients on Dialysis	Jassal et al., (2007) [48]
Time Until Dialysis	CKD Stage 3	O'Hare et al., (2012) [87]
	CKD Stage 4 & 5	Landray, M. (2010) [57]

We use the expected time until reaching death or dialysis to estimate the probability of transitioning into a terminal state in 1.5, 3, 6 and 12 months given patient's health status $(p_{k,i}, p'_{k,i}, p_{6,i})$. Landray [57] reports that the mean time until dialysis for patients in CKD stages 5 and 4 is approximately one and six years respectively. The mean time until dialysis for CKD Stage 3 is estimated using information provided by O'Hare [87]. They suggest that people experiencing slow disease progression observe a 7.7 $\pm$ 4.7 decrease in their eGFR every year. We make the conservative assumption that all CKD Stage 3 patients have eGFR = 60mL/min/1.73m^2 and will progress at a rate of 3 units per year. This means that on average it will take approximately eighteen years for a patient in CKD Stage 3 to go on

dialysis. Average life expectancy for patients across all CKD groups is reported in [30, 115]. Finally, the mean length of life for patients on dialysis is given by Jassal [48]. To calculate the probability of an event occurring given that a certain time has passed we assume that the time until terminal state is exponentially distributed with a mean equal to the estimates provided in the articles from Table 14.

4.5 NUMERICAL RESULTS

4.5.1 Empirical Results

Section 4.4 provided the inputs for our MDP. We now employ the MDP model to identify the optimal appointment strategies for each patient's subcategory over a pre-determined planning horizon. As noted earlier, the NICE guidelines are a special case of our MDP policy. Therefore, we compare our results with the guidelines suggested by NICE. We find that the data-informed decisions produced by the MDP outperform the NICE guidelines.

In all examined results, we set the time horizon to two years. This seemingly small value is chosen because the optimal policies across the examined cases are not sensitive to changes in N, when $N > 4$ (i.e. one year). Namely, the results across the cases below are almost identical when N is set to 40 (i.e. ten years). This observation is in agreement with the sensitivity analysis results in §4.5.2. We also select the two-years-time horizon setting for ease of illustration (§4.5.1.1 - 4.5.1.4).

4.5.1.1 Comorbidity Comparison We will first examine four common patient types, who have CKD Stage 3 with (1) no comorbidities, (2) diabetes, (3) heart failure and (4) diabetes and heart failure. All patients are 75 years old at the time the scheduling decisions are made. We focus on patients with CKD Stage 3 first because approximately 88.82% of the patient cases in our dataset fall into this category.

One would expect that individuals who belong to the first group would need the least amount of care. Similarly, patients who belong to the fourth group would require more

monitoring due to their disease complications. Still, if we follow the NICE guideline, which is solely based on one's kidney disease stage, we would conclude that patients from all four groups should be treated identically and seen on average once every six months, since all of them have CKD Stage 3.

Contrary to this generic policy, our model suggests that the sicker patients should be seen sooner than the healthier individuals. Thus, when the planning horizon is set to two years, our MDP suggests examining CKD Stage 3 patients with no comorbidities once a year, CKD Stage 3 patients with diabetes or heart failure once every six months, and CKD Stage 3 patients with both comorbidities once every three months. This policy is consistent across all age groups for the first three patient types, which suggests that their results are not age sensitive. However, younger patients (e.g. between 60 and 68 years old) with CKD Stage 3, heart failure and diabetes are scheduled to see a doctor every three months the first year and just once the following year. Therefore, the model recognizes that in more complex cases older patients require more care.

Table 15: Appointment Frequencies

Time periods	Ages 60 – 74	Ages 75 – 83	Ages 84 – 88
1 – 6 months	1	4	4
6 – 12 months	4	4	4
12 – 18 months	1	1	4
18 – 24 months	4	4	4

4.5.1.2 Age Comparison We examine the optimal strategy for patients in CKD Stage 5 with diabetes and heart failure under a 2-year time horizon. Similar to the fourth group above (i.e. Stage 3 with diabetes and heart failure), older individuals in CKD Stage 5 with diabetes and heart failure require more appointments. Our model suggests seeing patients ages 60 through 74 in six months. The appointment is followed by four equally-spaced visits (every 1.5 months) in the second half of the year. This pattern is repeated in the second

year (Table 15, column 1). Patients between the ages of 75 and 83 should be seen more frequently: eight times in the first year and five in the second year (Table 15, column 2). Finally, patients between the ages of 84 and 88 need to visit the clinic once every 1.5 months in both years (Table 15, column 3).

Unlike their CKD Stage 3 counterparts, younger CKD Stage 5 patients (see Table 15, column 1) with diabetes and heart failure are scheduled for more intense monitoring (four appointments in six months) after a period of less demanding follow-up care (a single appointment in six months). This schedule is based on the data used to parametrize the model. Thus, past observations influence our model to hedge against the possibility of observing a disease deterioration after a period of limited monitoring.

4.5.1.3 Sickest Patients Surprisingly, patients in CKD Stage 5 with all three comorbidities (diabetes, heart failure and vascular disease) achieve optimal quality adjusted life days following the least frequent appointment visits. This might seem counter intuitive. However, such patients are at a high risk of kidney failure and death. Therefore, our results flag the need for change in patient's treatment policy. Depending on the dialysis initiation practices followed by clinicians, most patients in this group would soon begin dialysis and will not need as many future monitoring appointments. Once on dialysis, chronic kidney disease patients will only attend nephrology clinic to obtain their medication, and to maintain their eligibility for kidney transplant. These needs can be accommodated by attending annual appointments because some of the tests required prior to kidney transplant are good only for a year; to keep their medical records current patients need to get them redone annually. Furthermore, the oldest individuals in this subcategory are expected to have a very short life span. Therefore, palliative care rather than frequent doctor's appointments might be more appropriate for them.

Table 16: Age-Averaged Life Quality Improvement Achieved by the MDP model

Diabetes	Heart Failure	Vascular Disease	CKD	mean (sd)
0	0	0	3	18.50% (1.87%)
1	0	0	3	14.24% (0.46%)
0	1	0	3	22.63% (1.09%)
1	1	0	3	20.71% (0.71%)
0	0	1	3	19.63% (1.84%)
1	0	1	3	21.70% (1.03%)
0	1	1	3	19.86% (1.73%)
1	1	1	3	39.35% (1.14%)
0	0	0	4	25.48% (1.75%)
1	0	0	4	26.39% (0.41%)
0	1	0	4	24.74% (2.43%)
1	1	0	4	36.53% (1.22%)
0	0	1	4	27.30% (1.43%)
1	0	1	4	30.75% (0.86%)
0	1	1	4	32.52% (1.83%)
1	1	1	4	42.32% (1.27%)
0	0	0	5	60.15% (0.48%)
1	0	0	5	49.33% (0.75%)
0	1	0	5	81.32% (0.26%)
1	1	0	5	14.58% (0.65%)
0	0	1	5	53.86% (0.28%)
1	0	1	5	64.28% (1.05%)
0	1	1	5	62.48% (0.49%)
1	1	1	5	46.14% (0.71%)

4.5.1.4 NICE Policies' Evaluation We estimate patients' QALDs following the NICE guidelines, and then compare the results with those of our MDP. We compute the difference in QALDs between adopting the NICE guideline and our policy. We summarized the results in Table 16 by averaging the difference for each disease-based state across all age groups. In all, there are 24 patient states defined by: diabetes, heart failure, vascular disease, and CKD Stage. This allows us to observe the age-averaged improvement in QALDs if patients are monitored following the proposed MDP model. We find that our model outperforms the NICE policy across the entire state space and the minimum average QALDs improvement is 14.24%.

The NICE guidelines suggest very frequent appointment visits for patients with advanced kidney disease. However, Linn [70] reports that frequent doctor appointments are associated with lower treatment satisfaction, and are directly linked to reduced quality of life [83]. This implies that overly demanding monitoring strategies could cause further deterioration in one's already poor quality of life. For CKD Stage 5 patients (the last 8 rows in Table 16), our model not only maintains proper care across all patient groups, but also avoids excessive clinic visits among patients. This explains why CKD Stage 5 patients can significantly improve their QALD by following our policy.

Patients groups whose NICE policy resembles the MDP policy include:

- CKD Stage 3 with diabetes
- CKD Stage 4 without comorbidities
- CKD Stage 5 with diabetes and heart disease

Following the NICE policy, the three patient groups will still experience a suboptimal quality of life over the planning horizon. This is because unlike the MDP model, the NICE policy is not as adaptive, and does not consider the impact a possible disease progression might have on patient's quality of life. Our results highlight the importance of making disease monitoring decisions, while taking patient's possible future disease progression into account.

4.5.2 Sensitivity Analysis

For ease of illustration, we allow for two transition outcomes: (1) remain in the same transient state, (2) transition to k=7 (death). To understand how sensitive the optimal policy is to changes in the immediate and terminal rewards, we consider three separate transient state settings:

(i) CKD Stage 3 with no comorbidities, i.e. s = (d, h,v,k,age) = (0,0,0,3,65): (0003,65),
(ii) CKD Stage 4 with no comorbidities: (0004, 65),
(iii) CKD Stage 5 without dialysis and no comorbidities: (0005, 65).

These states are sufficient to examine the impact terminal and immediate rewards have on the policy suggestions.

We consider four planning horizon alternatives: one, two, three and four years. These alternatives are sufficient to address the effect our planning horizon has on the proposed MDP model.

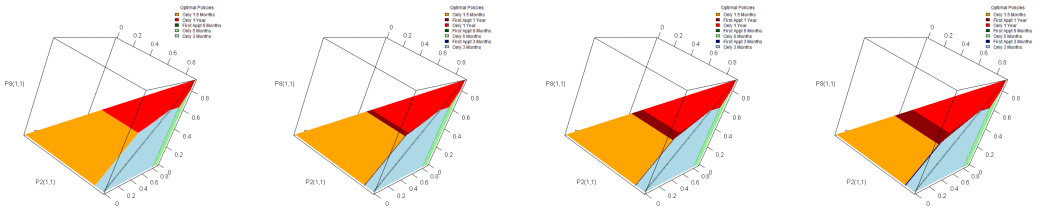
Table 17: Parameter Values

Parameter	Settings
Transient State	{(0003, 65), (0004, 65), (0005, 65)}
$p_{1_{(1,1)}}$	{0.1, 0.3, 0.5, 0.7, 0.9}
$p_{2_{(1,1)}}$	{0, ..., $p_{1_{(1,1)}} - 1/250$, $p_{1_{(1,1)}}$ }
$p_{4_{(1,1)}}$	{0, ..., $p_{2_{(1,1)}} - 1/250$, $p_{2_{(1,1)}}$ }
$p_{8_{(1,1)}}$	{0, ..., $p_{4_{(1,1)}} - 1/250$, $p_{4_{(1,1)}}$ }
N	{1 year, 2 years, 3 years, 4 years}

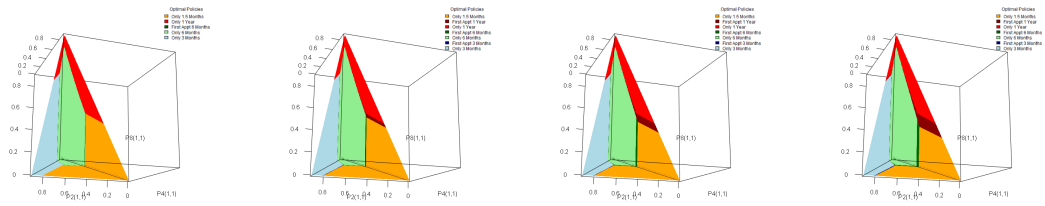
Our simplified model has two by two action dependent transition probability matrices, which are defined by a single parameter $p_{i_{(1,1)}}$ since $P_i = \begin{bmatrix} p_{i_{(1,1)}} & 1 - p_{i_{(1,1)}} \\ 0 & 1 \end{bmatrix}$. Recall $p_{i_{(s,m)}}$ is the probability of transitioning from state s to state m given that action i is chosen. To estimate the effect of the transition probability changes on patient's appointment policies,

we vary the values of $p_{i_{(1,1)}}$ between 0 and 1 at an increment of 1/250 across all actions. We examine all settings where $0 \leq p_{8_{(1,1)}} \leq p_{4_{(1,1)}} \leq p_{2_{(1,1)}} \leq p_{1_{(1,1)}} \leq 1$, which implies that when a patient sees his doctor frequently the chance of staying in the same state is relatively higher due to the short time interval between appointments.

The settings in Table 17 show that we consider a total of 38,902,500 ($3*[2,925$[1] $+73,150+333,375+908,600+1,923,825]*4$) model perturbations. They are presented in a series of graphs, as seen in Appendix C.



(a) N = 1 year (b) N = 2 years (c) N = 3 years (d) N = 4 years



(e) N = 1 year (f) N = 2 years (g) N = 3 years (h) N = 4 years

Figure 10: s = (0,0,0,3,65) & $p_{1_{(1,1)}} = 0.9$

We use Fig. 10 as an example to illustrate our results in Appendix C, where s = (0,0,0,3,65) and $p_{1_{(1,1)}} = 0.9$. Thus, $0 \leq p_{8_{(1,1)}} \leq p_{4_{(1,1)}} \leq p_{2_{(1,1)}} \leq 0.9$. Across all graphs, $p_{1_{(1,1)}} = 0.9$ indicates that the probability of staying in the same state given that the patient is seen in 1.5 months equals 90%. Each column (figure pair) corresponds to a different planning horizon (i.e. 1, 2, 3, 4 years). Each pair represent the same results viewed from different angles. The x-, y- and z-axis specify the values of $p_{2_{(1,1)}}, p_{4_{(1,1)}}, p_{8_{(1,1)}}$, i.e. the prob-

[1]There exist 25 unique $p_{2_{(1,1)}}$ values between 0 and 0.1 ($p_{1_{(1,1)}}$=0.1) and $\sum_{i=1}^{25}\sum_{j=1}^{i} j = 2,925$

ability of remaining in the same state given that the patient is seen in 3, 6, and 12 months, respectively. The orange, light blue, light green, and red optimal policy regions correspond to having equispaced appointments once every 1.5, 3, 6, and 12 months, respectively. The dark blue, dark green, and dark red areas signal that having irregularly-spaced appointments is optimal, where the first one is in 3, 6, and 12 months, respectively. We observe that the dark color areas expand with N and serve as separators between regions where equispaced appointments are optimal.

Each vertex of the pyramid corresponds to a different equispaced appointment policy. For example, the top vertex corresponds to $p_{1_{(1,1)}} = p_{2_{(1,1)}} = p_{4_{(1,1)}} = p_{8_{(1,1)}} = 0.9$. Therefore, all policies have the same probability of progression and the model picks action a_8, which is the least demanding monitoring strategy. The bottom left vertex of the pyramid in Fig. 10a corresponds to $p_{2_{(1,1)}} = p_{4_{(1,1)}} = p_{8_{(1,1)}} = 0$ and $p_{1_{(1,1)}} = 0.9$, while the bottom right is associated with $p_{4_{(1,1)}} = p_{8_{(1,1)}} = 0$ and $p_{2_{(1,1)}} = p_{1_{(1,1)}} = 0.9$. Therefore, our model selects policies a_1 and a_2 as optimal for the left and right vertexes, respectively. Thus, patients have a high probability of staying alive and their appointment frequencies are kept to a minimum. The last vertex of the pyramid corresponds to $p_{8_{(1,1)}} = 0$ and $p_{1_{(1,1)}} = p_{2_{(1,1)}} = p_{4_{(1,1)}} = 0.9$. Its optimal policy is to see the patient every six months (a_4). All these observations are in line with our findings that when policy alternatives produce similar results, the least demanding one will be chosen.

From the graphs in Appendix C, we conclude that the time horizon effect does not play a role in 90.07% of our simulated settings. This justifies our decision to report only results associated with a two-year time horizon in sections 4.5-4.6. Furthermore, fixed-interval appointment strategies are optimal in 91.42%, 92.82%, and 94.42% of the settings, where the transient state s equal to (0,0,0,3,65), (0,0,0,4,65) and (0,0,0,3,65), respectively. Therefore, fixed-interval schedules recommended by the MDP could be used in standard hospital settings.

Our sensitivity analysis results show that when all four action alternatives (a_1, a_2, a_4, and a_8) have compatible probabilities of deterioration the model will select a_8. Additionally, seeing a doctor every six months is preferred over other actions when a_1, a_2, and a_4 produce comparable results and have probabilities of health deterioration significantly smaller than

the one associated with a_8. Seeing a patient every three months is optimal when actions a_1 and a_2 are comparable and significantly better than a_4 and a_8. The most frequent appointment policy (a_1) dominates only when there is a significant benefit of seeing a doctor very often (i.e. $p_{1_{(1,1)}} >> p_{i_{(1,1)}}$ for i $= \{2, 4, 8\}$)). This suggests that the proposed MDP model will choose the least demanding policy, when there is a negligible difference in the probability of entering the terminal state between the least and the more frequent appointment alternatives.

When the probability of remaining in the same state is very low ($p_{1_{(1,1)}} = 0.1$ across all time horizons; or $p_{1_{(1,1)}} = 0.3$ when the time horizon is greater than 1 year), the model suggests seeing a patient once per year (Fig. 18, 27-29). Thus, the MDP reduces the disutility of attending appointments for patients near the end of their lives. This observation is consistent with the finding in §4.5.1.3, where our model schedules the sickest patients as infrequently as possible.

4.6 DEMAND PLANNING

Markov Decision Models have been around for decades. The steady improvement in computer technologies has allowed researchers to use MDPs and solve a variety of complex healthcare problems, which have provided meaningful recommendations on how to personalize treatments to improve patients' quality of life. Still, such patient-level results are not common in demand planning.

The relationship between workload and patient outcomes is well documented. Tarnow-Mordi [113] suggests that limiting the number of medical providers can adversely affect patients' outcomes. Hospitals deliver services which require a highly trained and expensive workforce. Therefore, management needs to plan for a capacity change in the early stages of a planning horizon to have sufficient time to recruit more staff. Furthermore, a better understanding of the patient population could help determine what type of new patients could be accepted at a clinic if its resources are underutilized. Patient recruitment decisions should be made to ensure that providers' idle time is minimized and patients' needs are met.

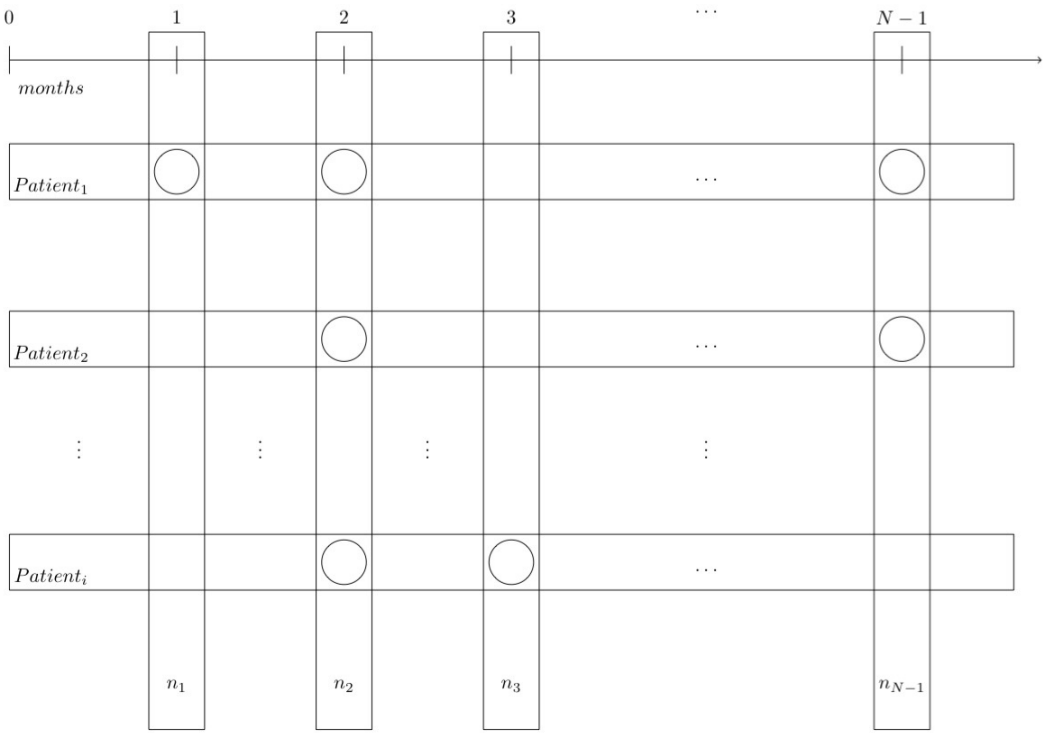


Figure 11: Aggregated Demand

We believe that leveraging the information obtained through a finite-planning-horizon MDP model can be instrumental in helping better anticipate demand for non-emergency care. This is important to ensure that the staff at a clinic can meet their patients' demand for services. Models for personalized care can help determine the optimal appointment schedule, which would maximize patients' quality of life. They can create an optimal visit plan over a prespecified time horizon, which would fit the needs of a patient in a specific health state. This patient-level information can be aggregated over the entire clinic population to estimate its expected demand for non-emergency care (Fig. 11). The aggregated demand represents the total number of patients who should be seen in 1.5 (n_1), 3 (n_2), 6 months from today (n_3), etc.

MDP results are used to guide the next decision maker's action given the current state of a system. We believe that using the remaining policy recommendations derived by an MDP could also be helpful, because even though such actions in the more distant future may not be optimal for each patient, when aggregated they can produce good overall demand-for-care

estimates. Our conjecture is in line with the Central Limit Theorem where given a large enough sample, policy discrepancies at the patient level will cancel each other out to produce satisfactory group approximations.

Because we would like to use the MDP to make capacity planning decisions we will focus on its performance when predicting demand for care 6 to 24 months into the future. Having estimates for the number of appointments needed that far in advance will allow management to decide if personnel recruitment is needed. Furthermore, this information will give the administrative staff enough time to hire one or more qualified clinicians.

We aim to show that our model will be appropriate for predicting the demand for care at a clinic. However, currently no clinic uses our MDP to make appointment decisions. Furthermore, we believe that the concern over malpractice lawsuits and poor service quality could result in patients being seen more frequently on average. Since our model does not consider such infrequent appointment deterrents, we would expect that its optimal policies will result in fewer appointments than the ones currently scheduled. Therefore, to evaluate how well our model predicts current practices we use its most frequent appointment visits suggested for each kidney state. We name this our conservative-MDP policy. Based on the results in §4.5, the conservative-MDP policy correspond to seeing a patient with CKD Stage 3, 4, and 5 once every 3, 6, and 1.5 months, respectively, in months 6 through 24 when the planning horizon is set to two years.

Following the suggestion made by our medical collaborator, we obtain patients' aggregate demand for care by recording appointments with doctors, who belong to one of seven specialties: (1) nephrology, (2) cardiology, (3) endocrinology, (4) hematology, (5) pulmonary, (6) primary care or (7) general internal medicine. Providers from these specialties are considered because they could prescribe treatments, which have a meaningful impact on the disease progression of a CKD patient.

To examine how well our MDP model predicts the demand for care given a patient population at a clinic, we obtain information on all chronic kidney disease patients treated within 5 diverse VA regions, who have CKD Stages 3, 4 or 5 (not on dialysis) as of January 1, 2009. The regions are selected because of their diverse CKD populations and geographical locations. We record each patient's first eGFR post January 1, 2009. We save the number

of appointments each patient had 6 through 12, 12 through 18 and 18 through 24 months after the baseline eGFR. The appointments we consider belong to any of the seven categories listed above. Finally, we aggregate the patient level information to get the total number of actual unique days with scheduled appointments between months 6 and 12, 12 and 18 and 18 and 24 past the eGFR baseline date.

Table 18: Aggregated Demand

Region	Time Period	Actual	Conservative-MDP	NICE
N. Florida & S. Georgia	6-12 months	13649	14087	8238
	12-18 months	13301	14087	8238
	18-24 months	13352	14087	8238
Central Florida	6-12 months	14957	13652	7897
	12-18 months	15360	13652	7897
	18-24 months	13721	13652	7897
S. Florida	6-12 months	5404	4872	2990
	12-18 months	5227	4872	2990
	18-24 months	5357	4872	2990
Texas	6-12 months	3389	2890	1714
	12-18 months	3281	2890	1714
	18-24 months	3191	2890	1714
Pennsylvania	6-12 months	7246	7365	4518
	12-18 months	6815	7365	4518
	18-24 months	6827	7365	4518

We estimate the number of appointments days using the NICE and our conservative-MDP policy suggestions. Accordingly, we multiple the number of patients who belong to CKD Stage 3, 4, and 5 at baseline by the kidney state dependent appointment frequencies suggested by each policy for periods 6 through 12, 12 through 18 and 18 through 24. For example, if there are one thousand CKD Stage 3 patients the conservative-MDP policy will

estimate that approximately 2∗1000 = 2000 appointments will be scheduled in each of the three time periods. At the same time, the NICE policy will estimate that there will be approximately 1∗1000=1000 appointments.

We observe that the actual demand for care across all three periods in all regions is relatively stable (Table 18). Both estimation methods (conservative-MDP and NICE) predict stable aggregate demand across all time periods. We conclude that our model provides good approximations for the demand of care, which represents a significant improvement over the NICE demand-for-care estimation strategy.

Our results show that appointment frequency strategies produced by MDPs can be useful when predicting the medium-term demand for care, conditional on knowing or being able to estimate clients' current health status. Therefore, such models would be good at approximating the demand for non-emergency appointments at clinics where patients are repeat customers.

4.7 CONCLUSION

Chronic disease monitoring strategies involve understanding patient's overall health state. In this research, we design a framework for optimizing such strategies, which would help improve the quality of life of a patient with a chronic condition. We showcase our framework using chronic kidney disease, which is selected due to its complexities, prevalence and cost of care. To develop a reliable model, we parameterize our MDP with the help of a supercomputer and using information on 68,514 patients treated for chronic kidney disease throughout the US. We show the benefits of following our framework, when monitoring CKD patients, by comparing it to a standard CKD population strategy, which does not account for patient's age and coexisting comorbidities.

This is the first chronic disease appointment optimization model, which is set up as a Markov Decision Process. It suggests that patients with the same chronic disease but different underlying complications require different care intensity. A similar observation was made by Gijsen [33], who stated that in past studies comorbidities were "consistently related

to health care utilization". This explains the results reported by Woodard [120] who note that the quality of care for more complex diabetes patients exceeds the one provided to individuals with fewer complications.

Furthermore, age impacts optimal policies primarily among patients with a more complex health status. Our model highlights that older individuals who are not close to kidney failure should be monitored more frequently than younger patients. These results confirm observations made in O'Hare [88], who suggest that for patients in CKD Stages 3 through 5 age plays a major role in disease progression and should be considered when designing CKD management strategies.

As noted earlier for patients near kidney failure and/or death the model suggests fewer monitoring appointments. This counterintuitive policy is sensible for individuals in CKD Stage 5 with multiple comorbidities because they are often faced with deciding between conservative (palliative) and active (dialysis) CKD treatment choices [24]. If the conservative treatment is chosen, regular monitoring appointments will be largely substituted with interactions with a palliative care specialist. If dialysis is initiated, the patient will begin attending a dialysis clinic and will leave our decision process.

Additionally, this is the first finite horizon discrete state MDP model, set from patient's perspective, whose goal is to maximize patient's total quality-adjusted life days and used for estimating demand for care. We demonstrate how the proposed MDP model can be used to predict demand for medium-term appointments, which is important when making staff hiring and patient recruitment decisions. We show that our model outperforms the NICE strategy and provides estimates which can be useful for clinic management and capacity planning. This is particularly important because of the decline in medical graduates with a nephrology subspecialty and the increase in the demand for the care they provide [90].

Our results have their limitations. The chronic disease used to showcase our framework produces results with a high medium-term demand-for-care accuracy and reasonable scheduling policies. Still, our model is developed using VA data which is primarily comprised of information on white male patients. This is the reason why gender and race were not considered in the state space of our MDP model. Therefore, our results should be used with caution when optimizing the appointment frequencies of patient types underrepresented in

our data set. This is particularly important for African Americans with CKD because they often experience faster disease progression [41].

Furthermore, our model does not suggest the type of appointments that should be scheduled in the future. This leads to an idea for future research, which will not only specify the optimal timing, but also the type of follow-up appointments given patients' current health state. Currently, we do not differentiate between appointments with different specialists, which could have an impact on patients' disease progression. Still, stratifying across doctors would make our MDP more realistic. This improvement could be achieved by updating the MDP action space to account for all appointment timing and doctor specialties we would like to consider.

We believe our work provides an important insight into optimizing the monitoring strategies for patients with chronic kidney disease. Still, there are several research directions we have not explored. First, we do not incorporate cost into our appointment optimization problem, which would be important when discussing patients without full healthcare coverage or designing policy recommendations from the societal perspective. Second, we do not examine the changes in appointment policies when patients are assumed to be risk averse rather than risk neutral. Third, we do not allow the timing between appointments to vary outside of our four action alternatives. Still, it will be interesting to study policies where it is possible for a patient to be seen as early as the next day.

5.0 PATIENT CHARACTERISTICS AND RESOURCE ALLOCATION FOR INDIVIDUALS WITH CHRONIC KIDNEY DISEASE.

5.1 INTRODUCTION

The median prevalence of chronic kidney disease (CKD) in the United States is approximately 7.2%, for people above the age of 30, and between 23.4% and 35.8% in people above 64[123]. As noted in a paper on CKD costs in the Medicare population [39] "the 2010 USRDS report shows that Medicare spent 29 billion dollars in 2009, or almost 6% of the annual Medicare budget, for people with ESRD" (end stage renal disease) [12]. The high demand for CKD care motivated us to investigate, empirically, using administrative data, whether patient characteristics are associated with the level of appointment utilizations. We use an approach similar to the one reported in Seligson et al. [104]. Our objective is to find patient clusters with greater numbers of outpatient appointments and hospitalizations, because such information might be of administrative use to improve patient prioritization policies.

5.2 METHODS

Patient Data

All models were developed using de-identified, administrative data, extracted from the Corporate Data Warehouse of the U.S. Department of Veterans Affairs. The extracted data set includes information on 29,735 CKD patients, treated at eleven VA facilities between January 1, 2009 and February 21, 2016. The data set includes patients' (1) comorbidities, CKD state and dialysis status, (2) lab values, (3) vital signs, and (4) appointment information.

We tracked four comorbidities: diabetes, heart failure, peripheral and/or cardiovascular disease (PVD/CVD), and cirrhosis. Our dataset contains information on individuals' CKD state, whether they are on dialysis, and whether they have been diagnosed with any of the four comorbidities of interest, as well as the time between the first recorded diagnosis, or treatment initiation, and death or the study end date. We track patients' diabetes status because it has been identified as the primary cause in 44% of kidney failures [26]. Patients' CKD states represent individuals' last explicitly specified CKD states. If an individual has an unspecified CKD state, we record it as unspecified. We followed both heart comorbidities, because CKD has been reported as an independent factor for cardiovascular disease outcomes [99]. We examine patients' cirrhosis status because it has been identified as a cause of inaccurate kidney filtration estimates [106].

We record four laboratory values, and four vital signs. Lab values consist of patients' estimated glomerular filtration rate (eGFR), albumin, potassium, and phosphate. Patients' vital signs include systolic (SBP) and diastolic blood pressure (DBP), pulse pressure, and weight. All characteristics, excluding phosphate, are summarized using five standard summary statistics: mean, median, standard deviation, minimum, and maximum. Approximately 40% of all patients have, at most, a single phosphate value on record. Therefore, for that lab value, we created a binary variable, which shows whether a patient has multiple phosphate lab measures on record. We examine patients' eGFR, because it is accepted as the best single measure of kidney function [108]. Low albumin could indicate patient proteinuria, inflammation, or inadequate nutrition. Elevated phosphate or potassium levels could result in a variety of heart complications [69]. Favorable values for blood and pulse pressure are indicative of good heart health [7]. Weight is used as a proxy for water retention.

We record the percentage of days a patient spends (1) hospitalized, (2) attending outpatient appointments, or (3) both. To obtain any of the three percentage values, we first calculate the number of days for which we have information on the patient, i.e., the distance between the first and last days with any records of appointments, comorbidity diagnoses, or laboratory results. We use the length of that interval to divide into the number of unique days a patient has (1) attended outpatient appointments, (2) been hospitalized, and (3) either been hospitalized or attended outpatient appointments.

Study Design

We built patient clusters for each VA facility, in order to compare and contrast hospitals located in different parts of the US. Each model was obtained using patients' laboratory value, vital sign, comorbidity, dialysis, and CKD status data. No appointment information data was used in the unsupervised clustering procedures, in order to test the hypothesis that groups determined by health status metrics translate into groups with significantly different resource utilization characteristics.

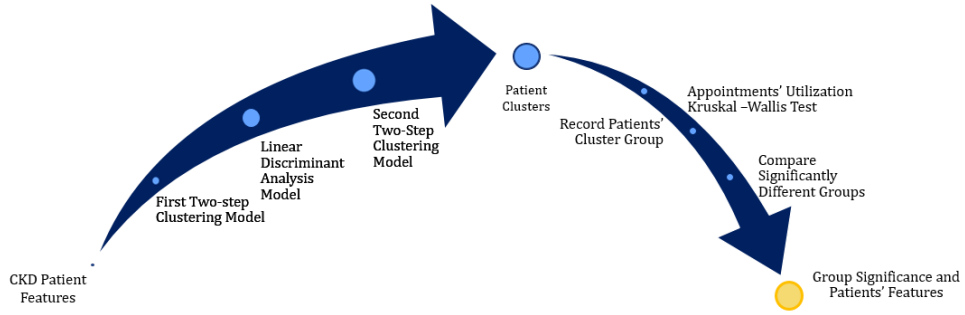


Figure 12: Analysis Graph

We used unsupervised clustering in order to avoid biases, such as that of selecting the number of possible patients' groups. We applied a Two-Step cluster analysis method [15]. The optimal cluster number was first picked with the help of information criteria (Akaike's Information Criterion (AIC) [2] or the Bayesian Information Criterion (BIC)[103]), which compared models with different numbers of clusters against one another by examining the trade-off between models' goodness-of-fit and their complexity. The unsupervised clustering method does not highlight the predictors that have a significant impact on patients' groupings. Therefore, in a second methodological step, we utilized a linear discriminant analysis (LDA) model [53] to determine the health characteristics that played a significant role in the cluster selection procedures. Once important predictors were flagged, they were used to re-calibrate the clusters, again using a Two-Step unsupervised clustering routine. This three step procedure, which involved (1) clustering, (2) significant predictor detection, and (3) re-clustering, was independently applied to all eleven VA facilities (Fig. 12).

Once patients were grouped, we used Kruskal-Wallis nonparametric tests to compare their clusters, based on the three appointment measures discussed above. We used the LDA predictors to compare and contrast all inter- and intra-facility groups with significantly different appointment utilizations (Fig. 12).

5.3 RESULTS

Overall, 32.46% of the patients in the data set had been diagnosed with diabetes; values at facilities ranged from 27.50% to 43.13%. 6.67% of the patients in the data set were on dialysis. Depending on the facility, this number was as small as 4.21% and as large as 10.46%. While only 11.41% of the people had been diagnosed with heart failure, 44.01% had some form of PVD/CVD. The facility located in Massachusetts had the highest incidence rate of both heart-related comorbidities, 19.36% and 59.10%, respectively. With an incidence rate of 2.48%, cirrhosis was the least frequently observed comorbidity. Most patients were in CKD Stage 3 (39.38%), followed by CKD Stage unspecified (34.10%), Stage 5 (ESRD) (8.78%), Stage 4 (8.29%), Stage 2 (6.56%), and Stage 1 (1.96%) (Table 19). A small percentage of the patients (0.92%) did not have recorded CKD-related ICD9/ICD10 codes, and were not assigned to any of the CKD groups. Across all facilities, the majority of patients who had not been diagnosed with heart failure also had not been diagnosed with diabetes, regardless of their PVD/CVD status. In all but one of the California hospitals, more patients had been diagnosed with diabetes if they had been diagnosed with at least one of the two heart related comorbidities.

The mean lab values, and the mean vital sign descriptors, varied significantly across the VA facilities (p-value < 0.0001) (Table 20). For example, two Florida facilities, and one Texas facility, had significantly lower average albumin values, which could mean that more patients at those facilities suffered from proteinuria, inflammation, or malnutrition. One of the four Florida hospitals, and a hospital located in Massachusetts, serviced patients with significantly lower mean weights, as compared to the patients in all other facilities. Similarly, a facility located in southern Florida had a patient population with significantly

higher eGFR values, which may indicate that, on average, its patients had healthier kidneys than did individuals from the other ten VA facilities.

Table 19: Population Categorical Descriptors

	Classifiers	Overall	Min	Max
Comorbidities & dialysis status	PVD/CVD	44.01%	34.56%	59.10%
	diabetes	32.46%	27.50%	43.14%
	heart failure	11.62%	6.07%	19.36%
	dialysis	6.67%	4.21%	10.46%
	cirrhosis	2.48%	1.71%	3.91%
Chronic Kidney Disease Stage	3	39.38%	22.03%	46.77%
	Unspecified	34.10%	21.51%	49.61%
	5	8.78%	5.69%	13.19%
	4	8.29%	6.50%	10.64%
	2	6.56%	2.51%	14.30%
	1	1.96%	0.67%	3.35%

We applied our three step clustering method to the data from each facility. Due to missing data, between 9.13% and 0.84% of the facility-specific patient data was ignored in the cluster selection procedure. In all but one facility, using either the AIC or the BIC information criterion resulted in the selection of the same number of clusters. The number of clusters identified for each facility varied from two to six, and used between eleven and twenty significant LDA variables (Table 21). There were a total of forty-four unique predictors used across all facility models. The predictors that were selected in at least fifty percent of the models (i.e., six or more times) included the indicator variable for multiple phosphate measures, a diagnosis of PVD/CVD, diabetes and dialysis, time since onset (TSO) of PVD/CVD and dialysis, minimum albumin and eGFR values, and standard deviation of the potassium and SBP values.

The facilities for which patients were grouped into only two clusters had smaller patient counts. They contained the second, third, and fifth smallest CKD patient populations. Hospitals with patient populations clustered into three groups had the first, second, and eighth largest CKD patient populations, among the eleven facilities examined (Table 21).

Table 20: Population Mean Laboratory Values and Vital Signs

Mean Value	Overall	Min	Max
Albumin	3.86	3.68	4.05
eGFR	45.54	44.43	47.75
Phosphate	3.57	3.48	3.64
Potassium	4.42	4.35	4.46
SBP	134.06	133.24	135.19
DBP	72.44	71.39	74.03
Weight	200.50	196.00	204.59
Pulse Pressure	61.74	60.62	62.70

We checked to see if there was a significant difference among patient groups across all three appointment utilization metrics (Table 22) discussed in the patient data section above. The only hospital with no significant difference across patient clusters for any utilization measures was the one located in southern Florida (FL #2). As noted above, the patient population at that facility appeared to be healthier than the patient populations at the other facilities. The facility located in Texas was the only one that had only two significant utilization measures. The outpatient appointments may have been similar across Texas' clusters because of the relatively low median outpatient appointment utilizations (significantly lower than the utilizations at seven of the facilities). Finally, CA #1 patient clusters differed only in their hospitalization utilization, perhaps due to the larger variability in outpatient appointment utilizations across patients from both CA #1 clusters.

Table 23 shows a summary of all health-status characteristics whose rank ordering was consistent with the frequency ordering of the significantly different patient groups across all three utilization metrics in the eight facilities where such significant differences existed. For example, across all three utilization measures, MA group 1 required significantly less outpatient and inpatient care than did MA group 2. As noted in Table 23, the patients' diabetes state predictor was aligned with that utilization difference, because none of the patients in group 1 had diabetes and approximately 76.61% of the individuals in group 2 had this comorbidity. Thus, the greater incidence rate of diabetes was associated with increased utilization.

Table 21: Facility Cluster Results

Hospital	Clusters (AIC/BIC)	Population	LDA Predictors	Missing Information (%)
NV	6/3	1804	17/16	2.66%
PA	4	2545	16	3.03%
FL #1	4	4516	20	0.84%
TX	4	1026	16	2.24%
CA #2	4	3400	20	7.47%
NY	3	4706	20	8.12%
FL #3	3	5156	17	9.13%
FL #4	3	1683	20	1.84%
MA	2	1560	19	4.17%
FL #2	2	1712	16	1.46%
CA #1	2	1627	11	2.34%

The majority of the health-status predictors in Table 23 were comorbidity based. The predictors most often used to differentiate among clusters were whether a patient was diagnosed with diabetes, and PVD/CVD, and the time since the onset of dialysis. For example, the lower utilization groups (1 and 3) in New York had almost no diabetes patients, while all the more frequent users in group 2 were diagnosed with diabetes. Similarly, less frequent

users in the FL #3 facility (groups 1 and 2) had almost no cases of PVD/CVD, while frequent users (group 3) were almost all diagnosed with this comorbidity. A greater contrast occurs in the data from the facility in Nevada, where high utilization users (1 and 6) had PVD/CVD, while none of the patients in the low appointment utilization groups, 2 through 5, had the same comorbidity.

Table 22: Significant Difference in Appointment Utilizations

Hospital	3 Measures	2 Measures	1 Measure	0 Measures
MA	X			
NY	X			
PA	X			
FL #1	X			
FL #3	X			
FL #4	X			
NV	X			
CA #2	X			
TX		X		
CA #1			X	
FL #2				X

Only two facilities, MA and FL #4, had their clusters differentiated by multiple laboratory value predictors. The minimum potassium value (median DBP) had a significantly greater variability in MA than in five (seven) of the examined hospitals. The second facility, FL #4, had its clusters separated using three standard deviation-based predictors. The potassium (weight) standard deviation-based predictors had significantly lower mean values in FL #4 than in eight (seven) of the other ten facilities.

Table 23: LDA predictors matching K-W cluster appointment ranks

Predictor	NY	PA	MA	FL #1	FL #3	FL #4	NV	CA #2
PVD/CVD State	1		1		1	1	1	1
PVD/CVD TSO					1	1		1
Diabetes State	1	1	1			1	1	1
Diabetes TSO	1		1					1
Heart Failure State		1	1		1			1
Heart Failure TSO	1		1			1	1	
Cirrhosis State	1		1					1
Cirrhosis TSO	1		1					
Dialysis state	1		1		1	1		1
Dialysis TSO	1		1	1		1	1	1
Phosphate - multiple			1					
Potassium sd						1		
min potassium			1					
median pulse pressure			1			1		
Pulse pressure sd						1		
mean pulse pressure							1	
Weight sd						1		
median DBP			1					

5.4 DISCUSSION

The primary goal of this study was to provide new CKD monitoring suggestions, by teasing out patients' characteristics that were associated with more frequent outpatient appointments and hospitalizations. To do that, we compared and contrasted regions based

on characteristics that significantly contribute towards individuals' appointment utilization differences. PVD/CVD and diabetes states, and the time since dialysis onset were the three characteristics whose mean group values had orderings most consistent with the frequencies of outpatient appointments and hospitalizations (Table 22). Those three characteristics were selected in six regions; heart failure and the time since the onset of heart failure were selected in four.

There exists a plethora of work on developing risk stratification models, whose main goal is to highlight patient groups with more complex and, therefore, more costly health needs. Such models include the hierarchical condition categories (HCC), adjusted clinical groups (ACGs), elder risk assessment, chronic comorbidity count (CCC), Carlson comorbidity index, and Minnesota tiering [36]. The primary commonality of those risk adjustment models is that all of them contain some patient comorbidity information. Still, the way comorbidity information is incorporated into the models differs. Some measures count the number of chronic conditions that fit certain characteristics (CCC and Minnesota tiering). Others group them based on whether they are stable or unstable (ACGs).

Our observations suggest that policymakers should focus on the type of comorbidities, not only on their number. We believe that they should pay attention first to whether patients have chronic peripheral and/or cardiovascular disease and diabetes, and then on whether they have been diagnosed with heart failure and cirrhosis, because there was a clear relationship between the differences among group resource utilizations and the prevalence of those comorbidities among patients. Focusing on diabetes is particularly important, because between 2000 and 2008, the percentage of Veterans with renal failure and diabetes increased by 197% and 34%, respectively [122]. Therefore, failing to improve the treatment of renal failure patients with diabetes could result in a significant increase in resource demand. A possible way of addressing that problem might be to develop a policy aimed at preventing the onset of diabetes and PVD/CVD in CKD veterans.

While preventing the onset of comorbidities could be difficult to achieve, a more doable goal might involve the creation or the improvement of existing comorbidity-management rules for individuals with chronic kidney disease. An example of the potential associated with an update in CKD policies was reported by Kaiser Permanente Hawaii, where the CKD state

of patients referred to a nephrologist varied significantly [59]. A risk stratification model, and improved communication channels between generalists and nephrologists, addressed the variability problem, and resulted in a significant reduction of late nephrologist referrals, which substantially increased the number of patients with functioning arteriovenous fistulas at dialysis initiation. Similarly, Schmidt et al. [102] argue that, even though long term survival rates in early and late referrals are not significantly different, the distress caused by the unexpected diagnosis and dialysis initiation could be very emotionally and physically demanding, which makes early referrals preferable.

While many papers on early referrals of CKD patients focus on the importance of nephrologist care, our results highlight the importance of referring patients to cardiologists and endocrinologists, who are expected to provide better care for patients with vascular and diabetes complications. A rise in such referrals might result in an increase in demand for more specialized outpatient appointments, but may help prevent hospitalizations caused by sub-optimal disease management.

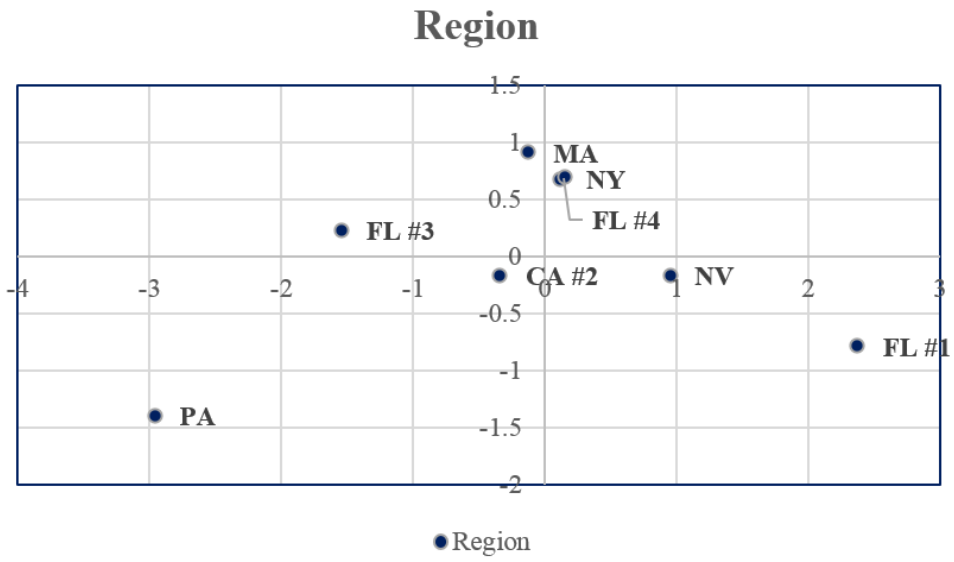


Figure 13: Regions Stimulus Coordinates

Unfortunately, our data set contains information only about Veterans with chronic kidney disease. Because the population of Veterans is predominantly male, and older than the

general population, it is possible that our findings may not apply as they stand to the entire US CKD population. Still, our results and methodology could be used to determine whether patients' laboratory results for the general population could be used to stratify individuals into distinct groups, to see if patient clusters have different resource utilization rates.

Furthermore, we observed geographical variability in clusters and clustering predictors, which might be partially explained by the regional variability of patient populations, because dissimilarities are driven by complex, and possibly interacting, factors. Note that close geographical proximity (FL #1 and FL #4) does not necessarily imply similarity (Fig. 13).

The degree of medical professionals' adherence to guidelines is an attractive attribute, which we have not measured in this study, and which might further explain the differences across facilities. This would suggest the need for further examination of whether hospitalizations across facilities are significantly different after controlling for patients' characteristics, as well as practitioners' adherence to guidelines.

In this study, we showed that, regardless of the facility's geographical location, patients' comorbidity states are usually the primary contributors towards individuals' appointment utilizations. The two comorbidities with the greatest regional selection rates are PCV/CVD and diabetes, followed by heart failure and cirrhosis. Those selection frequencies highlight the need for the improvement of chronic kidney disease referral policies, not only between generalists and nephrologists, but also between generalists and both cardiologists and endocrinologists. The data appear to suggest that more specialized care might be beneficial in reducing the number of hospitalizations among CKD patients with PVD/CVD and diabetes comorbidities.

6.0 SUMMARY AND FUTURE WORK

To improve the care they provide, doctors should personalize appointment frequencies, based on patient's health status, disease progression and appointments' impact. Therefore, accurate disease trajectory and life expectancy forecasts are crucial when managing a chronic disease. In my dissertation, I present novel methods used to quantify the effect of appointment frequencies given patient's health status. The models suggest optimal and personalized follow-up appointment strategies, which can help doctors select treatment policies for the chronically ill. The models can also estimate the expected aggregate demand for care at a chronic disease clinic in the near future. Such predictions can be utilized by management when making employee or patient recruitment decisions.

The optimization model confirms some generally accepted treatment notions; for example, older patients and individuals with one or multiple comorbidities require more intense care. Additionally, our results show that using a myopic policy for assigning follow-up appointments is detrimental to patient's quality of life. Such a policy fails to account for the patient's expected disease trajectories and disutility of attending appointments, which is especially pronounced among older and sicker individuals who live in rural regions. Furthermore, the models use readily available past appointment utilization data to flag patients in need for palliative care preparation. This is particularly important because current prioritization policies are based on subjective rather than data-driven rules.

Our disease monitoring framework has been applied to patients with moderate to severe chronic kidney disease. All models in the framework are parametrized using data on a large geographically diverse population of U.S. Veterans. The close nature of the VA healthcare system allows us to obtain thorough records of our patients. Still, our data source has its limitations, the main one being the lack of gender, age and race diversity among Veterans,

who are primarily older white males. Therefore, our results should be used with caution when selecting treatment options for patients who fall into any demographic category underrepresented in our patient population.

The models do not consider the quality of life implications associated with appointments' financial burden, which could be significant if a patient is a working adult with no access to free health care. Our framework does not incorporate cost data because most VA patients are retired and have service-based health coverage, therefore their disutility associated with seeing a doctor is only measured by the time spend on going to, attending and returning from an appointment.

The work presented in this dissertation provides valuable insight into the impact appointments have on one's disease trajectory. Our results suggest several avenues for future research. Our current framework offers recommendations on when a chronic kidney disease patient should be seen next. Still, our models do not specify the specialty of the medical professional/s the patient needs to see. Such a model, which provides the specialty of the doctor/s, is not present in the current literature. To design it we should estimate the effect of appointment type and frequency on patient's disease progression. This information will help optimally allocate limited healthcare provider resources and suggest backup appointment strategies if waiting times for a doctor from a recommended specialty are too long. Additionally, our current model assumes that a patient follows all recommendations made by the doctors. This assumption is often unrealistic and could also be relaxed. Lastly, our model can be updated to incorporate cost of care estimates, which would be useful when optimizing monitoring strategies for a chronic condition from the population prospective.

APPENDIX A

TWO STEP MODEL

We fitted all thirty-eight models to the training set in order to predict whether a case had more accurate forecasts when 1-NN, unequally weighted or equally weighted multiple neighbors were used. The model with the most accurately predicted and balanced results was the logistic regression model with a forward likelihood ratio selection (Table 24a-24b). It was accurate in 245 of the 500 test set cases (49%). It predicted correctly 117 of the unequally weighted, 43 of the equally weighted and 85 of the 1-NN cases. The model used a small fraction of the predictors considered in the analysis (Table 25a-25b).

Three hundred and three of the training cases were correctly classified as having a preference for the non-equally weighted forecasting approach. They were used to develop a model for predicting cases with preference for unequally weighted nearest neighbors, where the options were 2n-, 3n- and 4n-NN. We picked a multinomial logistic regression model with filtered forward likelihood ratio selection because of its performance on the test and training sets and its ability to predict cases across all three groups (Table 26a-26b). The filtering was done to remove predictors with almost identical values, which would cause numerical instability if used as independent variables. Similarly, to the first step model, over fifty percent of the predictors used were dissimilarity based (Table 27).

One hundred and thirty-eight training cases were correctly classified as having a preference for the equally weighted multiple nearest neighbors' techniques. Using the selection method applied when determining non-equally weighted nearest neighbor preferences, we predicted which cases fell into the 2e-, 3e-, and 4e-NN categories. A C5.0 model with

ten minimum numbers of records per child branch proved the best (Table 28a-28b). It had among the best total correct prediction percentages and achieved a good balance in correctly classifying cases across all three preference groups.

Table 24a: Step 1 Model Selection

Model	Weighted (0)	Non-Weighted (1)	1-NN	Total Percentage
SVM Poly=3 C=1	91	55	59	41.00
SVM Poly=3 C=2	91	55	59	41.00
SVM Poly=3 C=3	91	55	59	41.00
SVM Poly=3 C=4	91	55	59	41.00
SVM Poly=3 C=5	91	55	59	41.00
SVM Poly=3 C=6	91	55	59	41.00
SVM Poly=3 C=7	91	55	59	41.00
SVM Poly=3 C=8	91	55	59	41.00
SVM Poly=3 C=9	91	55	59	41.00
SVM Poly=3 C=10	91	55	59	41.00
C5 M=2	140	42	16	39.60
C5 M=5	45	39	121	41.00
C5 M=10	36	42	108	37.20
C5 M=15	40	9	143	38.40
C5 M=20	40	4	150	38.80

Table 24b: Step 1 Model Selection

Model	Weighted (0)	Non-Weighted (1)	1-NN	Total Percentage
Logit Forward LR	117	43	85	49.00
Logit Backward LR	105	56	39	40.00
LDA	108	52	68	45.60
SVM RBF C=1	105	41	51	39.40
SVM RBF C=2	100	40	55	39.00
SVM RBF C=3	102	40	56	39.60
SVM RBF C=4	98	43	58	39.80
SVM RBF C=5	98	42	61	40.20
SVM RBF C=6	89	43	58	38.00
SVM RBF C=7	92	43	57	38.40
SVM RBF C=8	91	44	57	38.40
SVM RBF C=9	91	44	56	38.20
SVM RBF C=10	91	43	55	37.80
SVM Poly=2 C=1	85	49	56	38.00
SVM Poly=2 C=2	86	51	60	39.40
SVM Poly=2 C=3	86	56	60	40.40
SVM Poly=2 C=4	86	56	57	39.80
SVM Poly=2 C=5	89	53	56	39.60
SVM Poly=2 C=6	92	56	58	41.20
SVM Poly=2 C=7	89	56	57	40.40
SVM Poly=2 C=8	91	57	58	41.20
SVM Poly=2 C=9	89	55	56	40.00
SVM Poly=2 C=10	89	56	56	40.20

Table 25a: Multinomial Logistic Regression Model, Step 1

Group	Predictor	Exp(B)	p-value
non-equal	Intercept		0.531
weights	[No Comorbidities=0]	1.434	0.01
	[No Comorbidities=1]		
	(eGFR peak 2-NN averaged dissimilarities)^2	0.682	0.296
	(Albumin 2-NN averaged dissimilarities)^2	4.254	0.012
	(DBP 4-NN averaged dissimilarities)^2	0.297	0.002
	(eGFR change)^2	0.623	0.028
	(max Phosphate)^2	1.012	0.062
	std eGFR peak dissimilarity diff b/w 1 & 2 NN	1.289	0.002
	std Phosphate dissimilarity diff b/w 1 & 2 NN	1.117	0.153
	std DBP dissimilarity diff b/w 1 & 3 NN	1.333	0
	std eGFR smooth t.s. dissimilarity diff b/w 1 & 4 NN	1.427	0
	std Albumin dissimilarity diff b/w 1 & 4 NN	1.796	0.001
	std Potassium dissimilarity diff b/w 1 & 4 NN	1.115	0.217
	std Weight dissimilarity diff b/w 1 & 4 NN	1.023	0.775
	std Potassium dissimilarity diff b/w 2 & 3 NN	1.312	0.003
	std (DBP dissimilarity diff b/w 1 & 2 NN)^2	1.288	0.011
	std (Albumin dissimilarity diff b/w 1 & 3 NN)^2	0.278	0
	std (Weight dissimilarity diff b/w 1 & 4 NN)^2	0.923	0.318
	std (eGFR smooth t.s. dissimilarity diff b/w 2 & 3 NN)^2	1.649	0.002
	std (Albumin dissimilarity diff b/w 2 & 4 NN)^2	16.386	0.001
	std (Albumin dissimilarity diff b/w 3 & 4 NN)^2	0.277	0.035
	std (SBP dissimilarity diff b/w 3 & 4 NN)^2	1.561	0.034

Table 25b: Multinomial Logistic Regression Model, Step 1

Group	Predictor	Exp(B)	p-value
equal	Intercept		0.461
weights	[No Comorbidities=0]	0.92	0.554
	[No Comorbidities=1]		
	(eGFR peak 2-NN averaged dissimilarities)^2	2.437	0.041
	(Albumin 2-NN averaged dissimilarities)^2	1.801	0.315
	(DBP 4-NN averaged dissimilarities)^2	0.407	0.028
	(eGFR change)^2	0.544	0.049
	(max Phosphate)^2	0.996	0.575
	std eGFR peak dissimilarity diff b/w 1 & 2 NN	1.161	0.079
	std Phosphate dissimilarity diff b/w 1 & 2 NN	0.905	0.18
	std DBP dissimilarity diff b/w 1 & 3 NN	0.875	0.096
	std eGFR smooth t.s. dissimilarity diff b/w 1 & 4 NN	1.128	0.167
	std Albumin dissimilarity diff b/w 1 & 4 NN	0.866	0.159
	std Potassium dissimilarity diff b/w 1 & 4 NN	0.886	0.164
	std Weight dissimilarity diff b/w 1 & 4 NN	0.752	0.005
	std Potassium dissimilarity diff b/w 2 & 3 NN	1.19	0.037
	std (DBP dissimilarity diff b/w 1 & 2 NN)^2	0.937	0.452
	std (Albumin dissimilarity diff b/w 1 & 3 NN)^2	0.899	0.272
	std (Weight dissimilarity diff b/w 1 & 4 NN)^2	0.757	0.015
	std (eGFR smooth t.s. dissimilarity diff b/w 2 & 3 NN)^2	1.343	0.076
	std (Albumin dissimilarity diff b/w 2 & 4 NN)^2	4.04	0.038
	std (Albumin dissimilarity diff b/w 3 & 4 NN)^2	0.561	0.268
	std (SBP dissimilarity diff b/w 3 & 4 NN)^2	1.654	0.017

Table 26a: Non-equally Weighted NN Model Selection

Model	2n-NN	3n-NN	4n-NN	Total Percentage
SVM Poly=3 C=1	11	9	27	40.17
SVM Poly=3 C=2	11	9	27	40.17
SVM Poly=3 C=3	11	9	27	40.17
SVM Poly=3 C=4	11	9	27	40.17
SVM Poly=3 C=5	11	9	27	40.17
SVM Poly=3 C=6	11	9	27	40.17
SVM Poly=3 C=7	11	9	27	40.17
SVM Poly=3 C=8	11	9	27	40.17
SVM Poly=3 C=9	11	9	27	40.17
SVM Poly=3 C=10	11	9	27	40.17
C5 M=2	16	4	24	37.61
C5 M=5	25	3	14	35.90
C5 M=10	24	1	14	33.33
C5 M=15	24	4	14	35.90
C5 M=20	25	5	15	38.46

Table 26b: Non-equally Weighted NN Model Selection

Model	2n-NN	3n-NN	4n-NN	Total Percentage
Logit Forward LR	12	8	32	44.44
Logit Forward LR filtered	6	7	47	51.28
Logit Backward LR	13	11	18	35.90
LDA	9	14	22	38.46
SVM RBF C=1	9	10	34	45.30
SVM RBF C=2	11	8	31	42.74
SVM RBF C=3	12	9	27	41.03
SVM RBF C=4	14	8	28	42.74
SVM RBF C=5	14	9	30	45.30
SVM RBF C=6	14	9	33	47.86
SVM RBF C=7	12	8	32	44.44
SVM RBF C=8	12	9	35	47.86
SVM RBF C=9	12	9	35	47.86
SVM RBF C=10	10	9	35	46.15
SVM Poly=2 C=1	11	10	26	40.17
SVM Poly=2 C=2	11	10	26	40.17
SVM Poly=2 C=3	11	10	26	40.17
SVM Poly=2 C=4	11	10	26	40.17
SVM Poly=2 C=5	11	10	26	40.17
SVM Poly=2 C=6	11	10	26	40.17
SVM Poly=2 C=7	11	10	26	40.17
SVM Poly=2 C=8	11	10	26	40.17
SVM Poly=2 C=9	11	10	26	40.17
SVM Poly=2 C=10	11	10	26	40.17

Table 27: Multinomial Logistic Regression Model, Step 2 non-equally weighted nearest neighbors

Group	Predictor	Exp(B)	p-value
4n-NN	Intercept		0.004
	(min SBP)^2	1	0.01
	(min DBP)^2	1	0.112
	(days in-patient)^2	1	0.15
	std Phosphate dissimilarity diff b/w 2 & 3 NN	1.363	0.04
	std eGFR smooth t.s. dissimilarity diff b/w 2 & 4 NN	2.734	0
	std eGFR peak dissimilarity diff b/w 2 & 4 NN	1.611	0.009
	std eGFR smooth t.s. dissimilarity diff b/w 3 & 4 NN	1.304	0.376
3n-NN	Intercept		0.061
	(min SBP)^2	1	0.001
	(min DBP)^2	1	0.006
	(days in-patient)^2	1	0.754
	std Phosphate dissimilarity diff b/w 2 & 3 NN	2.413	0
	std eGFR smooth t.s. dissimilarity diff b/w 2 & 4 NN	2.911	0.001
	std eGFR peak dissimilarity diff b/w 2 & 4 NN	0.892	0.58
	std eGFR smooth t.s. dissimilarity diff b/w 3 & 4 NN	0.36	0.002

Table 28a: Equally Weighted NN Model Selection

Model	2e-NN	3e-NN	4e-NN	Total Percentage
SVM Poly=3 C=1	6	5	6	39.53
SVM Poly=3 C=2	6	5	6	39.53
SVM Poly=3 C=3	6	5	6	39.53
SVM Poly=3 C=4	6	5	6	39.53
SVM Poly=3 C=5	6	5	6	39.53
SVM Poly=3 C=6	6	5	6	39.53
SVM Poly=3 C=7	6	5	6	39.53
SVM Poly=3 C=8	6	5	6	39.53
SVM Poly=3 C=9	6	5	6	39.53
SVM Poly=3 C=10	6	5	6	39.53
C5 M=2	5	3	10	41.86
C5 M=5	2	5	10	39.53
C5 M=10	8	6	6	46.51
C5 M=15	16	4	1	48.84
C5 M=20	5v	5	8	41.86

Table 28b: Equally Weighted NN Model Selection

Model	2e-NN	3e-NN	4e-NN	Total Percentage
Logit Forward LR	8	6	5	44.19
Logit Forward LR filtered	6	5	5	37.21
Logit Backward LR	9	2	3	32.56
LDA	7	5	2	32.56v
SVM RBF C=1	5	5	3	30.23
SVM RBF C=2	7	5	4	37.21
SVM RBF C=3	7	5	4	37.21
SVM RBF C=4	7	4	4	34.88
SVM RBF C=5	7	4	6	39.53
SVM RBF C=6	7	5	6	41.86
SVM RBF C=7	7	5	6	41.86
SVM RBF C=8	7	6	6	44.19
SVM RBF C=9	7	6	6	44.19
SVM RBF C=10	7	5	6	41.86
SVM Poly=2 C=1	7	5	6	41.86
SVM Poly=2 C=2	7	5	6	41.86
SVM Poly=2 C=3	7	5	6	41.86
SVM Poly=2 C=4	7	5	6	41.86
SVM Poly=2 C=5	7	5	6	41.86
SVM Poly=2 C=6	7	5	6	41.86
SVM Poly=2 C=7	7	5	6	41.86
SVM Poly=2 C=8	7	5	6	41.86
SVM Poly=2 C=9	7	5	6	41.86
SVM Poly=2 C=10	7	5	6	41.86

Table 29: C5.0 Model, Step 2 equally weighted nearest neighbors

Nodes
standardized SBP dissimilarity diff b/w 1 & 2 NN
Smooth GFR slope
standardized SBP dissimilarity diff b/w 1 & 3 NN
standardized (phosphate dissimilarity diff b/w 3 & 4 NN)2
standardized eGFR peak dissimilarity diff b/w 1 & 4 NN
Average smooth weight
standardized weight dissimilarity diff b/w 3 & 4 NN
standardized eGFR smooth t.s. dissimilarity diff b/w 1 & 2 NN
standardized (weight dissimilarity diff b/w 1 & 3 NN)^2

APPENDIX B

MDP PROOFS

Note: $[A^{\circ(n)}]_{(i,j)} = a^n_{(i,j)}$
Lemma 2:

Proof. Let n = 1
$[(P_1^2 - P_1^1)v_N^*]_s - P_1^2(s,s)[r_1^{\circ(2)}]_s \geq [(P_1^1 - P_1^0)v_N^*]_s - P_1^1(s,s)[r_1^{\circ(1)}]_s$
$[P_1^2 v_N^* - P_1 v_N^*]_s - P_1^2(s,s)[r_1^{\circ(2)}]_s \geq [P_1 v_N^* - v_N^*]_s - P_1(s,s)[r_1]_s$
But $[P_1 v_N^*]_s = [v_{N-1}^* - r_1]_s$
$[P_1^2 v_N^* - P_1 v_N^* + r_1]_s + P_1(s,s)[r_1]_s - P_1^2(s,s)[r_1^{\circ(2)}]_s \geq [v_{N-1}^* - v_N^*]_s$
$[P_1^2 v_N^* - v_{N-1}^*]_s + P_1(s,s)[r_1]_s \geq [v_{N-1}^* - v_N^*]_s + P_1^2(s,s)[r_1^{\circ(2)}]_s$
But $[v_{N-2}^*]_s = max[r_1 + P_1 r_1 + P_1^2 v_N^*; r_2 + P_2 v_N^*]_s$,
$\Rightarrow [v_{N-2}^*]_s \geq [r_1 + P_1 r_1 + P_1^2 v_N^*]_s \geq [P_1^2 v_N^* + P_1 r_1]_s \geq [P_1^2 v_N^*]_s + P_1(s,s)[r_1]_s$
$[v_{N-2}^* - v_{N-1}^*]_s \geq [P_1^2 v_N^* - v_{N-1}^*]_s + P_1(s,s)[r_1]_s \geq [v_{N-1}^* - v_N^*]_s + P_1^2(s,s)[r_1^{\circ(2)}]_s$
But $P_1^2(s,s)[r_1^{\circ(2)}]_s \geq 0$
$\Rightarrow [v_{N-2}^* - v_{N-1}^*]_s \geq [v_{N-1}^* - v_N^*]_s$
Let n = t
$[(P_1^t - P_1^{t-1})v_N^*]_s - P_1^t(s,s)[r_1^{\circ(t)}]_s \geq [(P_1^{t-1} - P_1^{t-2})v_N^*]_s - P_1^{t-1}(s,s)[r_1^{\circ(t-1)}]_s$
$[(P_1^t + P_1^{t-2})v_N^*]_s \geq [2P_1^{t-1}v_N^*]_s + P_1^t(s,s)[r_1^{\circ(t)}]_s - P_1^{t-1}(s,s)[r_1^{\circ(t-1)}]_s$
$[(P_1^t + P_1^{t-2})v_N^*]_s \geq [2P_1^{t-1}v_N^*]_s + [r_1^{\circ(t-1)}]_s(P_1^t(s,s)[r_1]_s - P_1^{t-1}(s,s))$
If $[r_1]_s \in [P_1(s,s)^{-1}, 365/8) \bigcup \{0\} \Rightarrow [r_1^{\circ(t-1)}]_s(P_1^t(s,s)[r_1]_s - P_1^{t-1}(s,s)) \geq 0$
$[v_t^*]_s \geq [f_1(r_1) + P_1^t v_N^*]_s \geq [P_1^t v_N^*]_s$
where $[f_1(r_1) + P_1^t v_N^*]_s$ is the value of $v_t^*(s, disu_{appt})$ if the optimal policy was having constant appointments of type a_1 between periods t and N and $[f_1(r_1)]_s \geq 0$.
$[v_{t-2}^*]_s \geq [f_2(r_1) + P_1^{t-2} v_N^*]_s \geq [P_1^{t-2} v_N^*]_s$
where $[f_2(r_1) + P_1^{t-2} v_N^*]_s$ is the value of $v_{t-2}^*(s, disu_{appt})$ if the optimal policy was having constant appointments of type a_1 between periods t - 2 and N and $[f_2(r_1)]_s \geq 0$.
$\Rightarrow [v_t^* + v_{t-2}^*]_s \geq [(P_1^t + P_1^{t-2})v_N^*]_s \geq [2P_1^{t-1}v_N^*]_s + [r_1^{\circ(t-1)}]_s(P_1^t(s,s)[r_1]_s - P_1^{t-1}(s,s))$
Based on (3) & (4) $[2P_1^{t-1}v_N^*]_s + [r_1^{\circ(t-1)}]_s(P_1^t(s,s)[r_1]_s - P_1^{t-1}(s,s)) \geq 2[v_{t-1}^*]_s$
$[v_{t-2}^* + v_t^*]_s \geq 2[v_{t-1}^*]_s \Rightarrow [v_{t-2}^* - v_{t-1}^*]_s \geq [v_{t-1}^* - v_t^*]_s$
Based on (2) $[r_1 + (P_1 - I)v_N^*]_s \geq 0 \Rightarrow [v_{N-1}^* - v_N^*]_s \geq 0$

Therefore by induction $[v^*_{t-2} - v^*_{t-1}]_s \geq [v^*_{t-1} - v^*_t]_s \geq ... \geq [v^*_{N-1} - v^*_N]_s \geq 0$
Therefore, $[v^*_{n-1} - v^*_n]_s \geq 0$ & non-increasing in n. □

Lemma 3:

Proof. By lemma 2 $[v^*_{n-1} - v^*_n]_s \geq 0$ & non-increasing in n.
If $[P_j]_s$ - convex then $f([P_j]_s) = [P_j(v^*_{n-1} - v^*_n)]_s$ - convex.
According to lemma 1 where $i^{'} = 1$ and $n = N - q + i$
$[P_1(v^*_{N-q+i-1} - v^*_{N-q+i})]_s \geq [P_i(v^*_{N-q+i-1} - v^*_{N-q+i})]_s$
Based on lemma 2
$[P_1(v^*_{N-q+1-1} - v^*_{N-q+1})]_s \geq [P_1(v^*_{N-q+i-1} - v^*_{N-q+i})]_s$, because $i = 2, 4, 8$ i.e. $i > 1$
$\Rightarrow [P_1(v^*_{N-q+1-1} - v^*_{N-q+1})]_s \geq [P_i(v^*_{N-q+i-1} - v^*_{N-q+i})]_s$
$[P_1 v^*_{N-q+1-1} - P_1 v^*_{N-q+1}]_s \geq [P_i v^*_{N-q+i-1} - P_i v^*_{N-q+i}]_s$
$[P_i v^*_{N-q+i} - P_1 v^*_{N-q+1}]_s \geq [P_i v^*_{N-q+i-1} - P_1 v^*_{N-q+1-1}]_s$
$[P_i v^*_{N-q+i} - P_1 v^*_{N-q+1}]_s \geq [P_i v^*_{N-(q+1)+i} - P_1 v^*_{N-(q+1)+1}]_s$
□

Proposition 1:

Proof. For a_1 to be optimal in all time periods we need to prove that
$[r_1 - r_i]_s \geq [P_i v^*_{N-n+i} - P_1 v^*_{N-n+1}]_s$ for $i \in \{2, 4, 8\}$ & $\forall n$
By induction:
(*) Let n = i (the smallest possible value in these settings)
$[r_1 - r_i]_s \geq [P_i v^*_N - P_1 v^*_{N-i+1}]_s$
true based on the assumption listed in the proposition.
(**) Let n = q, assume that the following statement is correct:
$[r_1 - r_i]_s \geq [P_i v^*_{N-q+i} - P_1 v^*_{N-q+1}]_s$
(***) Let n = q + 1; we need to prove that the following inequality holds:
$[r_1 - r_i]_s \geq [P_i v^*_{N-(q+1)+i} - P_1 v^*_{N-(q+1)+1}]_s$
By lemma 3 $[P_i v^*_{N-q+i} - P_1 v^*_{N-q+1}]_s \geq [P_i v^*_{N-(q+1)+i} - P_1 v^*_{N-(q+1)+1}]_s$
By (**) $[r_1 - r_i]_s \geq [P_i v^*_{N-q+i}]_s - [P_1 v^*_{N-q+1}]_s$
$\Rightarrow [r_1 - r_i]_s \geq [P_i v^*_{N-(q+1)+i}]_s - [P_1 v^*_{N-(q+1)+1}]_s$
$\Rightarrow$ By induction we can conclude that $[r_1 - r_i]_s \geq [P_k v^*_{N-n+i} - P_{i'} v^*_{N-n-1}]_s, \forall n$
$\Rightarrow [r_1 - r_i]_s \geq max_{i \neq 1}[r_i + P_i v^*_{N-n+i}]_s$
$\Rightarrow v^*_{N-n}(s) = max_{i^*=1,2,4,8}[r_{i^*} + P_{i^*} v^*_{N-n+i^*}]_s = [r_1 + P_1 v^*_{N-n+1}]_s, \forall n$
Therefore, if the assumptions made in proposition 1 hold, then a_1 is consistently optimal across the entire finite time horizon. □

Lemma 4

Proof. Let n = 1
$[(P_1 - I)P_1^1(v^*_N - r_8^{\circ(2)})]_s \leq [(P_1 - I)P_1^0(v^*_N - r_8^{\circ(1)})]_s$
$[P_1^2 v^*_N - v^*_{N-1} - (P_1 - I)P_1 r_8^{\circ(2)} + (P_1 - I)r_8]_s \leq [v^*_{N-1} - v^*_N]_s$
$[P_1^2 v^*_N - v^*_{N-1} + (I - P_1)(P_1 r_8^{\circ(2)} - r_8)]_s \leq [v^*_{N-1} - v^*_N]_s$
From (4) $[P_1^2 v^*_N]_s \geq [P_2 v^*_N]_s$
From (3) $[(P_1 - I)(P_1 r_8^{\circ(2)} - r_8)]_s \geq max([r_2]_s, 2[r_1]_s) \geq max([r_2]_s, [r_1]_s + [P_1 r_1]_s)$

Therefore $[v^*_{N-2}]_s \leq [P_1^2 v^*_N + (I - P_1)(P_1 r_8^{\circ(2)} - r_8)]_s$
$[v^*_{N-2} - v^*_{N-1}]_s \leq [P_1^2 v^*_N + (I - P_1)(P_1 r_8^{\circ(2)} - r_8) - v^*_{N-1}]_s \leq [v^*_{N-1} - v^*_N]_s$
$[v^*_{N-2} - v^*_{N-1}]_s \leq [v^*_{N-1} - v^*_N]_s$
Let n = t
$[(P_1 - I)P_1^t(v^*_N - r_8^{\circ(t+1)})]_s \leq [(P_1 - I)P_1^{t-1}(v^*_N - r_8^{\circ(t)})]_s$
$[P_1^{t+1} v^*_N + P_1^{t-1} v^*_N]_s + [(I - P_1)(P_1^t r_8^{\circ(t+1)} - P_1^{t-1} r_8^{\circ(t)})]_s \leq [2P_1^t v^*_N]_s \leq 2[v^*_{N-t}]_s$
From (3) $[P_1^t v^*_N]_s \geq [P_1^i P_2^j P_4^k P_8^l v^*_N]_s$ where t = i + j*2 + k*4 + l*8
From (4) $[(I - P_1)(P_1^t r_8^{\circ(t+1)} - P_1^{t-1} r_8^{\circ(t)})]_s \geq 2t * max\{[r_1]_s, \frac{[r_2]_s}{2}, \frac{[r_4]_s}{4}, \frac{[r_8]_s}{8}\}$
$[v^*_{N-t+1} + v^*_{N-t-1}]_s \leq [P_1^{t+1} v^*_N + P_1^{t-1} v^*_N]_s + [(I - P_1)(P_1^t r_8^{\circ(t+1)} - P_1^{t-1} r_8^{\circ(t)})]_s \leq 2[v^*_{N-t}]_s$
$[v^*_{N-t+1} + v^*_{N-t-1}]_s \leq 2[v^*_{N-t}]_s \Rightarrow [v^*_{N-t-1} - v^*_{N-t}]_s \leq [v^*_{N-t} - v^*_{N-t+1}]_s$
Based on (2) $[r_1 + (P_1 - I)v^*_N]_s \leq 0 \Rightarrow [v^*_{N-1} - v^*_N]_s \leq 0$
Therefore by induction $[v^*_{N-t-1} - v^*_{N-t}]_s \leq [v^*_{N-t} - v^*_{N-t+1}]_s \leq ... \leq [v^*_{N-1} - v^*_N]_s \leq 0$
Therefore, $[v^*_{n-1} - v^*_n]_s \leq 0$ & non-decreasing in n. □

Lemma 5:

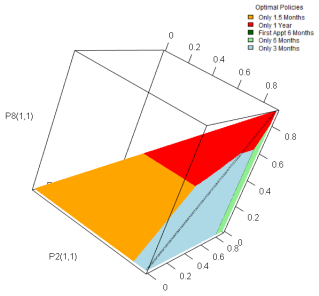
Proof. By lemma 4 $[v^*_{n-1} - v^*_n]_s \leq 0$ & non-decreasing in n.
If $[P_j]_s$ - concave then $f([P_j]_s) = [P_j(v^*_{n-1} - v^*_n)]_s$ - convex.
According to lemma 1 where $i' = 1$ and $n = N - q + i$
$[P_8(v^*_{N-q+i-1} - v^*_{N-q+i})]_s \geq [P_i(v^*_{N-q+i-1} - v^*_{N-q+i})]_s$
Based on lemma 4
$[P_8(v^*_{N-q+8-1} - v^*_{N-q+8})]_s \geq [P_8(v^*_{N-q+i-1} - v^*_{N-q+i})]_s$, because $i = 1, 2, 4$ i.e. $i < 8$
$\Rightarrow [P_8(v^*_{N-q+8-1} - v^*_{N-q+8})]_s \geq [P_i(v^*_{N-q+i-1} - v^*_{N-q+i})]_s$
$[P_8 v^*_{N-q+8-1} - P_8 v^*_{N-q+8}]_s \geq [P_i v^*_{N-q+i-1} - P_i v^*_{N-q+i}]_s$
$[P_i v^*_{N-q+i} - P_8 v^*_{N-q+8}]_s \geq [P_i v^*_{N-q+i-1} - P_8 v^*_{N-q+8-1}]_s$
$[P_i v^*_{N-q+i} - P_8 v^*_{N-q+8}]_s \geq [P_i v^*_{N-(q+1)+i} - P_8 v^*_{N-(q+1)+8}]_s$
□

Proposition 2:

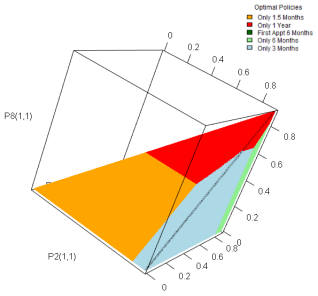
Proof. The proof is almost identical to the one in Proposition 1. □

APPENDIX C

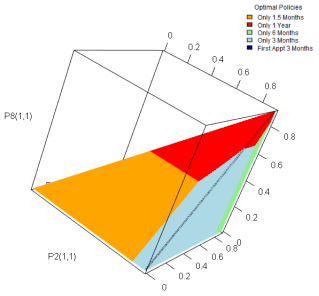
MDP GRAPHS



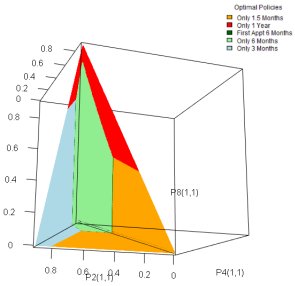
(a) s = (0,0,0,3,65)



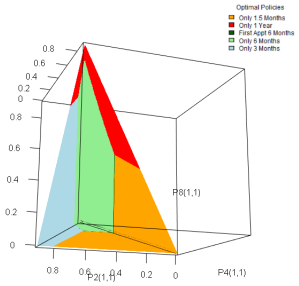
(b) s = (0,0,0,4,65)



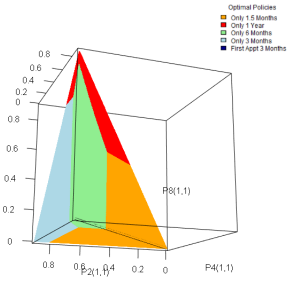
(c) s = (0,0,0,5,65)



(d) s = (0,0,0,3,65)

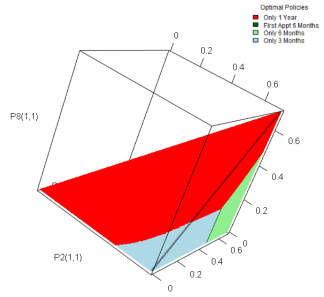


(e) s = (0,0,0,4,65)

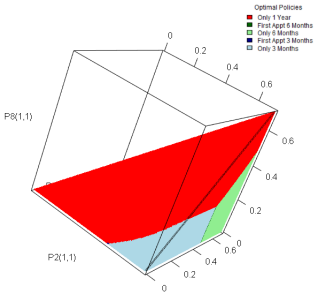


(f) s = (0,0,0,5,65)

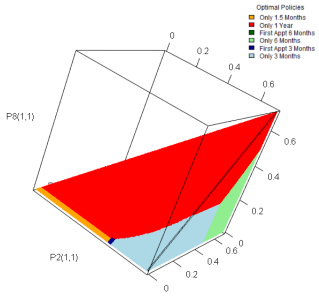
Figure 14: $p_{1_{(1,1)}} = 0.90$ & N = 1 year



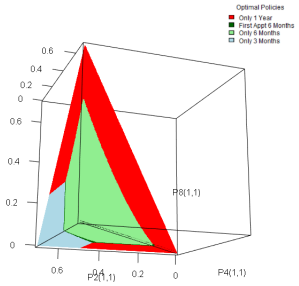
(a) s = (0,0,0,3,65)



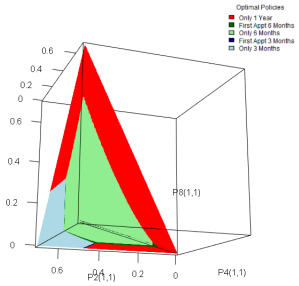
(b) s = (0,0,0,4,65)



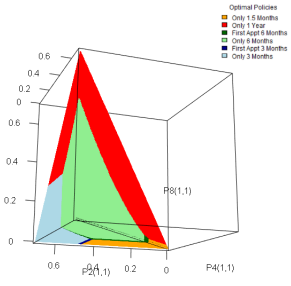
(c) s = (0,0,0,5,65)



(d) s = (0,0,0,3,65)

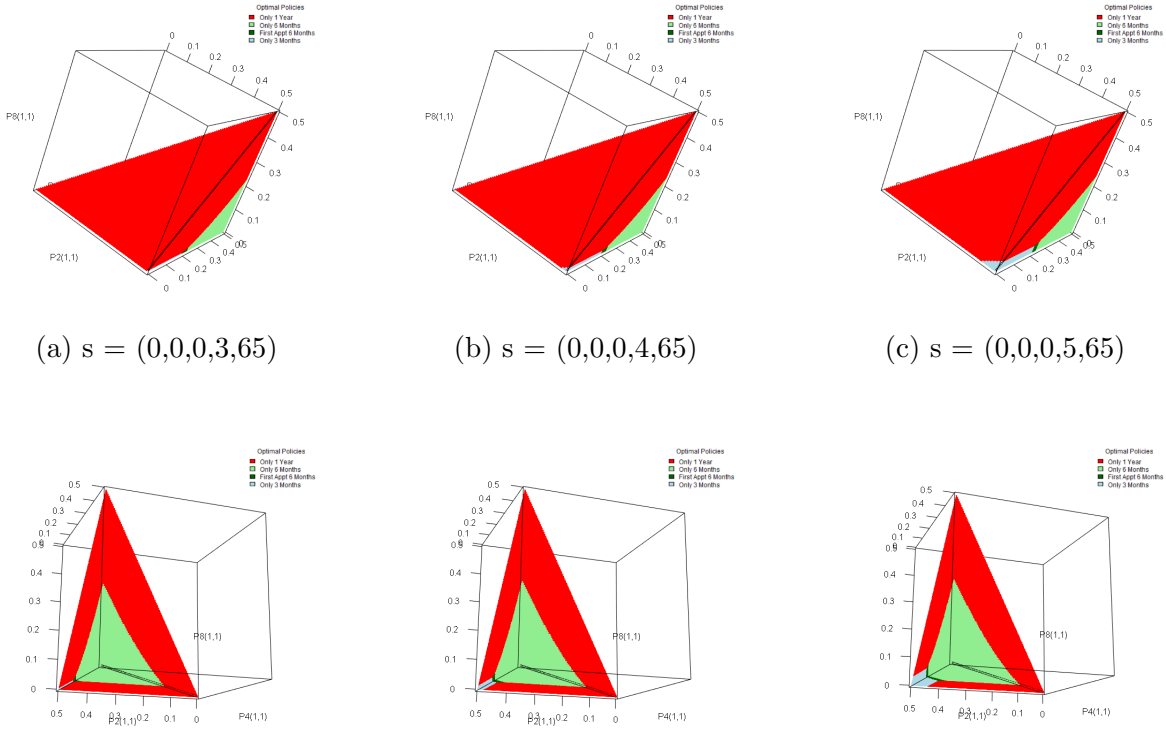


(e) s = (0,0,0,4,65)



(f) s = (0,0,0,5,65)

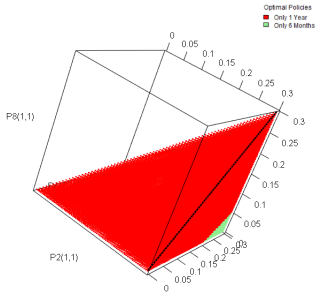
Figure 15: $p_{1_{(1,1)}} = 0.70$ & N = 1 year



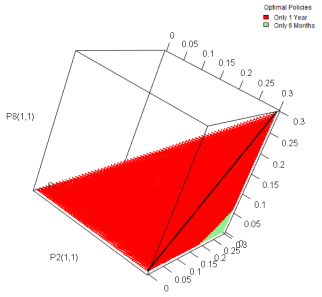
(a) s = (0,0,0,3,65) (b) s = (0,0,0,4,65) (c) s = (0,0,0,5,65)

(d) s = (0,0,0,3,65) (e) s = (0,0,0,4,65) (f) s = (0,0,0,5,65)

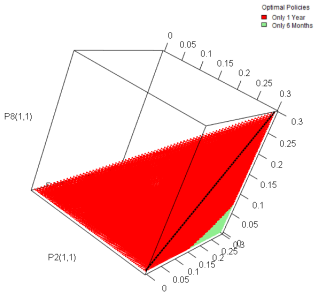
Figure 16: $p_{1_{(1,1)}} = 0.50$ & N = 1 year



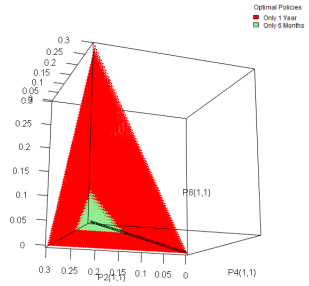
(a) s = (0,0,0,3,65)



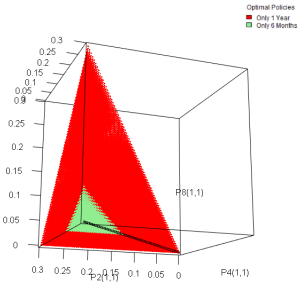
(b) s = (0,0,0,4,65)



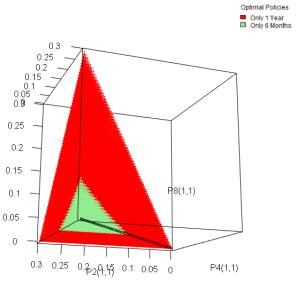
(c) s = (0,0,0,5,65)



(d) s = (0,0,0,3,65)

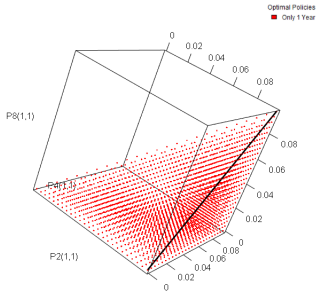


(e) s = (0,0,0,4,65)

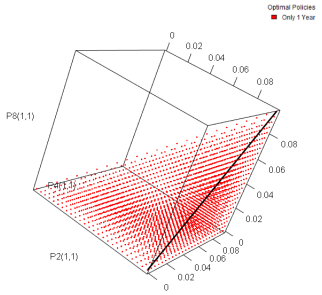


(f) s = (0,0,0,5,65)

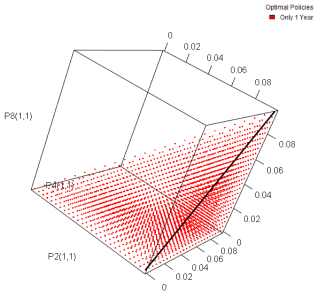
Figure 17: $p_{1_{(1,1)}} = 0.30$ & N = 1 year



(a) s = (0,0,0,3,65)

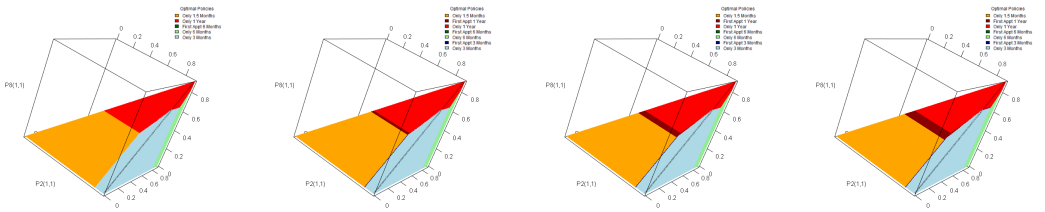


(b) s = (0,0,0,4,65)

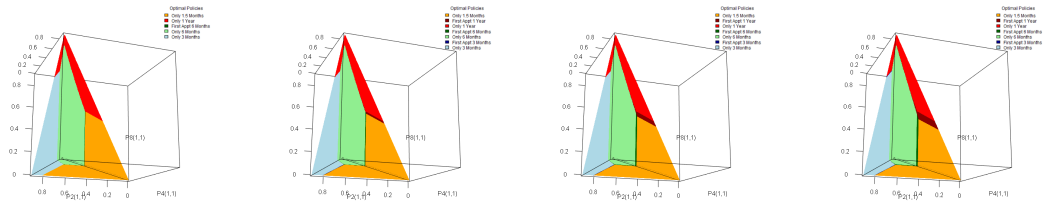


(c) s = (0,0,0,5,65)

Figure 18: $p_{1_{(1,1)}} = 0.10$ & $\forall N$

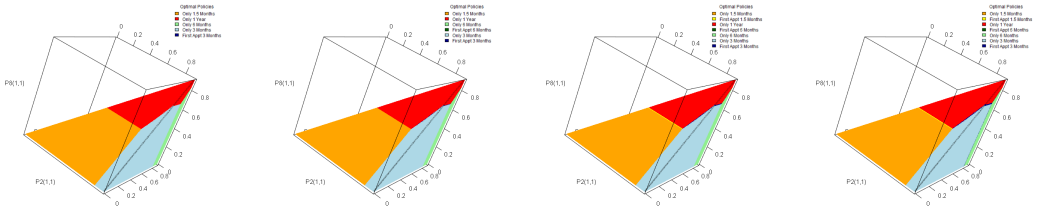


(a) N = 1 year (b) N = 2 years (c) N = 3 years (d) N = 4 years

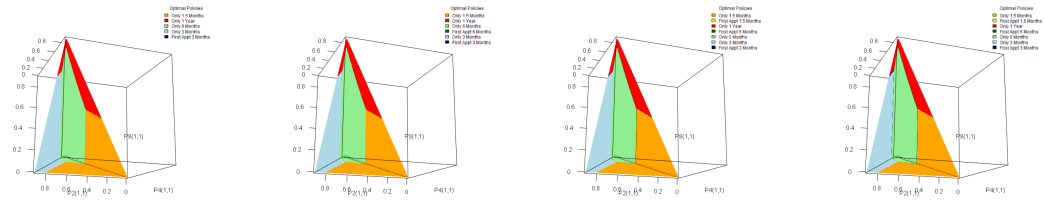


(e) N = 1 year (f) N = 2 years (g) N = 3 years (h) N = 4 years

Figure 19: s = (0,0,0,4,65) & $p_{1_{(1,1)}} = 0.9$

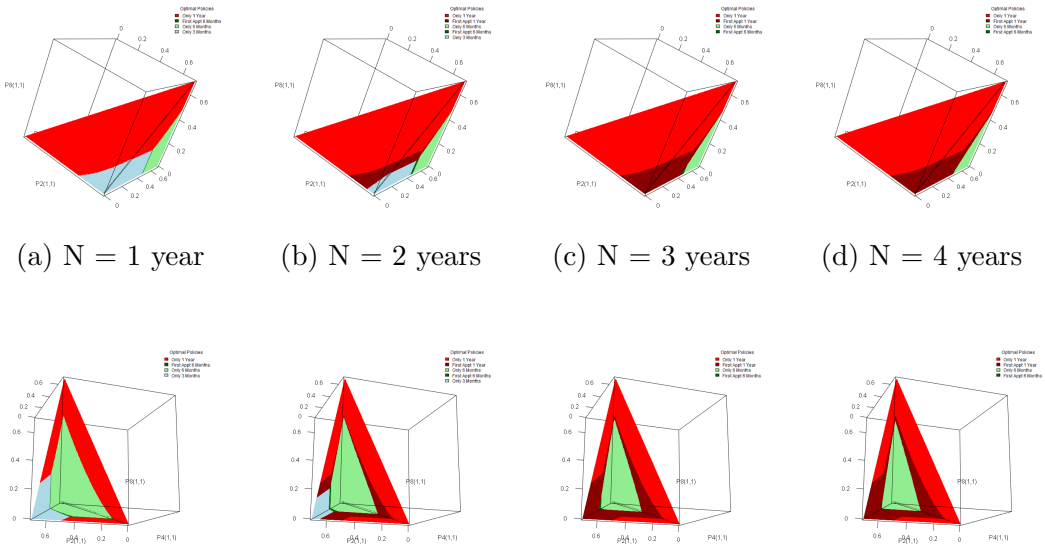


(a) N = 1 year (b) N = 2 years (c) N = 3 years (d) N = 4 years



(e) N = 1 year (f) N = 2 years (g) N = 3 years (h) N = 4 years

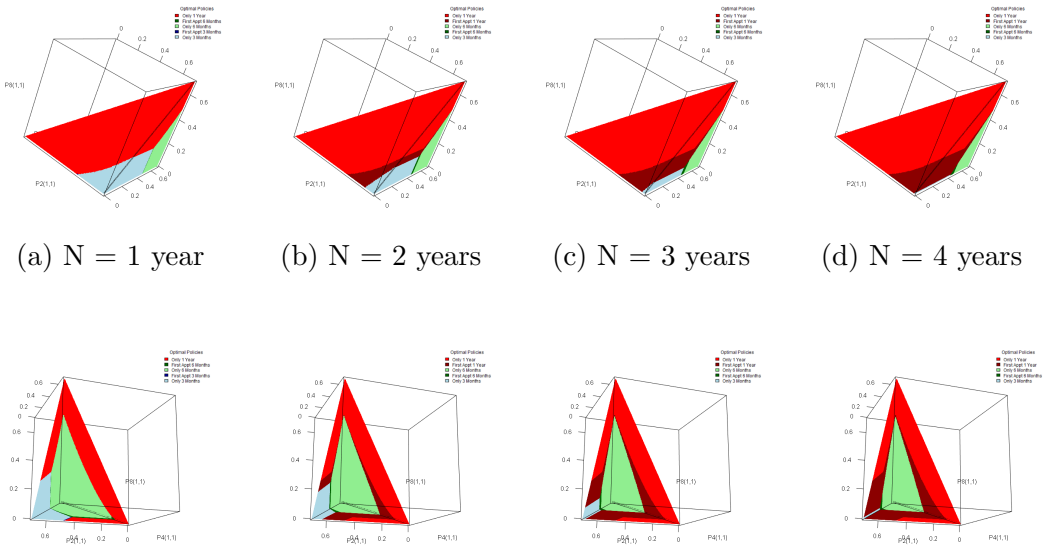
Figure 20: s = (0,0,0,5,65) & $p_{1_{(1,1)}} = 0.9$



(a) N = 1 year (b) N = 2 years (c) N = 3 years (d) N = 4 years

(e) N = 1 year (f) N = 2 years (g) N = 3 years (h) N = 4 years

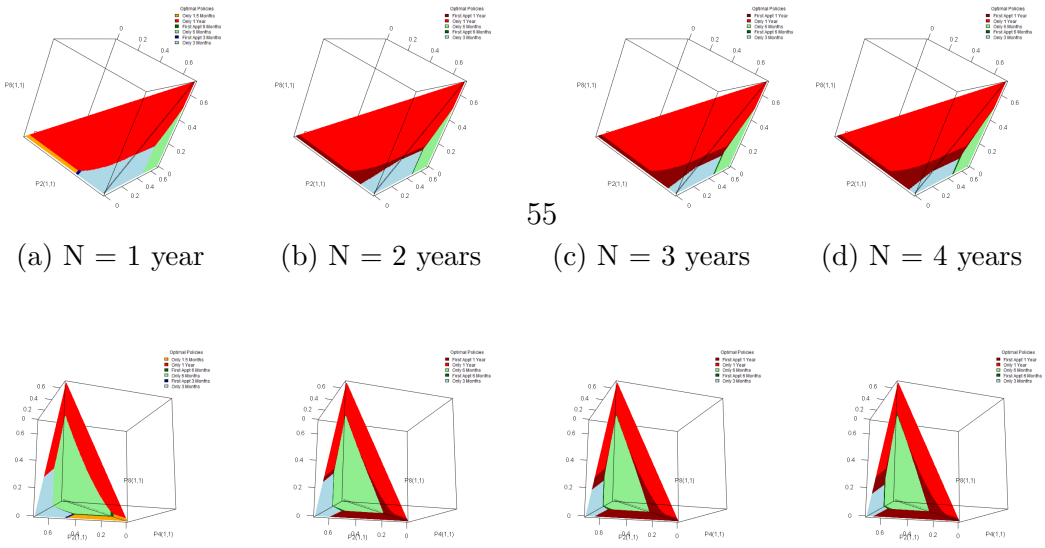
Figure 21: s = (0,0,0,3,65) & $p_{1_{(1,1)}} = 0.7$



(a) N = 1 year (b) N = 2 years (c) N = 3 years (d) N = 4 years

(e) N = 1 year (f) N = 2 years (g) N = 3 years (h) N = 4 years

Figure 22: s = (0,0,0,4,65) & $p_{1_{(1,1)}} = 0.7$

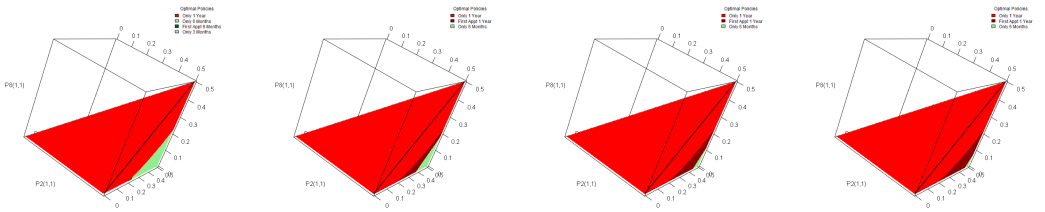


55

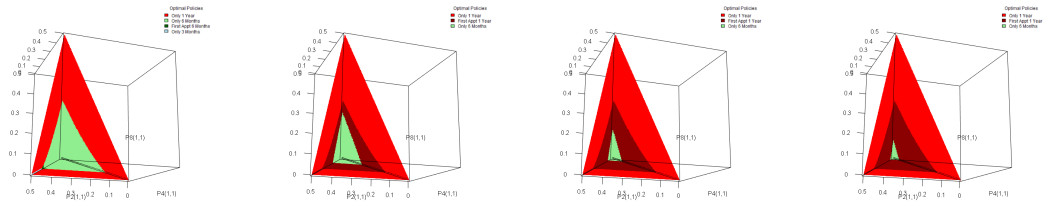
(a) N = 1 year (b) N = 2 years (c) N = 3 years (d) N = 4 years

(e) N = 1 year (f) N = 2 years (g) N = 3 years (h) N = 4 years

Figure 23: s = (0,0,0,5,65) & $p_{1_{(1,1)}} = 0.7$

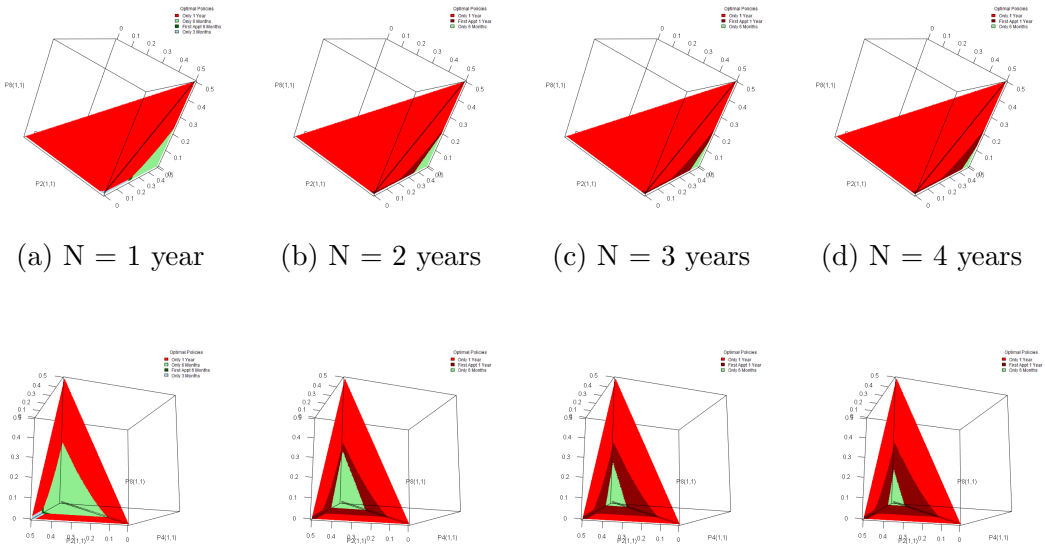


(a) N = 1 year (b) N = 2 years (c) N = 3 years (d) N = 4 years



(e) N = 1 year (f) N = 2 years (g) N = 3 years (h) N = 4 years

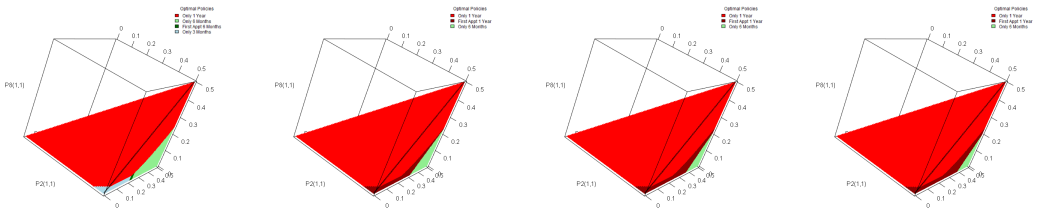
Figure 24: s = (0,0,0,3,65) & $p_{1_{(1,1)}} = 0.5$



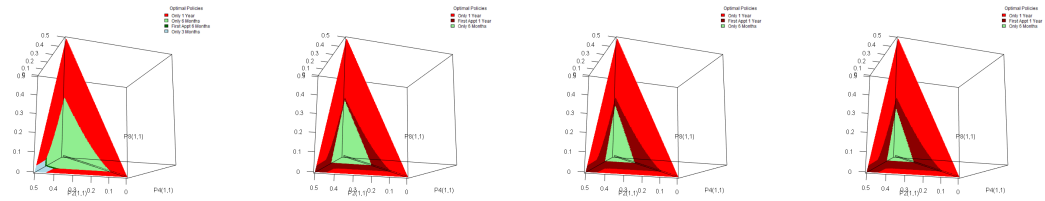
(a) N = 1 year (b) N = 2 years (c) N = 3 years (d) N = 4 years

(e) N = 1 year (f) N = 2 years (g) N = 3 years (h) N = 4 years

Figure 25: s = (0,0,0,4,65) & $p_{1_{(1,1)}} = 0.5$

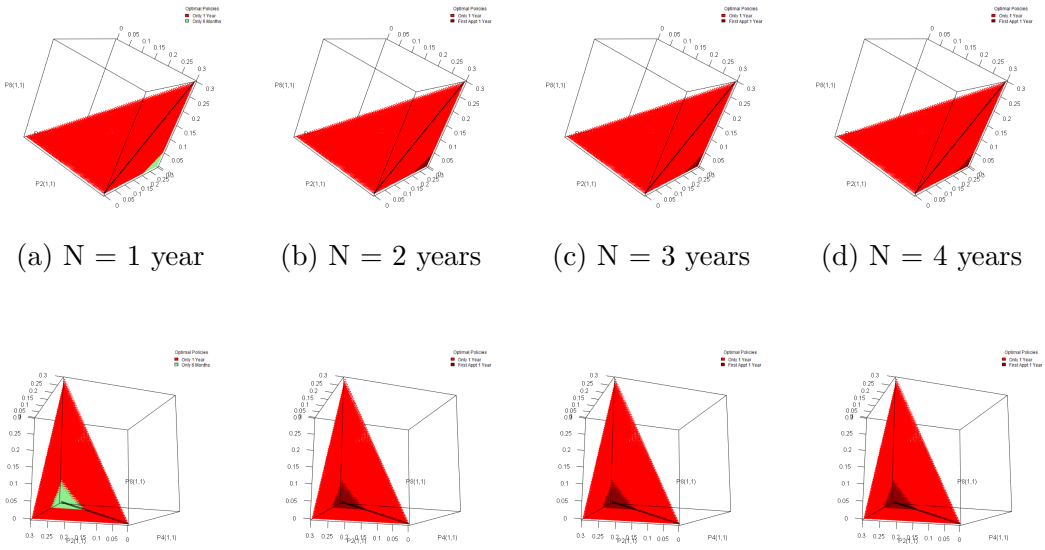


(a) N = 1 year (b) N = 2 years (c) N = 3 years (d) N = 4 years



(e) N = 1 year (f) N = 2 years (g) N = 3 years (h) N = 4 years

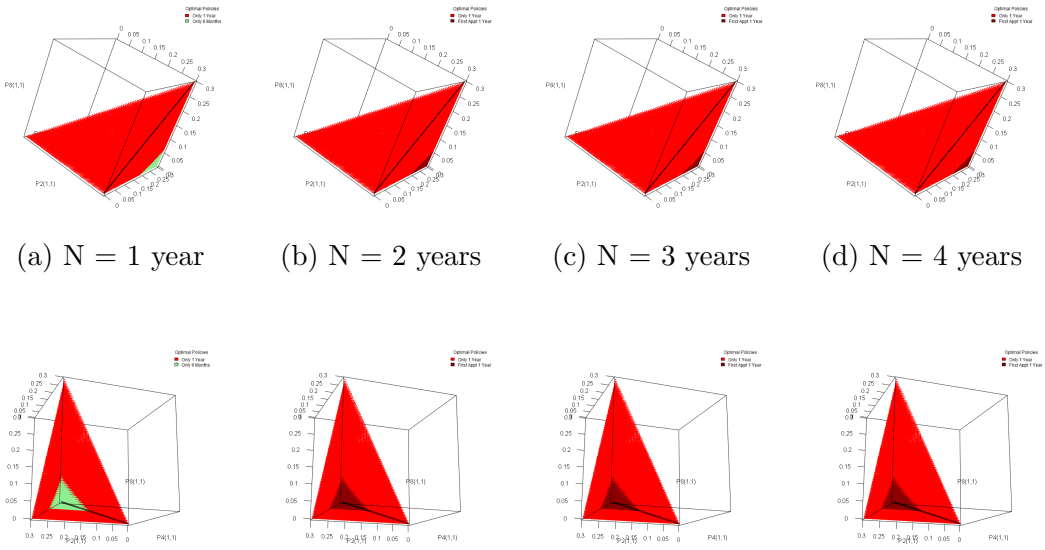
Figure 26: s = (0,0,0,5,65) & $p_{1_{(1,1)}} = 0.5$



(a) N = 1 year (b) N = 2 years (c) N = 3 years (d) N = 4 years

(e) N = 1 year (f) N = 2 years (g) N = 3 years (h) N = 4 years

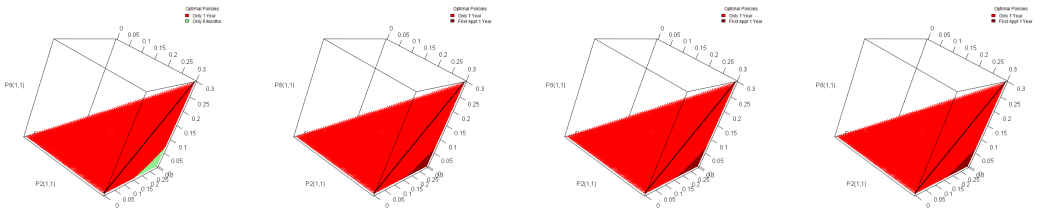
Figure 27: s = (0,0,0,3,65) & $p_{1_{(1,1)}} = 0.3$



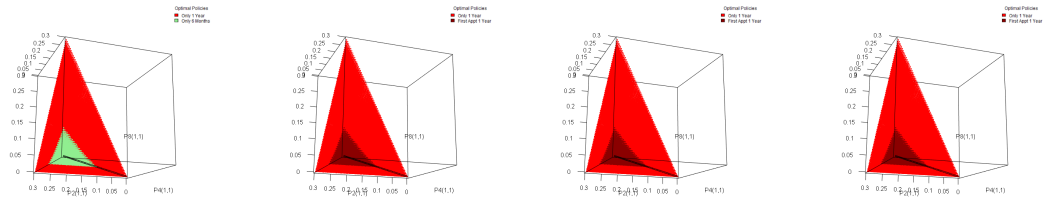
(a) N = 1 year (b) N = 2 years (c) N = 3 years (d) N = 4 years

(e) N = 1 year (f) N = 2 years (g) N = 3 years (h) N = 4 years

Figure 28: s = (0,0,0,4,65) & $p_{1_{(1,1)}} = 0.3$



(a) N = 1 year (b) N = 2 years (c) N = 3 years (d) N = 4 years



(e) N = 1 year (f) N = 2 years (g) N = 3 years (h) N = 4 years

Figure 29: s = (0,0,0,5,65) & $p_{1_{(1,1)}} = 0.3$

Bibliography

[1] Hyunchul Ahn, Kyoung-jae Kim, and Ingoo Han. Global optimization of feature weights and the number of neighbors that combine in a case-based reasoning system. *Expert Systems*, 23(5):290–301, 2006.

[2] Hirotugu Akaike. Akaike's information criterion. In *International Encyclopedia of Statistical Science*, pages 25–25. Springer, 2011.

[3] Oguzhan Alagoz, Cindy L Bryce, Steven Shechter, Andrew Schaefer, Chung-Chou H Chang, Derek C Angus, and Mark S Roberts. Incorporating biological natural history in simulation models: empirical estimates of the progression of end-stage liver disease. *Medical Decision Making*, 25(6):620–632, 2005.

[4] Oguzhan Alagoz, Lisa M Maillart, Andrew J Schaefer, and Mark S Roberts. The optimal timing of living-donor liver transplantation. *Management Science*, 50(10):1420–1430, 2004.

[5] Bradley P Allen. Case-based reasoning: Business applications. *Communications of the ACM*, 37(3):40–43, 1994.

[6] Nagesh S Anavekar, John JV McMurray, Eric J Velazquez, Scott D Solomon, Lars Kober, Jean-Lucien Rouleau, Harvey D White, Rolf Nordlander, Aldo Maggioni, Kenneth Dickstein, et al. Relation between renal dysfunction and cardiovascular outcomes after myocardial infarction. *New England Journal of Medicine*, 351(13):1285–1295, 2004.

[7] Nishkanta Arulkumaran, Ramaswamy Diwakar, Zaheer Tahir, Maha Mohamed, Juan Carlos Kaski, and Debasish Banerjee. Pulse pressure and progression of chronic kidney disease. *JN journal of nephrology*, 23(2):189, 2010.

[8] Turgay Ayer, Oguzhan Alagoz, and Natasha K Stout. Or forum—a pomdp approach to personalize mammography screening decisions. *Operations Research*, 60(5):1019–1034, 2012.

[9] Kerstin Bach, Odd Erik Gundersen, Christian Knappskog, and Pinar Öztürk. Automatic case capturing for problematic drilling situations. In *International Conference on Case-Based Reasoning*, pages 48–62. Springer, 2014.

[10] Shahina Begum, Mobyen Uddin Ahmed, Peter Funk, Ning Xiong, and Mia Folke. Case-based reasoning systems in the health sciences: a survey of recent trends and developments. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 41(4):421–434, 2011.

[11] S Berchtold, DA Keim, and HP Kriegel. An index structure for high-dimensional data. *Readings in multimedia computing and networking*, page 451, 2001.

[12] M Bethesda. Us renal data system, usrds 2011 annual data report: Atlas of chronic kidney disease and end-stage renal disease in the united states. *National Institute of Diabetes and Digestive and Kidney Diseases*, 10, 2011.

[13] Leo Breiman, Jerome Friedman, Charles J Stone, and Richard A Olshen. *Classification and regression trees*. CRC press, 1984.

[14] Jane E. Brody. Kidney Disease, an Underestimated Killer. https://well.blogs.nytimes.com/2013/07/15/kidney-disease-an-underestimated-killer/, 2013. [Online; accessed 30-July-2017].

[15] Tom Chiu, DongPing Fang, John Chen, Yao Wang, and Christopher Jeris. A robust and scalable clustering algorithm for mixed type attributes in large database environment. In *Proceedings of the seventh ACM SIGKDD international conference on knowledge discovery and data mining*, pages 263–268. ACM, 2001.

[16] Josef Coresh, Brad C Astor, Tom Greene, Garabed Eknoyan, and Andrew S Levey. Prevalence of chronic kidney disease and decreased kidney function in the adult us population: Third national health and nutrition examination survey. *American Journal of Kidney Diseases*, 41(1):1–12, 2003.

[17] Josef Coresh, G Laura Wei, Geraldine McQuillan, Fredrick L Brancati, Andrew S Levey, Camille Jones, and Michael J Klag. Prevalence of high blood pressure and elevated serum creatinine level in the united states: findings from the third national health and nutrition examination survey (1988-1994). *Archives of internal medicine*, 161(9):1207–1216, 2001.

[18] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine learning*, 20(3):273–297, 1995.

[19] Paul E Drawz and Mark E Rosenberg. Slowing progression of chronic kidney disease. *Kidney international supplements*, 3(4):372–376, 2013.

[20] Naihua Duan, Willard G Manning, Carl N Morris, and Joseph P Newhouse. A comparison of alternative models for the demand for medical care. *Journal of business & economic statistics*, 1(2):115–126, 1983.

[21] Jean Duchon. Splines minimizing rotation-invariant semi-norms in sobolev spaces. *Constructive theory of functions of several variables*, pages 85–100, 1977.

[22] Khaled El Emam, Saida Benlarbi, Nishith Goel, and Shesh N Rai. Comparing case-based reasoning classifiers for predicting high risk software components. *Journal of Systems and Software*, 55(3):301–320, 2001.

[23] Fatih Safa Erenay, Oguzhan Alagoz, and Adnan Said. Optimizing colonoscopy screening for colorectal cancer prevention and surveillance. *Manufacturing & Service Operations Management*, 16(3):381–400, 2014.

[24] Robert G Fassett, Iain K Robertson, Rose Mace, Loren Youl, Sarah Challenor, and Rosalind Bull. Palliative care in end-stage kidney disease. *Nephrology*, 16(1):4–12, 2011.

[25] Centers for Disease Control, Prevention (CDC, et al. Prevalence of chronic kidney disease and associated risk factors–united states, 1999-2004. *MMWR. Morbidity and mortality weekly report*, 56(8):161, 2007.

[26] Centers for Disease Control, Prevention, et al. National diabetes statistics report: estimates of diabetes and its burden in the united states, 2014. *Atlanta, GA: US Department of Health and Human Services*, 2014, 2014.

[27] National Center for Health Statistics (US et al. Health, united states, 2015: with special feature on racial and ethnic health disparities. 2016.

[28] John Fox, Sanford Weisberg, and Douglas Bates. car: Companion to applied regression. r package version 2.0-2, 2010.

[29] Mitchell Gail, Roger Williams, David P Byar, and Charles Brown. How many controls? *Journal of chronic diseases*, 29(11):723–731, 1976.

[30] Ron T Gansevoort, Ricardo Correa-Rotter, Brenda R Hemmelgarn, Tazeen H Jafar, Hiddo J Lambers Heerspink, Johannes F Mann, Kunihiro Matsushita, and Chi Pang Wen. Chronic kidney disease and cardiovascular risk: epidemiology, mechanisms, and prevention. *The Lancet*, 382(9889):339–352, 2013.

[31] Colin C Geddes, Jonathan G Fox, Marjory EM Allison, J Michael Boulton-Jones, and Keith Simpson. An artificial neural network can select patients at high risk of developing progressive iga nephropathy more accurately than experienced nephrologists. *Nephrology Dialysis Transplantation*, 13(1):67–71, 1998.

[32] Lothar Gierl, Mathias Bull, and Rainer Schmidt. Cbr in medicine. In *Case-Based Reasoning Technology*, pages 273–297. Springer, 1998.

[33] Ronald Gijsen, Nancy Hoeymans, François G Schellevis, Dirk Ruwaard, William A Satariano, and Geertrudis AM van den Bos. Causes and consequences of comorbidity: a review. *Journal of clinical epidemiology*, 54(7):661–674, 2001.

[34] Irina Gorodetskaya, Stefanos Zenios, Charles E Mcculloch, Alan Bostrom, Chi-Yuan Hsu, Andrew B Bindman, Alan S Go, and Glenn M Chertow. Health-related quality of

life and estimates of utility in chronic kidney disease. *Kidney international*, 68(6):2801–2808, 2005.

[35] Antonin Guttman. *R-trees: a dynamic index structure for spatial searching*, volume 14. ACM, 1984.

[36] Lindsey R Haas, Paul Y Takahashi, Nilay D Shah, Robert J Stroebel, Matthew E Bernard, Dawn M Finnie, and James M Naessens. Risk-stratification methods for identifying patients for care coordination. *The American journal of managed care*, 19(9):725–732, 2013.

[37] Milos Hauskrecht and Hamish Fraser. Planning treatment of ischemic heart disease with partially observable markov decision processes. *Artificial Intelligence in Medicine*, 18(3):221–244, 2000.

[38] Jonathan E Helm, Mariel S Lavieri, Mark P Van Oyen, Joshua D Stein, and David C Musch. Dynamic forecasting and control algorithms of glaucoma progression for clinician decision support. *Operations Research*, 63(5):979–999, 2015.

[39] Amanda A Honeycutt, Joel E Segel, Xiaohui Zhuo, Thomas J Hoerger, Kumiko Imai, and Desmond Williams. Medical costs of ckd in the medicare population. *Journal of the American Society of Nephrology*, 24(9):1478–1483, 2013.

[40] Chi-yuan Hsu, David W Bates, Gilad J Kuperman, and Gary C Curhan. Blood pressure and angiotensin converting enzyme inhibitor use in hypertensive patients with chronic renal insufficiency. *American journal of hypertension*, 14(12):1219–1225, 2001.

[41] Chi-yuan Hsu, Feng Lin, Eric Vittinghoff, and Michael G Shlipak. Racial differences in the progression from chronic renal insufficiency to end-stage renal disease in the united states. *Journal of the American Society of Nephrology*, 14(11):2902–2907, 2003.

[42] William W Hung, Joseph S Ross, Kenneth S Boockvar, and Albert L Siu. Recent trends in chronic disease, impairment and disability among older adults in the united states. *BMC geriatrics*, 11(1):47, 2011.

[43] Rob J Hyndman and Anne B Koehler. Another look at measures of forecast accuracy. *International journal of forecasting*, 22(4):679–688, 2006.

[44] Kidney Disease Outcomes Quality Initiative et al. Kidney disease outcomes quality initiative (k/doqi) clinical practice guidelines for chronic kidney disease: evaluation, classification, and stratification. *Am J Kidney Dis*, 39(2 Suppl 1):S1–S266, 2002.

[45] Ajay Israni, Cynthia Korzelius, Raymond Townsend, and Douglas Mesler. Management of chronic kidney disease in an academic primary care clinic. *American journal of nephrology*, 23(1):47–54, 2003.

[46] Eric P Jack and Thomas L Powers. A review and synthesis of demand management, capacity management and performance in health-care services. *International Journal of Management Reviews*, 11(2):149–174, 2009.

[47] Jacek Jarmulak, Susan Craw, and Ray Rowe. Self-optimising cbr retrieval. In *Tools with Artificial Intelligence, 2000. ICTAI 2000. Proceedings. 12th IEEE International Conference on*, pages 376–383. IEEE, 2000.

[48] Sarbjit Vanita Jassal, Lilyanna Trpeski, Naisu Zhu, Stanley Fenton, and Brenda Hemmelgarn. Changes in survival among elderly patients initiating dialysis from 1990 to 1999. *Canadian Medical Association Journal*, 177(9):1033–1038, 2007.

[49] Eric S Johnson, Micah L Thorp, Robert W Platt, and David H Smith. Predicting the risk of dialysis and transplant among patients with ckd: a retrospective cohort study. *American journal of kidney diseases*, 52(4):653–660, 2008.

[50] Jovan Karamata. Sur une inégalité relative aux fonctions convexes. *Publications de l'Institut mathematique*, 1(1):145–147, 1932.

[51] Eamonn J Keogh and Michael J Pazzani. Scaling up dynamic time warping to massive datasets. In *European Conference on Principles of Data Mining and Knowledge Discovery*, pages 1–11. Springer, 1999.

[52] Young-Ju Kim and Chong Gu. Smoothing spline gaussian regression: more scalable computation via efficient approximation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 66(2):337–356, 2004.

[53] William R Klecka. *Discriminant analysis*. Number 19. Sage, 1980.

[54] John P Klein and Melvin L Moeschberger. *Survival analysis: techniques for censored and truncated data*. Springer Science & Business Media, 2005.

[55] Rafal Kustra and Adam Zagdanski. Incorporating gene ontology in clustering gene expression data. In *Computer-Based Medical Systems, 2006. CBMS 2006. 19th IEEE International Symposium on*, pages 555–563. IEEE, 2006.

[56] Kolodner J. L. *Case-Based Reasoning*. Morgan Kaufmann, San Mateo, CA.

[57] Martin J Landray, Jonathan R Emberson, Lisa Blackwell, Tanaji Dasgupta, Rosita Zakeri, Matthew D Morgan, Charlie J Ferro, Susan Vickery, Puja Ayrton, Devaki Nair, et al. Prediction of esrd and death among people with ckd: the chronic renal impairment in birmingham (crib) prospective cohort study. *American Journal of Kidney Diseases*, 56(6):1082–1094, 2010.

[58] Tamar Lasky and Paul D Stolley. Selection of cases and controls. *Epidemiologic reviews*, 16(1):6–17, 1994.

[59] Brian J Lee and Ken Forbes. The role of specialists in managing the health of populations with chronic illness: the example of chronic kidney disease. *BMj*, 339:b2395, 2009.

[60] Chris P Lee, Glenn M Chertow, and Stefanos A Zenios. A simulation model to estimate the cost and effectiveness of alternative dialysis initiation strategies. *Medical decision making*, 26(5):535–549, 2006.

[61] Chris P Lee, Glenn M Chertow, and Stefanos A Zenios. Optimal initiation and management of dialysis therapy. *Operations Research*, 56(6):1428–1449, 2008.

[62] HY Lee and KN Park. Methods for determining the optimal number of cases to combine in an effective case based forecasting system. *Korean Journal of Management Research*, 27(5):1239–1252, 1999.

[63] Claude Lefévre. Optimal control of a birth and death epidemic process. *Operations Research*, 29(5):971–982, 1981.

[64] Andrew S Levey, Josef Coresh, Ethan Balk, Annamaria T Kausz, Adeera Levin, Michael W Steffes, Ronald J Hogg, Ronald D Perrone, Joseph Lau, and Garabed Eknoyan. National kidney foundation practice guidelines for chronic kidney disease: evaluation, classification, and stratification. *Annals of internal medicine*, 139(2):137–147, 2003.

[65] Andrew S Levey, Josef Coresh, Tom Greene, Lesley A Stevens, Yaping Lucy Zhang, Stephen Hendriksen, John W Kusek, and Frederick Van Lente. Using standardized serum creatinine values in the modification of diet in renal disease study equation for estimating glomerular filtration rate. *Annals of internal medicine*, 145(4):247–254, 2006.

[66] Gang-Guo Li and Zheng-Zhi Wang. Evaluation of similarity measures for gene expression data and their correspondent combined measures. *Interdisciplinary Sciences: Computational Life Sciences*, 1(1):72–80, 2009.

[67] Liang Li, Brad C Astor, Julia Lewis, Bo Hu, Lawrence J Appel, Michael S Lipkowitz, Robert D Toto, Xuelei Wang, Jackson T Wright, and Tom H Greene. Longitudinal progression trajectory of gfr among patients with ckd. *American journal of kidney diseases*, 59(4):504–512, 2012.

[68] Yi Lin and Shoucheng Ouyang. *Irregularities and prediction of major disasters*. CRC Press, 2010.

[69] Donald AB Lindberg, Betsy L Humphreys, Alexa T McCray, et al. The unified medical language system. *IMIA Yearbook*, pages 41–51, 1993.

[70] Margaret W Linn, Bernard S Linn, and Shayna R Stein. Satisfaction with ambulatory care and compliance in older patients. *Medical care*, 20(6):606–614, 1982.

[71] Helmut Lütkepohl. *New introduction to multiple time series analysis*. Springer Science & Business Media, 2005.

[72] MS MacGregor and MW Taal. Clinical practice guidelines: Detection, monitoring and care of patients with ckd. *UK Renal Association*, 2011.

[73] Paolo Magni, Silvana Quaglini, Monia Marchetti, and Giovanni Barosi. Deciding when to intervene: a markov decision process approach. *International Journal of Medical Informatics*, 60(3):237–253, 2000.

[74] Lisa M Maillart, Julie Simmons Ivy, Scott Ransom, and Kathleen Diehl. Assessing dynamic breast cancer screening policies. *Operations Research*, 56(6):1411–1427, 2008.

[75] W Mendenhall. *Statistics for Management and Economics*. Duxbury Press, Boston.

[76] George A Miller. The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological review*, 63(2):81, 1956.

[77] Stefania Montani and Luigi Portinale. Accounting for the temporal dimension in case-based retrieval: A framework for medical applications. *Computational Intelligence*, 22(3-4):208–223, 2006.

[78] Fliss EM Murtagh, Neil S Sheerin, Julia Addington-Hall, and Irene J Higginson. Trajectories of illness in stage 5 chronic kidney disease: a longitudinal study of patient symptoms and concerns in the last year of life. *Clinical Journal of the American Society of Nephrology*, 6(7):1580–1590, 2011.

[79] Sreerama K. Murthy, Simon Kasif, and Steven Salzberg. A system for induction of oblique decision trees. *JAIR*, 2:1–32, 1994.

[80] Mary Natrella. Nist/sematech e-handbook of statistical methods. 2010.

[81] Peter Nemenyi. Distribution-free multiple comparisons. In *Biometrics*, volume 18, page 263. INTERNATIONAL BIOMETRIC SOC 1441 I ST, NW, SUITE 700, WASHINGTON, DC 20005-2210, 1962.

[82] Zlatana D Nenova and Jerrold H May. Determining an optimal hierarchical forecasting model based on the characteristics of the data set: Technical note. *Journal of Operations Management*, 44:62–68, 2016.

[83] A Nicolucci, D Cucinotta, S Squatrito, A Lapolla, N Musacchio, S Leotta, Letizia Vitali, Angela Bulotta, Paolo Nicoziani, Gustavo Coronel, et al. Clinical and socio-economic correlates of quality of life and treatment satisfaction in patients with type 2 diabetes. *Nutrition, Metabolism and Cardiovascular Diseases*, 19(1):45–53, 2009.

[84] Markus Nilsson and Mikael Sollenborn. Advancements and trends in medical case-based reasoning: An overview of systems and system development. In *FLAIRS Conference*, pages 178–183, 2004.

[85] Daniel J Niven, Luc R Berthiaume, Gordon H Fick, and Kevin B Laupland. Matched case-control studies: a review of reported statistical methodology. *Clin Epidemiol*, 4:99–110, 2012.

[86] National Institutes of Health et al. Kidney disease statistics for the united states. *Washington: NHI*, 2012.

[87] Ann M O'Hare, Adam Batten, Nilka Ríos Burrows, Meda E Pavkov, Leslie Taylor, Indra Gupta, Jeff Todd-Stenberg, Charles Maynard, Rudolph A Rodriguez, Fliss EM Murtagh, et al. Trajectories of kidney function decline in the 2 years before initiation of long-term dialysis. *American Journal of Kidney Diseases*, 59(4):513–522, 2012.

[88] Ann M O'Hare, Andy I Choi, Daniel Bertenthal, Peter Bacchetti, Amit X Garg, James S Kaufman, Louise C Walter, Kala M Mehta, Michael A Steinman, Michael Allon, et al. Age affects outcomes in chronic kidney disease. *Journal of the American Society of Nephrology*, 18(10):2758–2765, 2007.

[89] Nisha I Parikh, Shih-Jen Hwang, Martin G Larson, James B Meigs, Daniel Levy, and Caroline S Fox. Cardiovascular disease risk factors in chronic kidney disease: overall burden and rates of treatment and control. *Archives of internal medicine*, 166(17):1884–1891, 2006.

[90] Mark G Parker, Tod Ibrahim, Rachel Shaffer, Mitchell H Rosner, and Bruce A Molitoris. The future nephrology workforce: will there be one? *Clinical Journal of the American Society of Nephrology*, 6(6):1501–1506, 2011.

[91] Syed A Pasha and Philip HW Leong. Cluster analysis of high-dimensional high-frequency financial time series. In *Computational Intelligence for Financial Engineering & Economics (CIFEr), 2013 IEEE Conference on*, pages 74–81. IEEE, 2013.

[92] Meenal B Patwardhan, Gregory P Samsa, David B Matchar, and William E Haley. Advanced chronic kidney disease practice patterns among nephrologists and non-nephrologists: a database analysis. *Clinical Journal of the American Society of Nephrology*, 2(2):277–283, 2007.

[93] JR Quinlan. C4. 5: Programs for empirical learning morgan kaufmann. *San Francisco, CA*, 1993.

[94] Amy Rak, Rupesh Raina, Theodore T Suh, Vinod Krishnappa, Jessica Darusz, Charles W Sidoti, and Mona Gupta. Palliative care for patients with end-stage renal disease: approach to treatment that aims to improve quality of life and relieve suffering for patients (and families) with chronic illnesses. *Clinical Kidney Journal*, page sfw105, 2016.

[95] Christian H Reinsch. Smoothing by spline functions. *Numerische mathematik*, 10(3):177–183, 1967.

[96] Antonio A Sánchez-Ruiz and Santiago Ontanón. Least common subsumer trees for plan retrieval. In *International Conference on Case-Based Reasoning*, pages 405–419. Springer, 2014.

[97] Rajiv Saran, Yi Li, Bruce Robinson, Kevin C Abbott, LY Agodoa, John Ayanian, Jennifer Bragg-Gresham, Rajesh Balkrishnan, JL Chen, Elizabeth Cope, et al. Us renal data system 2015 annual data report: Epidemiology of kidney disease in the united states. *American journal of kidney diseases: the official journal of the National Kidney Foundation*, 67(3 Suppl 1):A7, 2016.

[98] Robinson B et al. Saran R, Li Y. Usrds annual data report: Epidemiology of kidney disease in the united states. national institutes of health, national institute of diabetes and digestive and kidney diseases. Technical report, United States Renal Data System, 2015.

[99] Mark J Sarnak, Andrew S Levey, Anton C Schoolwerth, Josef Coresh, Bruce Culleton, L Lee Hamm, Peter A McCullough, Bertram L Kasiske, Ellie Kelepouris, Michael J Klag, et al. Kidney disease as a risk factor for development of cardiovascular disease. *Circulation*, 108(17):2154–2169, 2003.

[100] Andrew J Schaefer, Matthew D Bailey, Steven M Shechter, and Mark S Roberts. Modeling medical treatment using markov decision processes. In *Operations research and health care*, pages 593–612. Springer, 2005.

[101] Roger C Schank and Robert P Abelson. Scripts, plans, goals, and understanding: An inquiry into human knowledge structures (artificial intelligence series). 1977.

[102] Rebecca J Schmidt, Jennifer R Domico, Michael I Sorkin, and Gerald Hobbs. Early referral and its impact on emergent first dialyses, health care costs, and outcome. *American journal of kidney diseases*, 32(2):278–283, 1998.

[103] Gideon Schwarz et al. Estimating the dimension of a model. *The annals of statistics*, 6(2):461–464, 1978.

[104] David B Seligson, Steve Horvath, Tao Shi, Hong Yu, Sheila Tze, Michael Grunstein, and Siavash K Kurdistani. Global histone modification patterns predict risk of prostate cancer recurrence. *Nature*, 435(7046):1262–1266, 2005.

[105] Steven M Shechter, M Reza Skandari, and Nadia Zalunardo. Timing of arteriovenous fistula creation in patients with ckd: a decision analysis. *American Journal of Kidney Diseases*, 63(1):95–103, 2014.

[106] Deb S Sherman, Douglas N Fish, and Isaac Teitelbaum. Assessing renal function in cirrhotic patients: problems and pitfalls. *American journal of kidney diseases*, 41(2):269–278, 2003.

[107] M Reza Skandari, Steven M Shechter, and Nadia Zalunardo. Optimal vascular access choice for patients on hemodialysis. *Manufacturing & Service Operations Management*, 17(4):608–619, 2015.

[108] Homer William Smith. *The kidney: structure and function in health and disease*. Oxford University Press, USA, 1951.

[109] VS Subrahmanian. Principles of multimedia database systems. 1998.

[110] Patrick W Sullivan and Vahram Ghushchyan. Preference-based eq-5d index scores for chronic conditions in the united states. *Medical Decision Making*, 26(4):410–420, 2006.

[111] Patrick W Sullivan, William F Lawrence, and Vahram Ghushchyan. A national catalog of preference-based scores for chronic conditions in the united states. *Medical care*, pages 736–749, 2005.

[112] Navdeep Tangri, Lesley A Stevens, John Griffith, Hocine Tighiouart, Ognjenka Djurdjev, David Naimark, Adeera Levin, and Andrew S Levey. A predictive model for progression of chronic kidney disease to kidney failure. *Jama*, 305(15):1553–1559, 2011.

[113] WO Tarnow-Mordi, C Hau, A Warden, and AJ Shearer. Hospital mortality in relation to staff workload: a 4-year study in an adult intensive-care unit. *The Lancet*, 356(9225):185–189, 2000.

[114] Anjali Tiwari, Chin-Lin Tseng, Elizabeth FO Kern, Miriam Maney, Donald R Miller, and Leonard Pogach. Facility variation in utilization of angiotensin-converting enzyme inhibitors and angiotensin ii receptor blockers in patients with diabetes mellitus and chronic kidney disease. *American Journal of Managed Care*, 13(2):73–80, 2007.

[115] Tanvir Chowdhury Turin, Marcello Tonelli, Braden J Manns, Pietro Ravani, Sofia B Ahmed, and Brenda R Hemmelgarn. Chronic kidney disease and life expectancy. *Nephrology Dialysis Transplantation*, 27(8):3182–3186, 2012.

[116] Deborah P Waldrop and Abbie M Kirkendall. Comfort measures: a qualitative study of nursing home-based end-of-life care. *Journal of palliative medicine*, 12(8):719–724, 2009.

[117] Daniel E Weiner and Stephen L Seliger. Cognitive and physical function in chronic kidney disease. *Current opinion in nephrology and hypertension*, 23(3):291, 2014.

[118] D Randall Wilson and Tony R Martinez. Improved heterogeneous distance functions. *Journal of artificial intelligence research*, 6:1–34, 1997.

[119] Simon Wood. *Generalized additive models: an introduction with R*. CRC press, 2006.

[120] LeChauncy D Woodard, Cassie R Landrum, Tracy H Urech, Degang Wang, Salim S Virani, and Laura A Petersen. Impact of clinical complexity on the quality of diabetes care. *The American journal of managed care*, 18(9):508, 2012.

[121] JQ Xu, SL Murphy, KD Kochanek, and BA Bastian. Deaths: final data for 2013.

[122] Jean Yoon, Jennifer Y Scott, Ciaran S Phibbs, and Todd H Wagner. Recent trends in veterans affairs chronic condition spending. *Population health management*, 14(6):293–298, 2011.

[123] Qiu-Li Zhang and Dietrich Rothenbacher. Prevalence of chronic kidney disease in population-based studies: systematic review. *BMC public health*, 8(1):117, 2008.