Management of a Chronically Ill Population: An Operations Approach to Liver Cancer Screening

by

Elliot Lee

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Industrial and Operations Engineering)
in The University of Michigan
2016

Doctoral Committee:

Assistant Professor, Mariel S. Lavieri, Chair
Associate Professor Brian Denton
Assistant Professor Cong Shi
Assistant Professor Zeeshan Syed

ACKNOWLEDGEMENTS

First, I would like to thank my advisor, Dr. Mariel Lavieri, whose invaluable direction and meticulous revisions affected every presentation, every paper, and every document that ever carried my name. This work represents as much her work as it does mine, and I would have not enjoyed any amount of success without her.

Secondly, I would like to acknowledge Dr. Michael Volk who provided me with the opportunity to pursue this degree. I would especially like to thank him for his co-authorship and revisions of our papers. His participation in this research had little benefit for him, but provided me with much.

Thank you to my remaining committee members, Dr. Brian Denton, Dr. Cong Shi, and Dr. Zeeshan Syed, for your participation in my defense, as well as your comments and critiques of my work in the past.

I would like to thank my friends who helped me endure this long journey: Jivan Hawkinnchkreu, Brandon Pitts, Gregg Schell, Rose Figeuroa, and many others. Thank you for all your support.

My parents have provided untold amounts of physical, emotional, and financial support for my education, and any success in my life is theirs to share.

And lastly, I give thanks to God, who has sustained me when all other aforementioned sources of strength could not.

TABLE OF CONTENTS

CHAPTER

LIST OF FIGURES

Figure

LIST OF TABLES

Table

ABSTRACT

Management of a Chronically Ill Population: An Operations Approach to Liver Cancer Screening

by

Elliot Lee

Chair: Mariel Lavieri

We study how to perform medical surveillance of a population living with a chronic disease from an operations perspective. Our approach to the screening problem is the first to combine aspects of patient specific risk factors, heterogenous disease progression, as well limited screening resources shared by the population. Using clinical data from liver cancer as a motivating example, we (1) provide a new characterization of individualized risk for liver cancer through a nested case-control match study, then (2) demonstrate the utility of that individual biological information in screening decisions through the design and testing of reinforcement learning techniques, and then (3) model the problem as a family of restless bandits to gain structural insights into the problem, as well as derive an optimal policy to screen patients. Ultimately, we provide novel methods of screening a chronically ill population which are superior to current practice by adopting principles from a broad spectrum of operations methods.

CHAPTER I

Introduction

1.1 Background

Chronic diseases, such as cancer, heart disease, and diabetes make up 7 of every 10 deaths in the United States (*National Center for Health Statistics* (2016)). The U.S. National Center for Health Statistics defines a chronic disease as a disease that lasts 3 months or more, and cannot be prevented by vaccines nor cured by medication (*National Center for Chronic Disease Prevention and Health Promotion* (2015)). The CDC estimates that chronic diseases make up 86% of all healthcare spending (*Centers for Disease Control* (2016)). Due to this overwhelming burden, American healthcare has been forced to change its focus from traditional reactive care of acute illnesses to the preventative care of chronic illnesses.

The nature of preventative care proposes new challenges to clinicians. Commonly, a large population might be living with a chronic illness, but only a fraction of those diseases might lead to adverse events. Each patient's disease progresses uniquely, and this requires a simultaneous surveillance of the population. Furthermore, this surveillance can often be expensive and limited in availability. This poses a new sequential decision making problem that combines challenges of personalized medicine with resource allocation.

1.2 Motivation

Separate elements of this problem have been studied in isolation. The medical community has begun to investigate the relationships between biological information and a patient's risk to provide the opportunity for personalized medicine (*Lok et al.* (2009)). The operations research community has started to design medical decision making policies which tailor surveillance to optimize a single patient's health outcomes (*Zhang et al.* (2012), *Underwood et al.* (2012), *Maillart et al.* (2008)). However, in order to address the modern challenges of American healthcare, it still remains to incorporate these separate elements into a single, unifying framework for disease surveillance, where resources are both shared and limited.

We investigate the problem of hepatocellular carcinoma (HCC) screening as a prototypical example of a medical decision making problem with these challenges. There is a clear opportunity for improvement in the setting of HCC screening. The current recommended screening protocol in the United States is to screen all at-risk patients once every six months (*Bruix and Sherman* (2005)). This is a clearly inefficient policy, as it (1) treats all patients equally, and (2) does not take resource usage into account. The basic motivation behind this work was that incorporating individual disease progression, as well as constraints of limited resources, should improve the efficiency of screening.

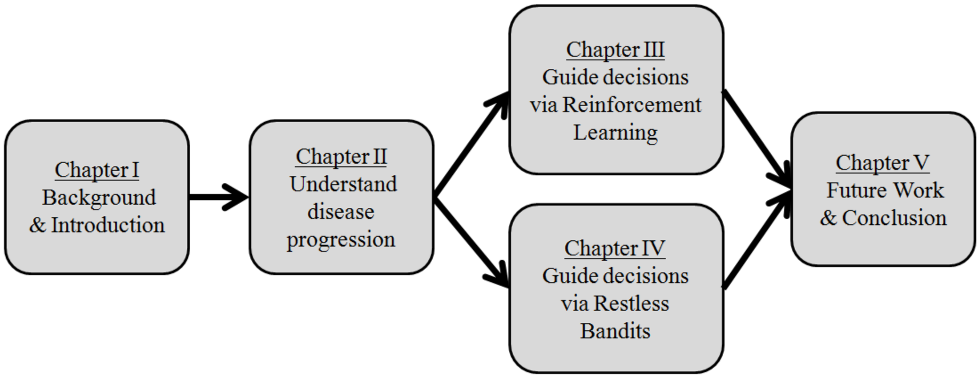


Figure 1.1: Overview of dissertation chapters.

Figure 1.1 gives an overview of the components of this thesis, and how they are organized in this dissertation. In Chapter II, we describe how we characterized and quantified each patient's individual risk for HCC. In Chapter III, we provide model-free algorithms to guide screening decisions. Chapter IV describes an analytical model for the HCC screening problem, as well as a tractable and optimal policy for this model. Lastly, we conclude with remarks and avenues for future work in Chapter V.

We now proceed to provide further details of the work accomplished in each chapter.

1.3 Chapter II: Characterizing Disease Progression

Chapter II seeks to understand how a patient's individual characteristics influence their progression to HCC. There are many known risk factors for HCC (*Lok et al.* (2009)), such as age, race, and smoking status. One risk factor, known as the alphafetoprotein (AFP), is particularly controversial. The AFP is a plasma protein mainly found in human fetuses, and whose exact function is still unknown (*Tomasi Jr* (1977)). Several studies point towards its usefulness in monitoring the health of a patient's liver (*Johnson* (2001), *Colli et al.* (2006)) but due to its noisy nature, the most recent standardized guidelines have advocated against the usage of AFP in diagnostic decisions (*Bruix and Sherman* (2005)). Prior to our study, the AFP remained a vaguely useful, but underutilized source of information.

We retrospectively analyzed patient data from the Hepatitis C Antiviral Long-Term Treatment against Cirrhosis (HALT-C) clinical trial. We performed a nested case-control matched study to find which patient characteristics correlated with the screening outcome of developing HCC.

We identified a statistically significant relationship between a patient's alphafetoprotein trends, and their risk of developing HCC. This relationship allowed us to translate past AFP observations into information that could be taken advantage of in future periods. The key to our discovery was to consider patterns of AFP over time, as opposed to only the most recent observation. In particular, (1) the rate of AFP rise over time, and (2) the fluctuations in AFP over time both proved to be more indicative of a patient's risk for HCC than the current AFP level alone. We found that incorporating the standard deviation of AFP and rate of AFP rise along with patient-specific risk factors improved the prognostic accuracy to an area under the receiver operating-characteristic curve (AUROC) of 0.81, compared to 0.76 when using the most recent AFP alone.

With a quantified relationship between AFP patterns and patient risk, we had discovered a key relationship that could be exploited. We now had the means to translate past AFP observations into predictions of HCC, which would allow us to make advantageous surveillance decisions in the future.

Key contributions:

- We quantify the AFP as a prognostic factor for HCC. (1) Rate of rise of AFP, and (2) fluctuations in AFP are much stronger predictors of HCC than the most recent level of AFP alone.
- We calibrate and validate our models with large clinical trial data.

- This work was published in Clinical Gastroenterology and Hepatology (*Lee et al.* (2012a)).

1.4 Chapter III: Reinforcement Learning Based Policies

Given the groundwork laid in Chapter II, in Chapter III we study how to allocate constrained screening resources across a population at risk for developing a disease by utilizing past AFP observations. The goal of this model is to increase the number of positive screening detections over a finite horizon using limited resources. The fundamental challenge is that a patient's risk of developing the disease depends on his/her AFP dynamics. However, knowledge of these dynamics must be learned over time. This greatly complicates the decision of how to allocate screening resources.

In Chapter III, we designed three classes of reinforcement learning policies designed to address the problem of simultaneously gathering and utilizing information across multiple patients. Reinforcement learning enjoys the advantages of making no assumptions about the system being tested, treating patients and their underlying disease as a "black box". With less assumptions, we were able to study the broader picture of the screening problem, albeit without any guarantees of optimality.

To test these newly designed policies, we created a case study based upon the screening for Hepatocellular Carcinoma (HCC). In this case study, a simulation was built which could gauge the performance of any hypothetical screening policy upon historical patient data. The purpose of this simulation was two-fold: Firstly, although we designed the mathematical structure of each reinforcement learning policy, the values of each policy's parameters remained open-ended. We employed the Indifference Zone Method (*Dudewicz and Dalal* (1975)) to optimize these parameters through simulation. Secondly, the resulting tuned policies were then tested within the simulation to gauge their performance, and to compare them against current practice.

We found that the best performing policy enjoyed a 8.6% increase in performance over current practice, while using the same amount of resources. Alteratively, the same best performing policy demonstrated a 16.5% reduction in screening costs, while detecting the same number of early stage cancers as current practice. Lastly, we conclude this chapter by studying how the benefits of learning-based screening policies differ across various levels of resource scarcity.

Key contributions:

- We provide proof of concept that current practice is vastly suboptimal; it can be outperformed by utilizing individual biological information.
- We establish that our problem is essentially a learning problem, as demonstrated by the success of reinforcement learning algorithms.
- The work presented in this chapter was published in *Lee et al.* (2012b) (simulation model) and *Lee et al.* (2012a) (reinforcement learning model).

1.5 Chapter IV: Restless Bandit Based Policies

In Chapter IV, we remain in the problem setting of a screening clinic where each patient's disease evolves stochastically, and there are limited screening resources shared by the population. Building upon the success of learning approaches in Chapter III, we model the problem as a family of restless bandits, with each patient's disease progression evolving as a partially observable Markov Decision Process (POMDP). We chose to model this problem as a POMDP to undertake a more rigorous and analytic approach to the screening problem.

We derived an optimal policy for this problem, and reduce its structural complexity for ease of understanding, as well as computational complexity. From this policy, we discussed several managerial insights about what characterizes more effective screening. Next, we developed a heuristic to approximate this policy for real-world implementation. In small-scale testing, this heuristic is as accurate as an exact solution over 99% of the time, while requiring 30% less computational time.

To calibrate and validate our work, we used two independent datasets: (1) the HALT-C trial, and (2) patient records from the University of Michigan Hospital, manually collected by our team with IRB approval (HUM00088566). The former was used to to train the parameters of the optimal policy, and the latter to build a computer simulation to act as a testbed for said policy. Over several clinic scenarios and iterations, we are able to show that our policy detects 22% more early stage cancers than current practice, while using the same amount of resource expenditure.

This chapter represents the culmination of our work. Not only do we establish an implementable policy for the medical community, we have contributed a rare example of a provably optimal solution to a restless bandit problem, thereby laying groundwork for this method to be used in other application domains.

Key contributions:

- We develop a novel approach to the screening problem which combines (1) individual biological information, (2) heterogenous disease evolution over time, and (3) shared and limited resources by the population.
- We derive a provably optimal solution to a restless bandit problem, which typically can only be solved through approximation.
- We provide an easily implementable form of our policy for use in practice.
- We calibrate and validate our models on two independent datasets.
- The work has been submitted for publication (*Lee et al.* (2016)).

1.6 Chapter V: Conclusions and Future Work

We conclude this thesis with Chapter V which summarizes this work, both in its accomplishments, shortcomings, and remaining questions of interest. We discuss potential areas of implementation for this work, and then suggest three natural avenues for future research.

CHAPTER II

Disease Progression

In this chapter, we discovered a relationship between individual patient characteristics and his/her risk of developing HCC. We hypothesized that patterns of alpha-fetoprotein (AFP) over time might offer prognostic information about HCC development.

This was a nested case-control study involving subjects from the Hepatitis C Antiviral Long-term Treatment against Cirrhosis (HALT-C) trial. 82 patients with HCC were matched 1:3 to controls without HCC, using bootstrapping to ensure similar follow-up time in both groups. The independent association with HCC was assessed for a) standard deviation of AFP, and b) rate of rise of AFP, in a multiple logistic regression which also included patient-specific risk factors such as age, platelet count, and smoking status.

In bivariable analysis, all three AFP metrics were associated with HCC development. Incorporating the standard deviation of AFP and rate of AFP rise, along with patient-specific risk factors improved the prognostic accuracy to an area under the receiver operating curve (AUROC) of 0.81, compared to 0.76 when using the most recent AFP alone.

This work was published in *Clinical Gastroenterology and Hepatology* in April 2013 (*Lee et al.* (2012a)), and has been cited 37 times since then. Beyond establishing a new relationship to be exploited for better informed decision making, it has provided clinicians with a new model of patient risk.

2.1 Background

Hepatocellular carcinoma (HCC) is now the second leading cause of cancer mortality worldwide (*Globocan* (2012)). The incidence in the United States is rising, in part due to the aging cohort of patients with chronic hepatitis C (*El-Serag et al.* (2003)). Although numerous treatment options exist, fewer than 50% of patients are diagnosed at an early enough stage to benefit from curative therapy (*Altekruse et al.* (2009)).

Patients with chronic hepatitis C and cirrhosis or advanced fibrosis are a group at high risk for developing HCC, with an annual incidence of 1-3% (*Davis et al.* (2010)). Current guidelines recommend surveillance of HCC with ultrasound and alpha-fetoprotein (AFP) every six months, but it is well recognized that this strategy detects only 60% of cases at early stage (*Singal et al.* (2009)). Therefore, better surveillance methods are needed.

One limitation of the current surveillance strategy is that each testing interval is viewed independently, without considering the history of prior testing. For example, most prior studies on AFP have assessed its performance based on the most recent value prior to HCC diagnosis, and have not considered any trends over time. In clinical practice, a rise in AFP is often viewed ominously, though limited data exists on the predictive value of this trend. A large fluctuation in AFP over time has been shown in some studies to confer increased risk, though it is unknown whether this pattern provides incremental prognostic value in addition to other proven risk factors (*Imaeda and Doi* (1992)). Therefore, the aim of this study was to determine whether patterns of AFP over time could be used to improve the accuracy of screening for HCC.

2.2 Description of the Data

We performed a nested case-control study of patients in the Hepatitis C Antiviral Long-term Treatment against Cirrhosis (HALT-C) trial. In HALT-C non-responders to prior antiviral therapy were enrolled and randomized to receive maintenance pegylated interferon or placebo. Inclusion criteria included advanced fibrosis (Ishak score ≥ 3) on biopsy, lack of suspicious mass on cross-sectional imaging, and AFP < 200. The HALT-C trial included 1025 patients followed for an average of 5.3 years.

Surveillance of patients in this trial was performed in three ways: Firstly, the patients were screened every 3 months for the first 3.5 years, then every 6 months thereafter on a voluntary basis. At each screening visit, the level of alpha-fetoprotein (AFP) concentration in the blood was measured. Secondly, each patient underwent an ultrasound imaging approximately every 6-12 months. Thirdly, a liver biopsy was performed on all trial participants at 1.5 and 3.5 years into the trial. HCC was diagnosed by cross-sectional imaging and biopsy. Tumors were staged based on the modified United Network of Organ Sharing TNM system. Early HCC was defined as tumor stage T1 (single lesion <2 cm in diameter) or T2 (single lesion between 2 and 5 cm or no more than 3 lesions each $<$ 3cm in diameter). As the HALT-C data are now de-identified and publicly available, the current study was exempt from IRB review at our institution.

2.3 Methods

HCC cases were matched to controls without HCC in a 1:3 ratio by length of follow-up time. This was accomplished by dividing the duration of the HALT-C trial into 10 mutually exclusive intervals of equal length. Cases and controls were matched by sampling from the patients who fell into the same interval of follow-up time. This criteria was chosen for matching in order to exclude bias caused by subjects with longer follow-up time having higher cumulative probability of HCC development, as well as more AFP values available. This matching process was repeated 1,000 times through bootstrapping, generating a new nested case-control group in each iteration. Any cases found to not have at least three candidate controls for matching were discarded from that iteration. Within each iteration, the cases were sequentially matched to controls without replacement. For this reason, the order in which cases were matched was randomized each time, so as to not give any case a higher likelihood of being discarded. However, replacement was used in between iterations, allowing for each matched case-control set to be viewed as independent of previous sets.

Using conditional logistic regression, we analyzed the association between development of HCC and the following two AFP patterns:

- Rate of rise, defined as $\frac{\text{AFP Final}-\text{AFP Baseline}}{\frac{\text{(Follow up duration)}}{\text{90 days}}}$ where AFP baseline is the patient's first recorded AFP reading subsequent to randomization in ng/ml, AFP final is either (a) the patient's last recorded AFP if they did not develop HCC, or (b) the last recorded AFP prior to diagnosis if they did develop HCC. Follow up duration is defined to be the difference between the dates of the two aforementioned AFP readings.

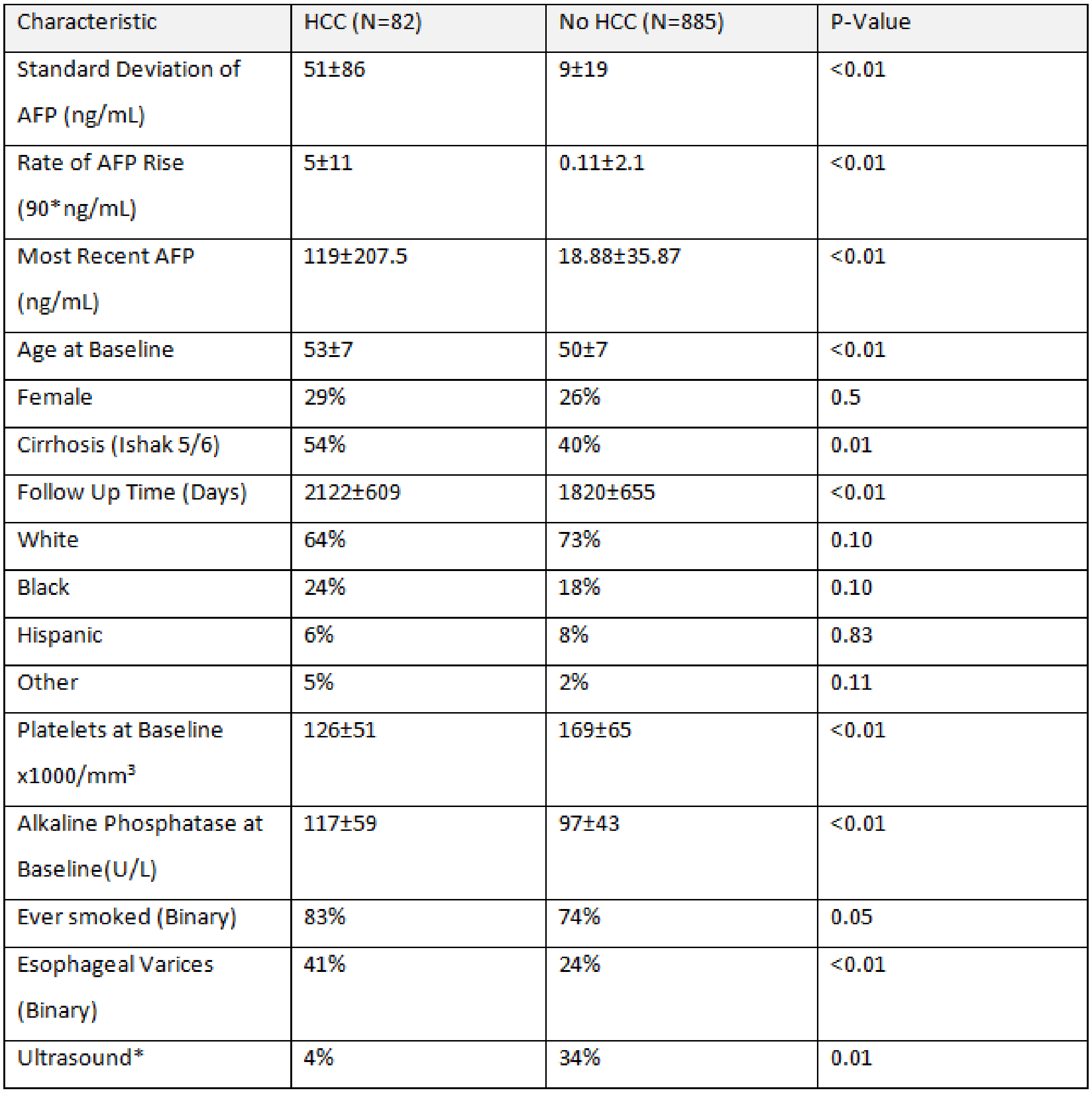
- Standard deviation of AFP, defined as the standard deviation of all AFP values recorded within each patient. The metric of standard deviation was chosen to capture the pattern of variability in AFP. Variability has been previously identified as behavior typical in patients at higher risk by *Imaeda and Doi* (1992).

Next, the independent significance of each pattern was assessed in multiple logistic regression. The initial model included the two AFP patterns as well as the most recent AFP value, and clinical and demographic risk factors for HCC as described in *Lok et al.* (2009). Since multiple variations of AFP were being assessed, model collinearity was determined by calculating the Variance Inflation Factor (VIF). Mean value imputation was performed on any missing values ($< 0.05\%$ of the dataset).The final model was chosen by stepwise backwards elimination, removing variables with $p < 0.1$ or VIF> 10. The area under the receiver-operating characteristic (AUROC) curve value was used to compare accuracy of the final model (history model) to a model which included only baseline risk factors and most recent AFP (no history model). Bootstrapping was used to generate confidence intervals around the AUROCs, and to provide p-values for comparison between the models. All calculations were performed using R (v 2.15.1)

2.4 Results

Out of 1050 subjects randomized in the HALT-C study, 83 were omitted from the current analysis for having < 5 AFP values available. Among the 967 subjects remaining, 82 developed HCC during the study period. Table 2.1 shows the clinical and demographic characteristics of these 82 cases and the remaining 885 subjects without HCC. For continuous variables, the mean standard deviation is shown, with p-values from a 2-sample t-test. For binary variables, the proportion is shown, with p-values from Fisher's exact test. During the screening period (time from enrollment to HCC diagnosis or end of follow-up), subjects had a median of 18 (range 5-23) AFP tests performed.

After matching each HCC case to 3 controls by duration of follow-up, the follow-up time in HCC cases and controls was 1670 and 1693 when averaged over 1,000 bootstrapped samples. All subsequently reported output statistics are the average of each statistic over these bootstrap samples. The distributions of the three AFP metrics, namely a) most recent AFP, b) standard deviation of AFP, and c) rate of rise of AFP, are shown in Figure 2.1, Figure 2.2, and Figure 2.3.



Characteristic	HCC (N=82)	No HCC (N=885)	P-Value
Standard Deviation of AFP (ng/mL)	51±86	9±19	<0.01
Rate of AFP Rise (90*ng/mL)	5±11	0.11±2.1	<0.01
Most Recent AFP (ng/mL)	119±207.5	18.88±35.87	<0.01
Age at Baseline	53±7	50±7	<0.01
Female	29%	26%	0.5
Cirrhosis (Ishak 5/6)	54%	40%	0.01
Follow Up Time (Days)	2122±609	1820±655	<0.01
White	64%	73%	0.10
Black	24%	18%	0.10
Hispanic	6%	8%	0.83
Other	5%	2%	0.11
Platelets at Baseline x1000/mm^3	126±51	169±65	<0.01
Alkaline Phosphatase at Baseline(U/L)	117±59	97±43	<0.01
Ever smoked (Binary)	83%	74%	0.05
Esophageal Varices (Binary)	41%	24%	<0.01
Ultrasound*	4%	34%	0.01

Table 2.1: Characteristics of patients with and without HCC.

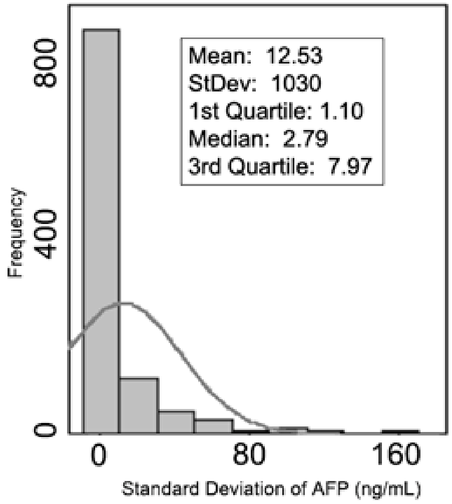


Figure 2.1: Distribution of the standard deviation of AFP over the studied population.

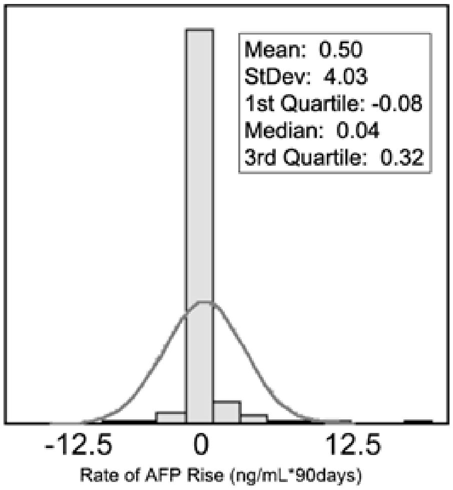


Figure 2.2: Distribution of rate of rise of AFP over the studied population.

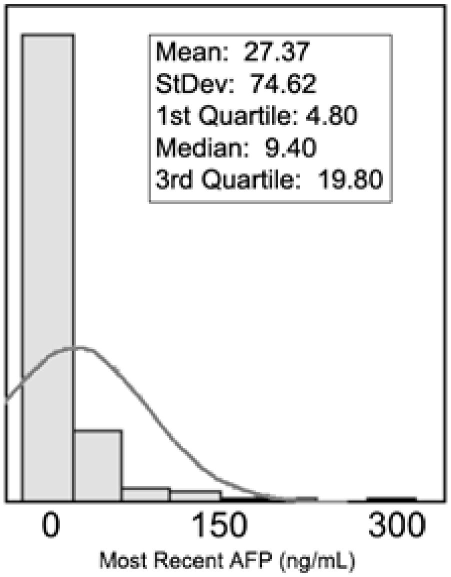
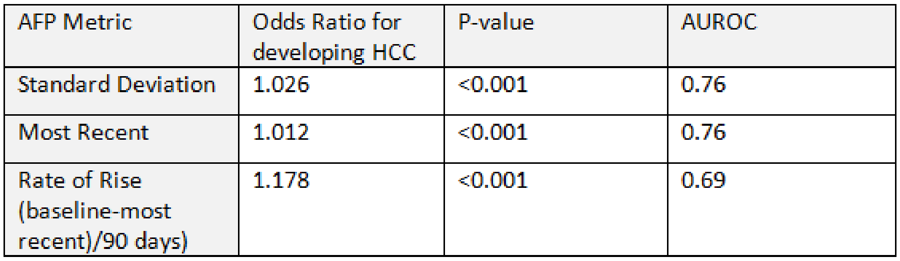


Figure 2.3: Distribution of most recent AFP over the studied population.

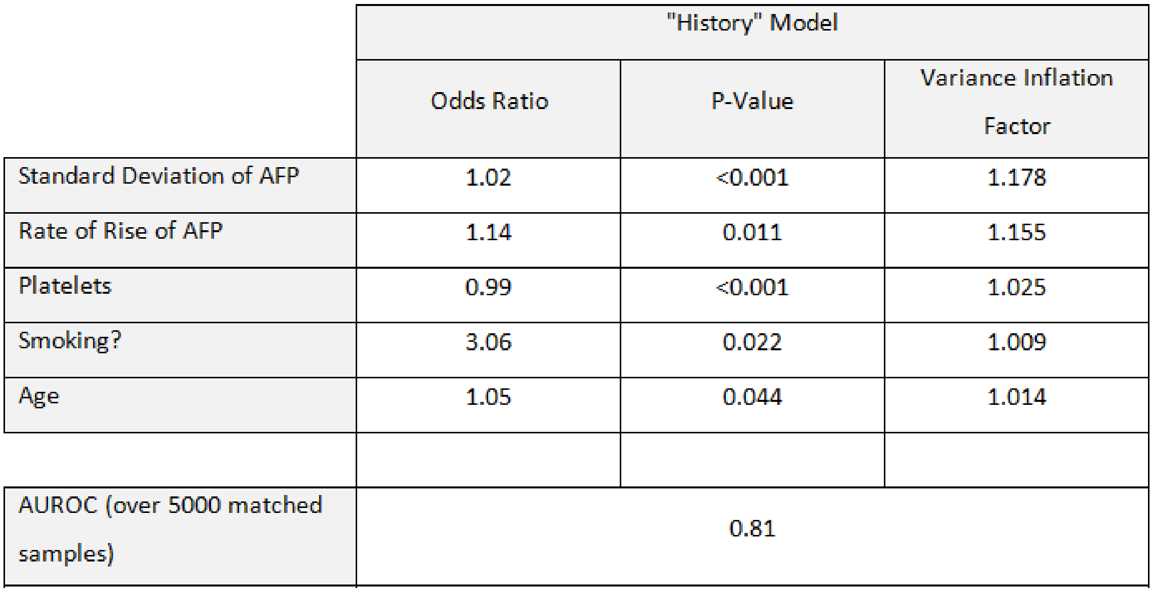
In bivariable analysis, all three metrics were associated with risk of HCC, as shown in Table 2.2.



AFP Metric	Odds Ratio for developing HCC	P-value	AUROC
Standard Deviation	1.026	<0.001	0.76
Most Recent	1.012	<0.001	0.76
Rate of Rise (baseline-most recent)/90 days)	1.178	<0.001	0.69

Table 2.2: Results of simple logistic regression of association between alpha-fetoprotein (AFP) metrics and hepatocellular carcinoma (HCC).

In initial multiple logistic regression, the most recent AFP and the rate of rise were moderately collinear, with VIF values of 3-4, and the most recent AFP was no longer significant with p=0.59 (but standard deviation remained statistically significant). Thus, the final model included standard deviation, rate of rise of AFP, as well as baseline age, platelet count, and smoking history, as shown in Table 3. The AUROC for this "history" model was 0.81, which was significantly higher than the "no history"



	"History" Model		
	Odds Ratio	P-Value	Variance Inflation Factor
Standard Deviation of AFP	1.02	<0.001	1.178
Rate of Rise of AFP	1.14	0.011	1.155
Platelets	0.99	<0.001	1.025
Smoking?	3.06	0.022	1.009
Age	1.05	0.044	1.014
AUROC (over 5000 matched samples)	0.81		

Table 2.3: Multiple logistic regression of patient variables associated with development of hepatocellular carcinoma.

model at 0.76 (p$<$ 0.001, 95% confidence interval of difference $[0.044, 0.046]$ obtained via a Wilcoxon Rank Sum test of the AUROC's calculated at each iteration for each model), as shown in Table 2.3 and Figure 2.4. Notice that the AUROC is higher for the "history" model than the traditional "no history" approach (p<0.001).

2.5 Sensitivity Analyses

Two sensitivity analyses were then conducted to test the strength of the findings across various assumptions. First, the analysis was repeated including only HCC cases with early stage disease and their controls matched by bootstrapping. Second, the final multiple logistic regression was repeated including ultrasound as a covariate, to determine whether the AFP patterns remained independently associated with development of HCC.

In sensitivity analysis, omitting cases with late-stage HCC from the analysis resulted an increase of less than 0.01 in AUROC for both the "history" and "no history"

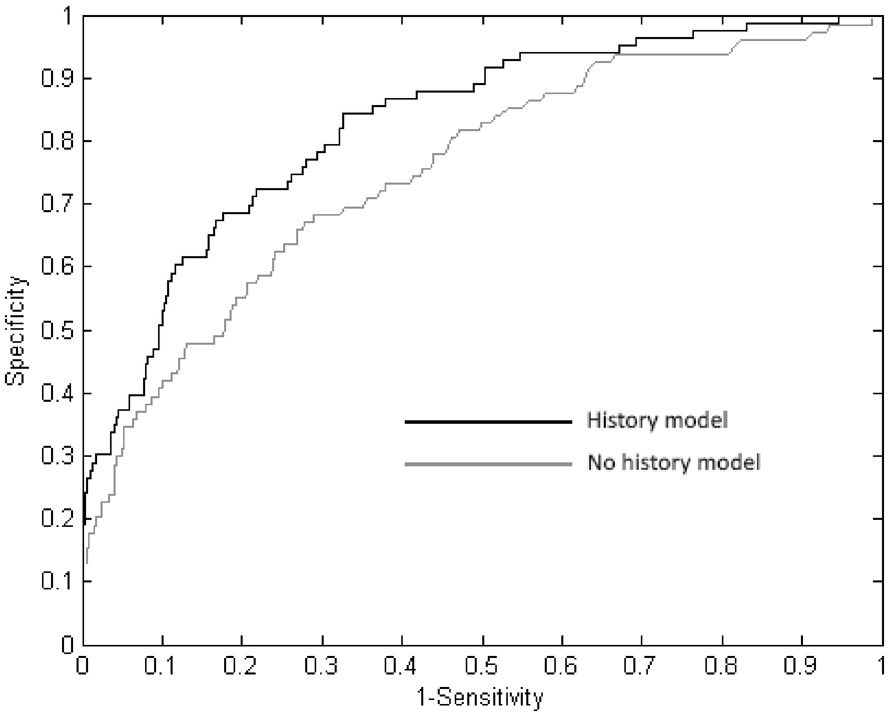


Figure 2.4: Comparison of AUROC is for the “history” and “no history” models.

models. Including ultrasound findings improved the AUROC of the “history” model from 0.81 to 0.86, but all variables remained statistically significant (data not shown). For sake of comparison, when ultrasound was analyzed alone, the AUROC was only 0.66.

2.6 Discussion

This study has shown that the pattern of AFP behavior over time in an individual patient is associated with development of HCC. Both the standard deviation and the rate of rise of AFP were independently associated with HCC, and incorporating these metrics along with patient-specific risk factors resulted in improved accuracy for HCC prediction when compared to the current method of using only the most recent AFP value. If confirmed in future studies, these findings could be used to develop individualized risk assessments for clinical use, which could then influence the frequency and type of further testing.

It is intuitively evident why the rate of rise of AFP might be associated with risk of HCC development, but what might explain the association with standard deviation? The biologic basis is unknown, but we theorize that fluctuations in AFP may reflect cycles of damage and regeneration within the liver, and that growth factors involved in regeneration could stimulate hepatocarcinogenesis. Future studies could perhaps measure fluctuation in growth factors in humans or mouse models of HCC, and these observations might thus shed light on mechanisms of HCC development.

The results of this study should be interpreted in the context of several limitations. This was a case-control study, albeit nested within a prospective cohort. Prior to considering these findings for clinical use, confirmation should be awaited from future prospective cohort studies with a priori hypotheses regarding patterns of AFP. While the two AFP metrics were independently associated with HCC development, their inclusion into a multivariable model with patient-specific risk factors only increased the AUROC from 0.76 to 0.81. This is a modest improvement, in part due to the AUROC of 0.76 with most recent AFP alone, which is higher than seen in most studies. Since this study was primarily aimed towards testing a hypothesis rather than validating a prognostic algorithm, we did not split the data to calculate a validation AUROC. Thus, the exact performance characteristics for these metrics should be viewed as preliminary. From a practical perspective these metrics will only be useful once a patient has been followed for at least 2 years in order for the patterns to emerge. Finally, this study included only patients with hepatitis C; it is unknown whether these associations would be present among patients with other chronic liver diseases. Despite these limitations, this is the first large study to demonstrate that patterns of AFP over time are independently associated with HCC development. The concept is appealing since it uses data already available, without extra cost or patient inconvenience.

In summary, we have shown that incorporating the history of prior AFP testing can improve accuracy for detecting HCC among patients with hepatitis C and advanced fibrosis or cirrhosis. Specific AFP patterns with prognostic significance included the degree of fluctuation, as measured by standard deviation, as well as the rate of rise over time. Future studies should validate these findings in a dedicated prospective cohort of patients with a variety of chronic liver diseases. In the future, these findings may be used to develop individualized testing strategies.

CHAPTER III

Reinforcement Learning Based Policies

In this chapter, we investigate the problem faced by a healthcare system wishing to allocate its constrained screening resources across a population at risk for developing a disease. A patient's risk of developing the disease depends on his/her biomedical dynamics. However, knowledge of these dynamics must be learned over time. Three classes of reinforcement learning policies are designed to address this problem of simultaneously gathering and utilizing information across multiple patients. We investigate a case study based upon the screening for Hepatocellular Carcinoma (HCC), and optimize each of the three classes of policies using the indifference zone method. A simulation is built to gauge the performance of these policies, and their performance is compared to current practice. We then demonstrate how the benefits of learning-based screening policies differ across various levels of resource scarcity and provide metrics of policy performance.

3.1 Introduction

Over six million Americans are estimated to be at risk for developing Hepatocellular Carcinoma (HCC) (*Wilkins et al.* (2010)). The incidence rate of HCC in the United States as of 2005 was 4.9 persons per 100,000, a rate which has tripled since 1975 (*Altekruse et al.* (2009)).

Early detection is highly correlated to patient health outcomes; less than 10% of the patients who are diagnosed with late-stage HCC survive beyond 5 years, whereas more than 50% of the patients diagnosed with early-stage HCC are disease free after 5 years (*Centers for Disease Control* (2010)). Therefore, the primary goal of HCC screening programs is to detect the development of the disease in the early stage.

We consider a screening program which has a limited screening capacity in each period. More precisely, the number of patients at risk for developing HCC outweighs the number of screenings available for administration in each period, and thus the problem of deciding which subset of the population to screen in each period arises.

This situation of a limited screening capacity could arise as a product of multiple scenarios. For instance, in highly overbooked screening clinics, the limited capacity results due to operational constraints of the clinics. Our approach addresses the challenge of finding a suitable screening program which accounts for their overbooked settings. Morever, capacity constraints could arise as a decision maker is faced with the problem of improving population-wide health outcomes without using additional resources beyond the current expenditure, such as in third world countries. This approach also becomes more relevant in the face of soaring costs in American Healthcare (*Bodenheimer et al.* (2009)).

The current recommended screening protocol in the United States is to screen all at-risk patients every six months. The full definition of the at-risk population for HCC is defined in (*Bruix et al.* (2001)). In this Chapter, we restrict that definition to be those patients with chronic Hepatitis C with advanced fibrosis, which are two key risk factors for HCC.

The main disadvantage of fixed interval screening is that it does not take into account the information learned sequentially, which differentiates patients at various levels of risk of developing HCC. More intelligible behavior would entail allocating resources according to the risk learned.

Reinforcement learning algorithms are well-suited to handle these problems of sequential learning under constrained resources, as it will be demonstrated in this paper. It is our goal to provide insight into what types of behaviors are characteristic of efficient screening for Hepatocellular Carcinoma.

3.1.1 HCC Screening

Two things occur when a patient is screened for HCC. An ultrasound image of the patient's liver is taken and examined by a doctor. The doctor will order more accurate tests (such as CT or MRI) if any suspicious features in the ultrasound suggest the development of a tumor.

Secondly, the patient's blood is measured for the alphafeto-protein (AFP) level. The AFP is a biomarker which is weakly correlated with HCC. Given that this correlation is weak, the AFP is not explicitly utilized in treatment or screening decisions (*Colli et al.* (2006)).

However, it has been shown that certain dynamics of the AFP, namely the standard deviation of a patient's AFP and the rate of rise of the patient's AFP, can significantly improve estimations of the patient's risk of developing HCC (*Lee et al.* (2012a)). The following list contains risk factors identified in that study (where * indicates values measured upon enrollment into the surveillance program)

- Age*
- Black ethnicity*
- Blood platelet count*
- Ever having been a smoker*
- Alkaline phosphatase*
- Presence of esophageal varices*

- Standard deviation of all AFP readings while under surveillance
- Rate of AFP rise over time, determined by ordinary least squares estimate of all AFP readings while under surveillance

To illustrate the nature of the last two risk factors, the following graph depicts sample AFP paths of patients in our dataset. Patients who eventually develop cancer tend to be characterized by AFP measurements which are wildly fluctuating, while simultaneously trending upwards. On the other hand, patients who do not develop HCC tend to have more stable, predictable AFP measurements.

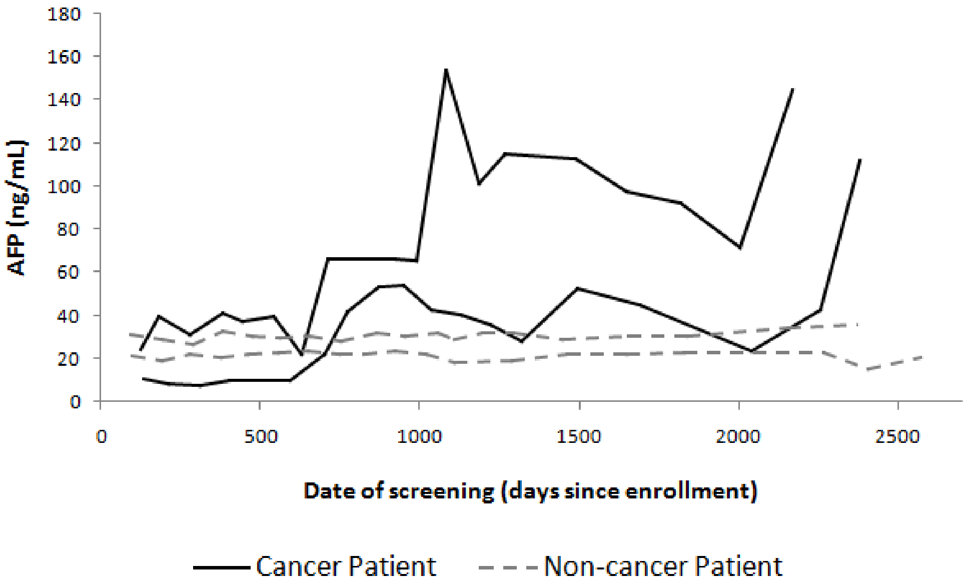


Figure 3.1: Sample AFP progressions for 2 patients who did and did not develop cancer.

Intuitively, detection rates could be maximized by shifting resources towards high risk patients, away from low risk patients. However, re-allocating resources according to patients' risk is not a straightforward problem due to the fact that, at any point in time, the decision maker holds imperfect knowledge of these dynamics. While baseline information about the patient provides some initial knowledge about the patient's risk of developing HCC, it is incomplete without understanding the patient's AFP dynamics. The challenge lies in the fact that knowledge of the AFP dynamics must be observed over multiple visits. Therefore, a sequential stage learning problem arises in the question of how to simultaneously gather and utilize this knowledge with limited resources.

3.2 Relevant Literature

The mathematical analysis, design, and optimization of medical screening policies is expansive, and has seen significant proliferation in the past decade. Comprehensive surveys of this field include *Stevenson* (1995), *Alagoz et al.* (2011), *Knudsen et al.* (2007), and *Pierskalla and Brailer* (1994).

Research in the field varies in goals, such as the cost-effectiveness analysis of existing screening programs (*Goldie et al.* (2004), *Frazier et al.* (2000), *Leshno et al.* (2003)) the cost-effectiveness analysis of proposed hypothetical screening programs (*Harper and Jones* (2005), *Davies et al.* (2002), *Kulasingam et al.* (2008)), and the optimization of new screening programs (*Preston and Smith* (2001), *Hanin et al.* (2001), *Lee and Zelen* (2003), *Lee and Zelen* (2008), *Parmigiani et al.* (2002), *Tsodikov et al.* (2006)). Our work is of the latter type, as aforementioned recent medical discoveries have presented the opportunity for novel policies to be considered.

Previous research also varies in how the objective is defined. Authors seek to maximize quality-adjusted life years gained (*Ayer et al.* (2012), *Chhatwal et al.* (2010), *Erenay et al.* (2014)), minimize negative health outcomes of a patient (*Parmigiani et al.* (2002), *Tsodikov et al.* (2006)), or minimize costs to the system (*Leshno et al.* (2003), *Myers et al.* (2000), *Frazier et al.* (2000)). Others yet consider a combination of the above, usually providing a pareto-optimal set of policies, as seen in *Rauner et al.* (2010), *Maillart et al.* (2008), and *Güneş et al.* (2004). As mentioned in the previous section, the stage at which HCC is diagnosed is heavily correlated to the patient's likelihood of survival. We analyze policies with respect to two performance metrics: (1) the proportion of all cancers detected in the early stage, and (2) the proportion of screening resources spent on cancer patients.

Depending on the goal and problem setting, a large number of methodologies have been proposed, including Markov chains (*Myers et al.* (2000), *Preston and Smith* (2001), *Goldie et al.* (2004), *Kulasingam et al.* (2008), *Harper and Jones* (2005)), simulation (*Davies et al.* (2002), *Frazier et al.* (2000), *Harper and Jones* (2005), *Clemen and Lacke* (2001)), Markov decision processes (*Chhatwal et al.* (2010), *Maillart et al.* (2008), *Leshno et al.* (2003)), partially observable Markov decision processes (*Zhang et al.* (2012), *Ayvaci et al.* (2012), *Zhang et al.* (2012), *Ayer et al.* (2009), *Erenay et al.* (2014)), hidden Markov chains (*Maillart et al.* (2008)), and other stochastic models (*Lee and Zelen* (2003), *Lee and Zelen* (2008), *Rauner et al.* (2010), *Helm et al.* (2015), *Schell et al.* (2014), *Piette et al.* (2013)). There are even less traditional methodologies used for analysis, such as game theory *Yaesoubi and Roberts* (2008) and queueing theory *Güneş et al.* (2004). For this work, simulation based optimization was chosen for its flexibility to analyze less traditional policies unable to be handled within the framework of the above methodologies.

Historically, researchers have often focused on simulating each patient's medical history by sampling from population parameters. This method is demonstrated in *Loeve et al.* (1999) in the simulation of colorectal cancer screening, and again in *Urban et al.* (1997) in the simulation of ovarian cancer screening. Our study differed primarily in choosing to use historical data to retroactively draw each patient's medical progression. Our approach has the advantage of having actual historical patients experience the proposed policy. It does not, however, have the robustness of a population-based simulation which can easily have a million unique, albeit fabricated, patients experience each policy.

Within the healthcare field, simulation based optimization has already seen much success. Zhang used a bisection search algorithm to find optimal capacity levels in long term care facilities (*Zhang and Puterman* (2013)). Romero employed built-in optimization packages with the simulation software Arena to find optimal operational parameters of a skin cancer clinic (*Romero et al.* (2013)). Dhamodharan utilized Monte Carlo sampling methods, along with their developed simulation model, to optimize the implementation of immunization services in rural areas (*Dhamodharan and Proano* (2012)). Simulation based optimization has not yet been used to optimize learning-based screening policies under constrained resources. We will rely on the indifference zone method to obtain optimal policies to address this question.

3.3 Problem Setting

We consider a healthcare system with a panel size of $i = 1, ..., n$ patients at risk of developing HCC. All patients at time $t = 0$ are known to begin in a cancer-free state. In our problem, the size of the population outweighs the number of screenings available, and it is the task of the decision maker (DM) to decide which subset of the population to screen. At each time $t = 0, 1, 2, .., T$, the DM can choose a subset of $k < n$ patients to screen.

Each patient i's risk of developing HCC can be measured by the following equation:

$$P(\text{HCC})_i = [1 + exp(-c_1 B_i - c_2 SD_i - c_3 RR_i)]^{-1} \tag{3.1}$$

- $P(\text{HCC})_i$ is the patient's lifetime cumulative probability of developing HCC,
- B_i is a vector of all static risk cofactors measured upon enrollment into surveillance (age, ethnicity, smoker, alkaline phosphatase, blood platelets, and esophageal varices),
- SD_i is the standard deviation amongst a patient's recorded AFP readings,
- RR_i is the least squares estimate for the rate of AFP rise over time amongst a

patient's recorded AFP readings,

- c_1 is a vector of the corresponding regression coefficients for all static risk factors, and
- c_2, and c_3 are regression coefficients for the AFP standard deviation and the rate of AFP rise over time.

Equation 3.1 calculates a patient's lifetime cumulative risk of developing HCC based upon several risk factors. For simplicity of notation, multiple risk factors which do not vary over time have been combined into a single quantity, B_i. The equation was determined through a nested case-control study in which risk factors for HCC development were assessed through conditional logistic regression in *Lee et al.* (2012a).

The knowledge of the DM at time t is captured by the state space variable:

$$(B_i, \hat{SD}_{i,t}, v_{i,t}, \hat{RR}_{i,t}, w_{i,t}) \tag{3.2}$$

$$\forall i = 1, .., n, \forall t = 0, 1, 2, ...T$$

Where the subscript t has been added to emphasize that the DM only holds estimates of these quantities for each patient i at each time t. $\hat{SD}_{i,t}$ is the sample standard deviation of all AFP observations for patient i up to, and including, time t. $\hat{RR}_{i,t}$ is the rate of AFP rise over time for patient i, estimated by ordinary least squares of all AFP readings up to, and including, time t. Note that this quantity can be negative. $v_{i,t}$ is the variance of the standard deviation estimate $\hat{SD}_{i,t}$, and $w_{i,t}$ is the variance of the rate of rise estimate $\hat{RR}_{i,t}$, both calculated by standard statistical methods.

The DM can utilize Equation 3.1 to obtain an estimate of patient i's risk at time

t by using the equation:

$$P(\hat{\text{HCC}})_{i,t} = [1 + exp(-c_1 B_i - c_2 \hat{SD}_{i,t} - c_3 \hat{RR}_{i,t})]^{-1} \tag{3.3}$$

Note that we have simply adapted Equation 3.1 by replacing unknown clinical quantities with sample estimates in order to approximate a patient's risk when said quantities are not perfectly known.

When the DM chooses to screen a patient i, two events occur in succession:

Firstly, the patient is revealed to be either (1) still cancer-free, (2) in early-stage cancer, (3) in late-stage cancer, or (4) dead, whether from cancer or other causes. If either outcome (2), (3), or (4) occur, the patient exits the system, and a new patient arrives in his/her place at time $t + 1$, thus maintaining a constant panel size n.

Secondly, the patient's AFP level is measured. This additional reading is then used to re-estimate the AFP related state space variables for that patient: $\hat{SD}_{i,t+1}$, $v_{i,t+1}$, $\hat{RR}_{i,t+1}$, and $w_{i,t+1}$.

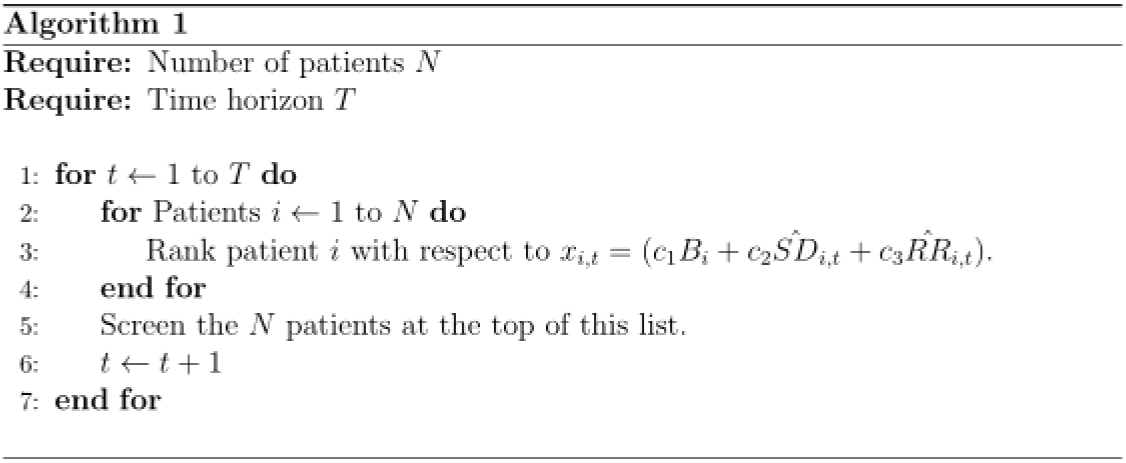
3.4 Reinforcement Learning Policies

3.4.1 Myopic Behavior

An intuitive policy to investigate would be to act myopically upon current estimates $P(\hat{\text{HCC}})_{i,t}$. The algorithm proceeds as follows:

Note that ranking patients according to $x_{i,t}$ is equivalent to ranking patients according to $P(\hat{\text{HCC}})_{i,t}$ because the function $(1+exp(-x))^{-1}$ is monotonically increasing in x.

Naturally, this will utilize the DM's knowledge accumulated thus far to maximize the number of cancers detected in the current stage. The downside to this policy is that it fails to expand the DM's knowledge set for future decisions by exploring other patients. This behavior is often referred to as "pure exploitation" and usually



```
Algorithm 1
Require: Number of patients $N$
Require: Time horizon $T$

for $t \leftarrow 1$ to $T$ do
    for Patients $i \leftarrow 1$ to $N$ do
        Rank patient $i$ with respect to $x_{i,t} = (c_1 B_i + c_2 \hat{SD}_{i,t} + c_3 \hat{RR}_{i,t})$.
    end for
    Screen the $N$ patients at the top of this list.
    $t \leftarrow t + 1$
end for
```

Figure 3.2: Pseudocode for myopic behavior algorithm.

performs suboptimally in various settings (*Sutton and Barto* (1998)).

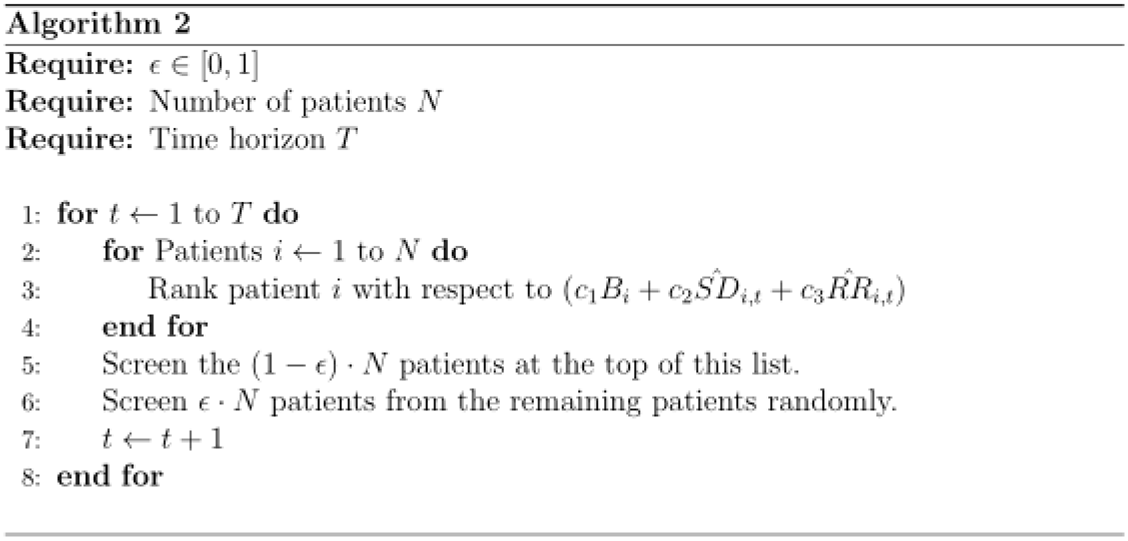
We consider three classes of reinforcement learning algorithms, each of which can be viewed as more intelligible modifications of this myopic behavior, with features to encourage exploration. We chose to study three distinct classes of algorithms to provide multiple perspectives on the value of learning within this problem setting, and thus increase the robustness of our conclusions.

3.4.2 ϵ-Greedy Strategies

The first class of reinforcement learning algorithms that we consider are ϵ-greedy strategies (*Watkins* (1989)). The algorithm (modified for our problem setting) proceeds as follows:

This strategy represents a slight modification of the myopic behavior as it reserves ϵ proportion of resources for exploration, and acts greedily with the remainder. It should be noted that myopic behavior is a special case of ϵ-greedy strategies, corresponding to $\epsilon = 0$.

Conversely, the special case of $\epsilon = 1$ is often referred to as "pure exploration", where all choices are made randomly. Both pure exploration and pure exploitation



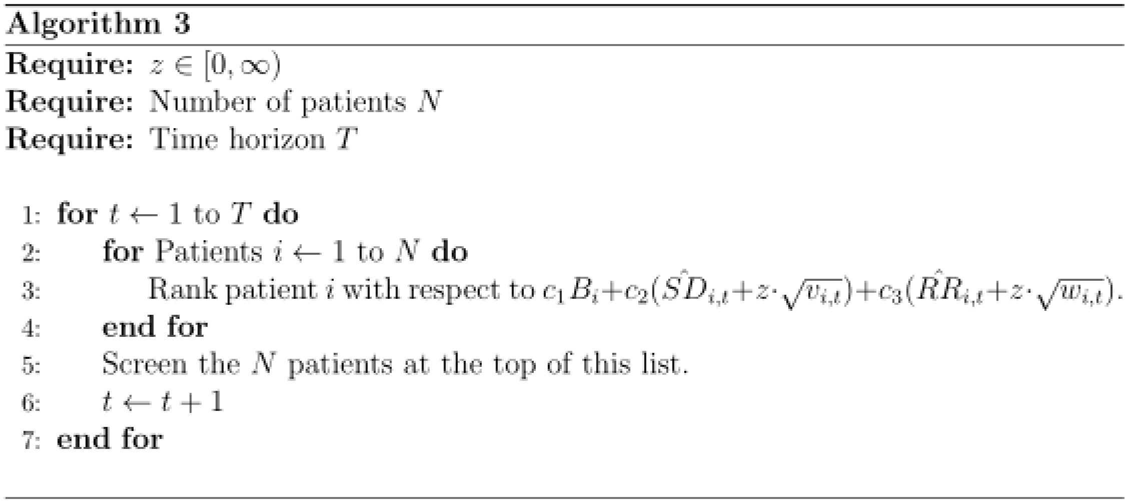
```
Algorithm 2
Require: $\epsilon \in [0, 1]$
Require: Number of patients $N$
Require: Time horizon $T$

for $t \leftarrow 1$ to $T$ do
    for Patients $i \leftarrow 1$ to $N$ do
        Rank patient $i$ with respect to $(c_1 B_i + c_2 \hat{SD}_{i,t} + c_3 \hat{RR}_{i,t})$
    end for
    Screen the $(1 - \epsilon) \cdot N$ patients at the top of this list.
    Screen $\epsilon \cdot N$ patients from the remaining patients randomly.
    $t \leftarrow t + 1$
end for
```

Figure 3.3: Pseudocode for epsilon greedy algorithm.

provide benchmark performances for other reinforcement learning techniques to compare against.

3.4.3 Interval Estimation Strategies

The second class of reinforcement learning algorithms that we consider are interval estimation strategies (*Kaelbling* (1993)). These algorithms proceed as follows:



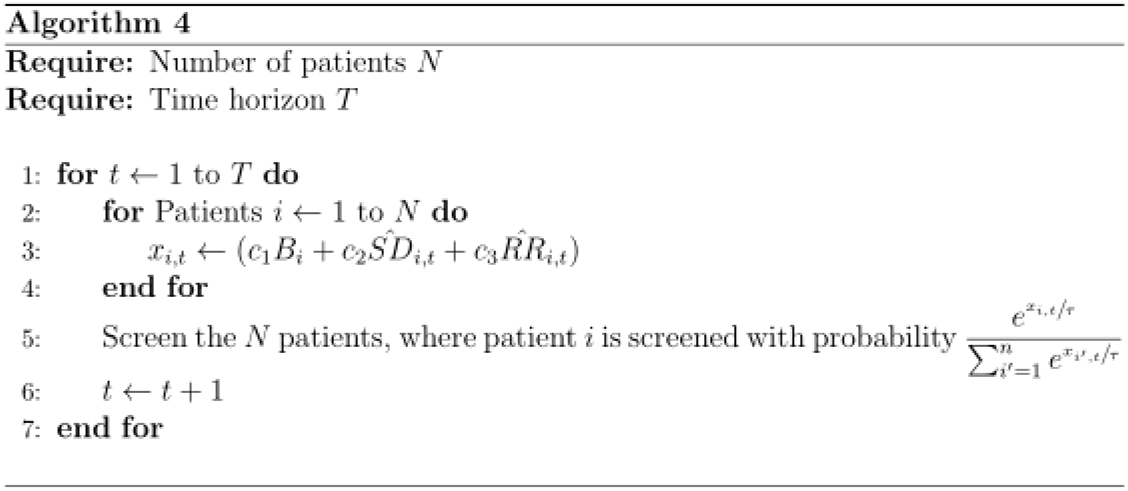
```
Algorithm 3
Require: $z \in [0, \infty)$
Require: Number of patients $N$
Require: Time horizon $T$

for $t \leftarrow 1$ to $T$ do
    for Patients $i \leftarrow 1$ to $N$ do
        Rank patient $i$ with respect to $c_1 B_i + c_2(\hat{SD}_{i,t} + z \cdot \sqrt{v_{i,t}}) + c_3(\hat{RR}_{i,t} + z \cdot \sqrt{w_{i,t}})$.
    end for
    Screen the $N$ patients at the top of this list.
    $t \leftarrow t + 1$
end for
```

Figure 3.4: Pseudocode for interval estimation algorithm.

Interval estimation is very similar in form to myopic behavior, but it encourages exploration by using a "distorted" risk score. Under very mild assumptions, patients whose risk of developing HCC is not currently known with high confidence are characterized by higher estimate variances, $v_{i,t}$ and $w_{i,t}$. Therefore by adding a multiplicative factor of the estimate variance to the risk score, the policy has artificially promoted patients with low knowledge up the list. The multiplicative factor, z, captures the incentive to explore as the relative importance of exploration. It should be noted that $z = 0$ corresponds to myopic behavior.

3.4.4 Boltzmann Exploration Strategies

The third class of reinforcement learning algorithms we consider are Boltzmann exploration strategies (*Luce* (1959)). These algorithms proceed as follows:



```
Algorithm 4
Require: Number of patients $N$
Require: Time horizon $T$

for $t \leftarrow 1$ to $T$ do
    for Patients $i \leftarrow 1$ to $N$ do
        $x_{i,t} \leftarrow (c_1 B_i + c_2 \hat{SD}_{i,t} + c_3 \hat{RR}_{i,t})$
    end for
    Screen the $N$ patients, where patient $i$ is screened with probability $\frac{e^{x_{i,t}/\tau}}{\sum_{i'=1}^{n} e^{x_{i',t}/\tau}}$
    $t \leftarrow t + 1$
end for
```

Figure 3.5: Pseudocode for Boltzmann exploration algorithm.

Boltzmann exploration gives all patients a positive probability of being screened, but with a probability which is weighted according to the DM's current risk estimates. The tuning parameter τ is known as the temperature. τ roughly captures the relative importance of exploration versus exploitation.

If $x_{i,t}$ is the original risk score, then the quantity $e^{x_{i,t}/\tau}$ can be thought of as a skewed risk score.

For example, when τ is very low, patients with only slightly differing original risk scores will have vastly differing skewed risk scores, due to the nature of the exponential function. Assessing patient risk according to the skewed risk score in this situation would be akin to artificially promoting exploitive behavior.

On the other hand, when τ is very high, the patients with vastly different original risk scores will have relatively similar skewed risk scores, again due to the nature of the exponential function. Therefore if we behave according to this skewed risk score, we have artificially promoted explorative behavior.

As $\tau \to \infty$, Boltzmann exploration approaches pure exploration.

3.5 Simulation

With three classes of candidate policies, a simulation was designed to serve as a testbed for empirical evaluation. The goal of this simulation was to receive proposed alternative screening policies as inputs, and then determine the number of cancers detected, as well as the resources used by each policy.

3.5.1 Description of the Data

The Hepatitis C Antiviral Long-term Treatment against Cirrhosis (HALT-C) trial included 1050 patients followed for an average of 5.3 years. Surveillance of patients in this trial was performed in three ways: Firstly, the patients were screened every 3 months for the first 3.5 years, then every 6 months thereafter on a voluntary basis. At each screening visit, the level of alpha-fetoprotein (AFP) concentration in the blood was measured. Secondly, each patient underwent an ultrasound imaging approximately every 6-12 months. Thirdly, a liver biopsy was performed on all trial participants at 1.5 and 3.5 years into the trial. Table 3.1 displays the relevant characteristics of the patients in the HALT-C dataset used in our study. For continuous variables, the mean ± standard deviation is shown, with p-values from a 2-sample t-test. For binary variables, the proportion is shown, with p-values from Fisher's exact test.

Characteristic	HCC (N=82)	NO HCC (N=885)	P Value
Age at Baseline (years)	53±7	50±7	< 0.01
Black (binary)	24%	18%	0.10
Platelets at Baseline x1000/mm^3	126±51	169±65	< 0.01
Ever Smoked (binary)	41%	24%	< 0.01
Alkaline Phosphataste at Baseline (U/L)	117±59	97±43	< 0.01
Esophageal Varices (binary)	4%	34%	0.01
Standard Deviation of AFP (ng/mL)	51 ± 86	9 ± 19	< 0.01
Rate of AFP Rise (90*ng/mL)	5±11	0.11±2.1	< 0.01

Table 3.1: Characteristics of patients with and without HCC.

Out of 1050 subjects, 83 were omitted from the current analysis for having < 5 AFP values available. Among the 967 subjects remaining, 82 developed HCC during the study period. During the screening period (time from enrollment to HCC diagnosis or end of follow-up), subjects had a median of 18 (range 5-23) AFP tests performed. It should be noted that the general American population has a lifetime risk for HCC of approximately 0.9%. Our dataset demonstrated a much higher cumulative incidence due to the fact that the eligibility requirements of the HALT-C trial included a history of chronic hepatitis C with advanced fibrosis, a key risk factor for HCC. As HCC screening is not currently recommended for the general public, it is appropriate to study the performance of these policies on this at-risk subset population.

3.5.2 Model

In this simulation, we record three statistics:

1. E, the total number of early stage cancers detected during the planning horizon,

2. L, the total number of late stage cancers detected during the planning horizon, and

3. X, the number of screenings spent on patients who would eventually develop cancer

The discrete event logic is graphically depicted in Figure 3.6. At all times, the simulation maintains two separate sets of patient data: (1) the simulation knowledge, and (2) the DM's knowledge. The latter is an incomplete subset of the former, which is further revealed through the DM's decisions of who to screen. The simulation begins at time $t = 0$ by using the Initial Panel Module to fill panel slots $i = 1, .., n$. Here, the simulation will randomly draw, with replacement, a patient history from the dataset to be this patient's simulated history. The result will be the creation of C, the set of patients who will develop cancer, and N, the set of patients who will not develop cancer in their lifetime. The DM's knowledge of these n patients at this point is limited to the baseline score, B. Also at this time, the DM knows every patient to be cancer-free, as the dataset which we used was a clinical trial whose enrollment criteria included being cancer-free at the beginning of surveillance.

Next the simulation runs the Policy Module, which receives the DM's knowledge of the current n patients as an input, and the DM chooses a subset K of patients to screen according to the current policy being evaluated. The value X is then increased by the number of screenings which were correctly spent on cancer patients, $|K \cap N|$.

For patients $i \notin K$ not chosen to be screened, the DM's knowledge of the patients will go unchanged until the next period. Patients chosen to be screened $i \in K$ enter the Imaging Module, which queries the simulation for the patient's current cancer state. If the patient never developed cancer during the course of HALT-C, the Imaging Module automatically outputs a cancer-free state. However, if the data indicates that this patient was detected to have a tumor of size $\bar{s}$ on date $\bar{t}$, we can estimate the tumor size s on the simulated date t according to the tumor doubling

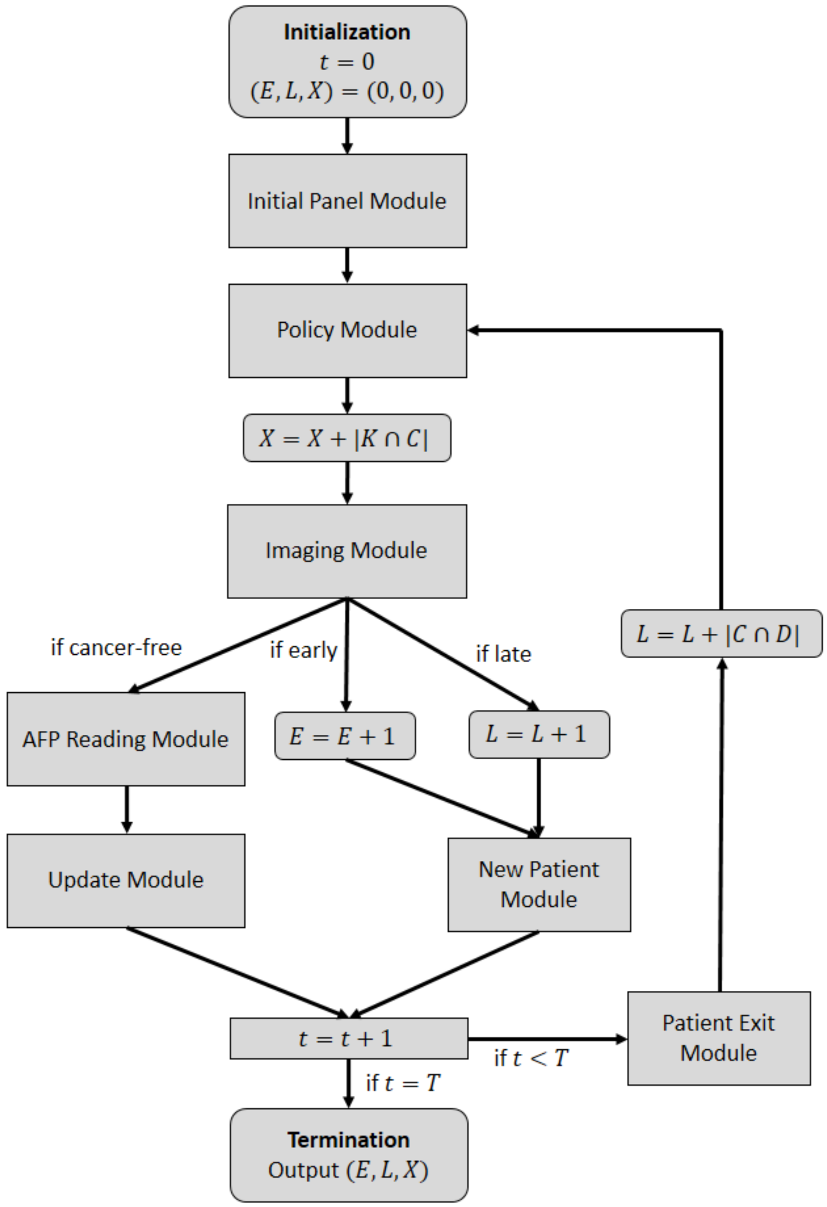


Figure 3.6: Discrete event simulation flow event logic.

time δ given in *Okada et al.* (1993) and the following doubling time equation:

$$s = 2^{\frac{t-\bar{t}}{\delta}} \cdot \bar{s} \qquad (3.4)$$

The Imaging Module then assigns a cancer state according to the following logic:

$$\text{State } = \begin{cases} \text{Early} & \text{if } t \geq \bar{t} \text{ and } 1 \leq s \leq 5 \\ \text{Late} & \text{if } t \geq \bar{t} \text{ and } 5 < s \\ \text{Cancer-Free} & \text{if Otherwise} \end{cases} \tag{3.5}$$

It should be noted that the Imaging Module assumes that all tumors between 1 cm and 5 cm in size are detected with perfect accuracy. This assumption of perfect accuracy is supported by the fact that it is standard procedure to follow up any ultrasound which reveals suspicious features with either a CT scan or MRI, both of which are highly accurate tests for tumor detection. This assumption could easily be relaxed in future work by incorporating the specificity and sensitivity of the respective tests.

The reason for assuming the tumor to be undetected on all dates $t < \bar{t}$ is to be as conservative as possible in our gauging in the performance of hypothetical policies. We could have utilized the same doubling time formula to determine the first date of tumor development, i.e the date at which the tumor was 1 cm in size. The tumor could theoretically be detected on any date after this first date of tumor development in the simulation. We however instead adopted our more stringent definition in order to make it as difficult as possible for the investigated screening policies to outperform the real-world detection rates.

If the patient is cancer-free, then the simulation assigns a new AFP reading for this patient through the AFP Reading Module. On the simulated date, the dataset is queried for a linear interpolation between the two closest AFP readings in date. This simulated AFP reading is then added to the DM's knowledge by appropriately updating all state space variables to reflect this new reading in the Update Module.

If the patient is assigned an early-cancer state, the value E is incremented by 1, and the patient is replaced according to the New Patient Module. The New Patient Module resets both the simulation knowledge of patient i to a new patient drawn from the dataset, as well as the DM's knowledge of patient i. Furthermore, we assume the DM receives a single AFP reading for this patient.

Similarly, if a patient is assigned a late-cancer state, the value L is incremented by 1, and the patient is replaced according to the same New Patient Module.

At the end of each period (with the exception of the final period), the Patient Exit Module is run to eliminate patients who depart the panel before the beginning of the next period. It does so by determining the departing subset $D \subset K$ whose time under surveillance in HALT-C has exceeded their duration in the simulation. These departures from surveillance include both outcomes of death or voluntary withdrawal from HALT-C. All patients who are eliminated are also replaced by a new patient before the beginning of the next period. In addition, all patients $i \in |D \cap C|$ who are eliminated through the Patient Exit Module also increment the penalty metric L by 1, as the policy has failed to identify a patient who had early stage cancer, and will now develop late-stage cancer outside of the simulated surveillance of that patient.

It should be noted that if the incoming replacement patient is drawn with equally likely probabilities, the population would become biased towards those patients with longer follow-up times. To maintain a patient panel which is probabilistically equivalent to that of HALT-C, the following probabilistic weights are used: if patients $j = 1, .., 967$ of HALT-C have follow-up times p_j, define $P = \sum_j p_j$. Patient q is chosen for replacement with probability

$$\frac{\frac{P}{p_q}}{\sum_j \frac{P}{p_j}} \tag{3.6}$$

This process of deciding who to screen, and simulating the outcome of those screenings repeats until the end of the planning horizon T. Upon termination the three final values of E, L, and X are reported to produce the following performance metrics:

1. The proportion of cancers detected in early stage $\frac{E}{E+L}$

2. The proportion of screenings spent on cancer patients $\frac{X}{K \times T}$

The usage of this historical simulation method is not without its shortcomings. This simulation is subject to inaccuracy if the HALT-C cohort consisted of many statistical outliers (with respect to disease progression sample paths). A typical simulation which fabricates patients from population parameters would not suffer from the misrepresentation of these statistical outliers. However, it should be noted that our simulation draws from a large dataset. Therefore, under some mild assumptions on the distribution of disease progression, it is unlikely that the HALT-C dataset is misrepresentative of the general American population. Additionally, HALT-C was a multi-center trial, with patients enrolled at 12 hospitals from all regions of America, further mitigating the possibility of a misrepresentative dataset.

3.5.3 Validation

Each iteration of the simulation begins by first randomly splitting the patients in the HALT-C data set into a training and validation set. The data in the training set is input into a conditional logistic regression to obtain coefficients c_1, c_2, and c_3 in equation (1). The patients in the validation set are used to populate the simulation. By this method, we avoid obtaining inflated estimates of policy performance, which would inevitably result by testing a policy on the same patients upon which the policy was built.

In testing the simulation outputs for agreement with real-world observations, the simulation predicted the number of early-cancers detected per year to be within 3% of the results observed in the HALT-C dataset.

The model was built with high face validity by discussing the discrete-event logic alongside people involved in the screening process. Our co-author, a practicing clinician at the University of Michigan Hospital, helped to validate our model. We also interviewed the receptionists at the hospital responsible for booking screenings for patients, ultrasound technicians who perform the screenings, and nurses at the blood draw clinic responsible for measuring and reporting the AFP back to the doctor. These interviews were meant to strengthen our understanding of the real-world flow of events at every step of the screening process.

Lastly, we unilaterally deviated simulation parameters to extreme scenarios for "sanity checks", and checked for intuitive agreement with expected outputs.

3.6 Tuning Parameter Optimization

To find the optimal levels of tuning parameters within each class of reinforcement learning algorithms, we employed the indifference zone method of *Dudewicz and Dalal* (1975). This discrete optimization via simulation method considers $m = 1, .., \ell$ discrete alternatives, where observations from population m are normally distributed $N(\mu_m, {\sigma^2}_m)$.

The procedure begins by sampling each of the ℓ alternatives and equal number of times, $n0$, via simulation. After the completion of this first stage, the sample variance of the simulation outcomes for each of the ℓ alternatives is calculated in order to determine the suitable number of additional samplings are needed for each alternative. After a second stage of sampling, a weighted average of the observations from the two stages is taken. The alternative m with the highest weighted averaged is declared to be within δ of the true best with probability P.

The main advantage of this method over the ranking and selection method developed in *Bechhofer* (1954) is that it does not require the assumption of ${\sigma^2}_1 = ... = {\sigma^2}_\ell := \sigma^2$, where σ^2 is known in advance. Initial analyses proved neither assumption to hold in this problem setting, thus encouraging the usage of the indifference zone method. Further details of this sampling procedure can be found in *Dudewicz and Dalal* (1975).

3.7 Results

3.7.1 Implementation Parameters

Decision epochs were chosen to be at equally spaced 90 day intervals, under clinical recommendations that under no circumstance would it be necessary to again screen a patient less than 90 days after being screened. The planning horizon was chosen to be $T = 30$ years arbitrarily by the authors, although this analysis was first performed over 10 years, and the difference in results were insignificant. Finally, a panel size of $n = 500$ was chosen to mimic the approximate size of the screening program at the University of Michigan Hospital. The indifference zone method was run at a confidence of $P = 95\%$ with an indifference zone width of $\delta = 0.25$. These parameters are the result of initial analyses on the computation time required, and were chosen to maximize accuracy given the resources available.

The calculations were performed using MATLAB v2013a's Parallel Computing Toolbox, at the University of Michigan's Center for Advanced Computing, on 24 computing cores (intel i7, 4GB RAM). A single iteration of the simulation requires approximately 30 seconds of computing time. Our simulation was run for approximately 400 iterations per each of the 34 policies, per each of the 5 resource constraint levels.

3.7.2 Policy Performance

The first analysis compared 5000 samples of current practice and pure exploration. Recall that pure exploration is equivalent to choosing a random subset of the population to screen in each period. Figure 3.7 is a histogram of the number of early cancers detected by these two policies. It is readily apparent that the two policies are highly similar in performance. This agrees with intuition, as both policies use no patient specific information, and treat all patients equally. Both policies can act as baselines to compare the performance of other policies against.

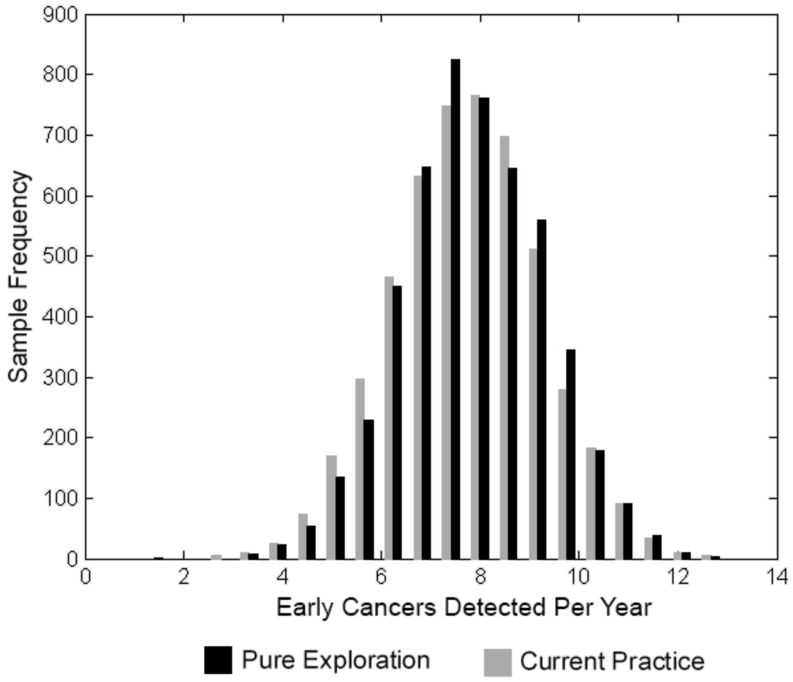


Figure 3.7: Comparison of performances of current practice and pure exploration.

We optimized each of the three classes of learning-based policies at five settings of resource constraints with respect to the proportion of cancers detected in early stage. If we let k be the number of screenings available to spend by the DM in each period, and n be the size of the patient panel, then $\frac{k}{n}$ is the measure of how constrained the problem is. We analyzed this problem at $\frac{k}{n}$ = 0.10,0.20, 0.30, 0.40, and 0.50 corresponding to five scenarios of varying resource scarcity.

ϵ-Greedy strategies were searched over the range $\epsilon = 0, 0.025, 0.050, 0.075, ..., 0.25$, Interval Estimation over $z = 1, 2, 3, ..., 10$, and Boltzmann Exploration over $\tau = .250, 0.275, .300, .325, ..., .750$. The search ranges were chosen at the discretion of the authors, after some initial experiments to find suitable candidates, and observing significant drop-offs in performance beyond those bounds. We optimized each of the three classes of policies at each of the five resource constraint levels. Table 3.2 shows the results of the tuning parameter optimization.

Resource Constraint Level $^k/_n$	ϵ-greedy ϵ	Interval Estimation z	Boltzmann Exploration τ
10%	0.025	1	0.250
20%	0.05	1	0.250
30%	0.10	1	0.300
40%	0.10	1	0.325
50%	0.25	5	0.400

Table 3.2: Optimal tuning parameters determined by the indifference zone method.

Recall that for each of the three classes of policies studied, higher tuning parameters represent more emphasis upon exploration. From these results, we can see that as resources become less constrained, the optimal balance between exploration and exploitation shifts towards exploration. Conversely, as resources become more constrained, greater exploitation is encouraged. The tendency to explore less in highly resource constrained settings is intelligible, as the DM does not have enough resources to learn anything of significance. Therefore, in highly resource constrained settings, it would be prudent to depend more upon the baseline risk information received by the DM when the patient entered screening.

Next we sought to compare the increase in performance of these optimized learning policies over current practice. Because the protocol of screening every patient every six months cannot be implemented in resource constrained settings, we created the equitable allocation policy to act as a policy equivalent in spirit to current practice. The equitable allocation policy uses whatever resources are available as fairly as possible, akin to current practice, by ensuring that all patients experience fixed interval screening at equal frequencies. Figure 3.8 displays our findings.

Our analysis determined Boltzmann exploration to be the policy which produced the most early stage cancer detections at every level of resource constraint. Furthermore, at every level of resource constraint, both myopic behavior and equitable allocation are dominated by any of the three reinforcement learning policies, thus

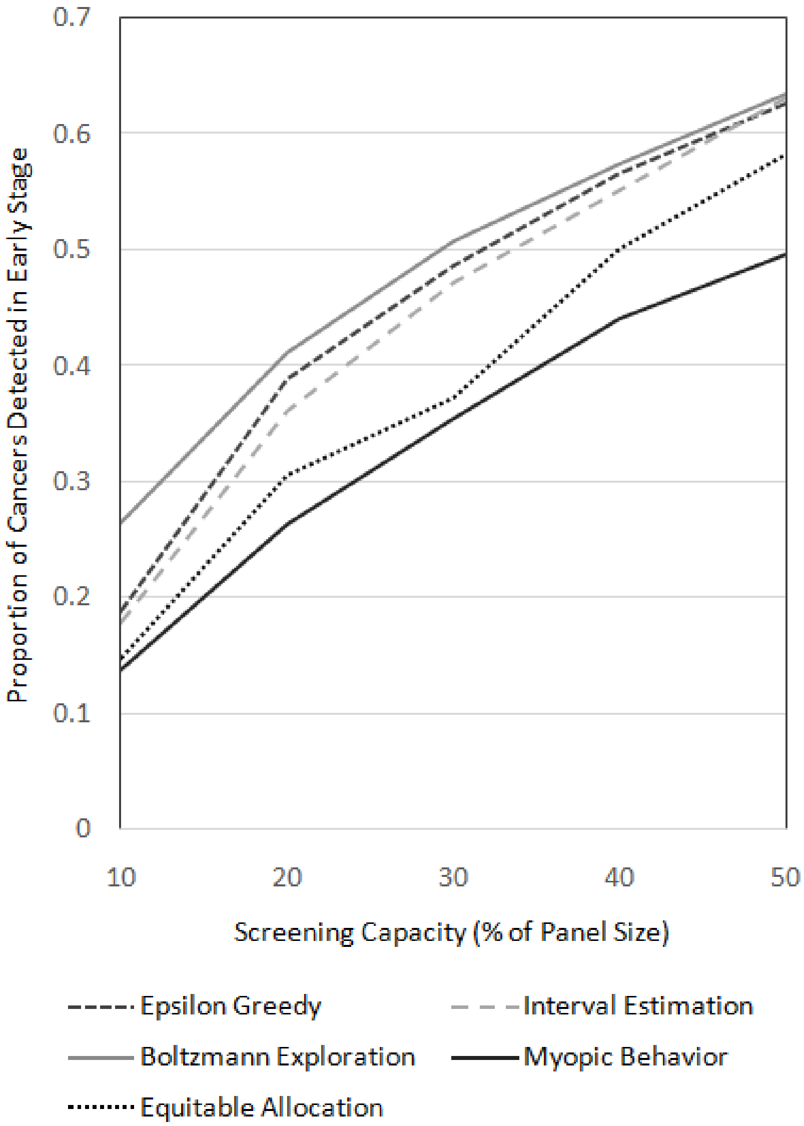


Figure 3.8: The performance of the optimal policies across various resource constraints.

demonstrating the importance of learning in our problem.

The most immediate benefit that can be drawn from these results is the increase in detection rates gained by switching to the best learning policy. Current practice screens 100% of the population every 180 days, so it stands to reason that its performance is equivalent to equitably screening 50% of the population every 90 days. The latter policy detects 58% of patients in early stage cancer. The best performing learning policy reaches a 63% detection rate at the same level of resource expenditure.

This represents a 8.6% increase in performance by switching from equitable allocation to the best performing learning policy.

Alternatively, we can analyze the cost-savings that can be achieved by switching to a learning based policy. The best performing learning policy only requires screening 41.75% of the population every 90 days to achieve the same level of performance as current practice, which screens 50% of the population every 90 days. This represents a 16.5% reduction in screening costs by switching from equitable allocation to the best performing learning policy, while maintaining the same level of performance.

As in many reinforcement learning applications, myopic behavior, or pure exploitation, is vastly suboptimal. This is due to the fact that myopic behavior can very easily become stuck in poor knowledge sets, and continue to incorrectly believe that certain subset of patients to be high risk. At 50% resource constraint setting, the best learning policy has a 27% increase in performance over myopic allocation of resources.

Another interesting feature of Figure 3.8 is the relative gap between the learning policies and the equitable policy seems to decrease in size as more resources become available. This agrees with intuition because the more scarce a resource becomes, the more benefit there stands to be gained by acting efficiently. The value of learning policies can be made apparent by comparing the best performing policy at each resource level with equitable allocation in Figure 3.9. Although the relative benefit of learning policies does generally decrease as more resources become available, it is still better than equitable allocation of resources.

We analyzed performance with respect to the percentage of screenings spent on cancer patients. This metric rewards a policy for screening the correct patients, even if those screenings did not immediately result in the detection of an early stage cancer. This metric is more concerned with the correct identification of high risk patients, and not necessarily the timing at which patients are screened. Although this metric

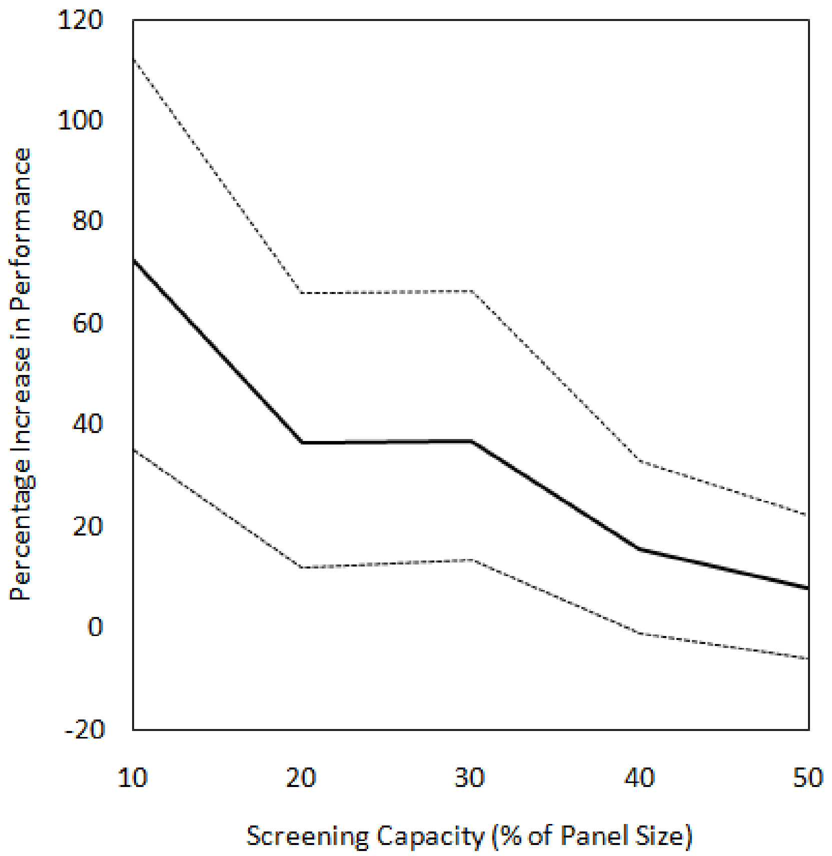


Figure 3.9: Increases in performance, with 90% sampling percentiles, across various resource constraints.

is less useful in a clinical setting, it is actually more closely aligned to the original purpose of the reinforcement learning algorithms studied. The results are displayed in Figure 3.10. Equitable allocation spends roughly 8% of its resources on cancer patients, across all resource levels. This is to be expected, as approximately 8% of the HALT-C dataset develops cancer, and thus approximately 8% of our simulated patient panel will eventually develop cancer. This metric more decisively demonstrates the advantage of learning-based policies over both equitable allocation, and myopic behavior. It also displays the decrease in relative benefit of learning-based policies as the capacity of the system increases.

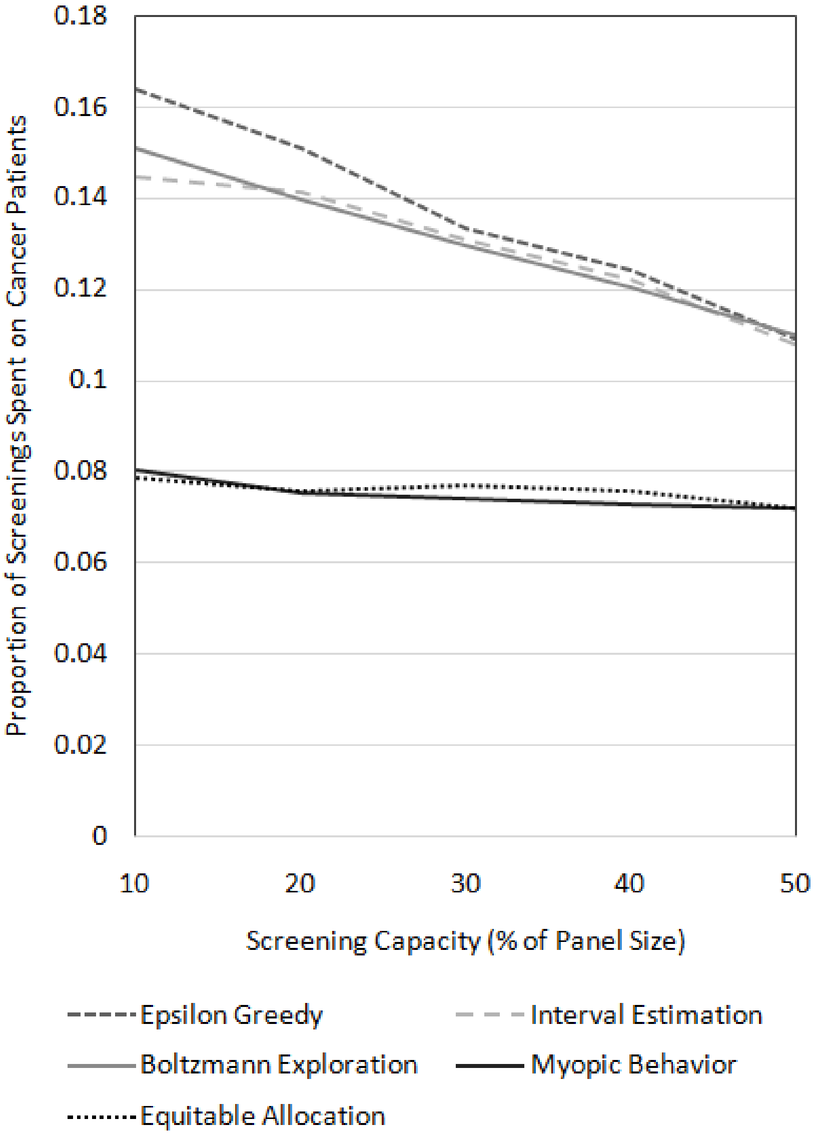


Figure 3.10: The proportion of screenings spent on cancer patients.

3.7.3 Policy Equity

Lastly, we investigated the affects of these policies from the patients' perspective. We sought to answer what a patient could expect to experience by participating in a screening regimen prescribed by our approach. We measured the 25th, 50th, and 75th percentile in days between subsequent screenings for each patient, then averaged these statistics across the population throughout the history of the simulation. These results are presented in Figure 3.11.

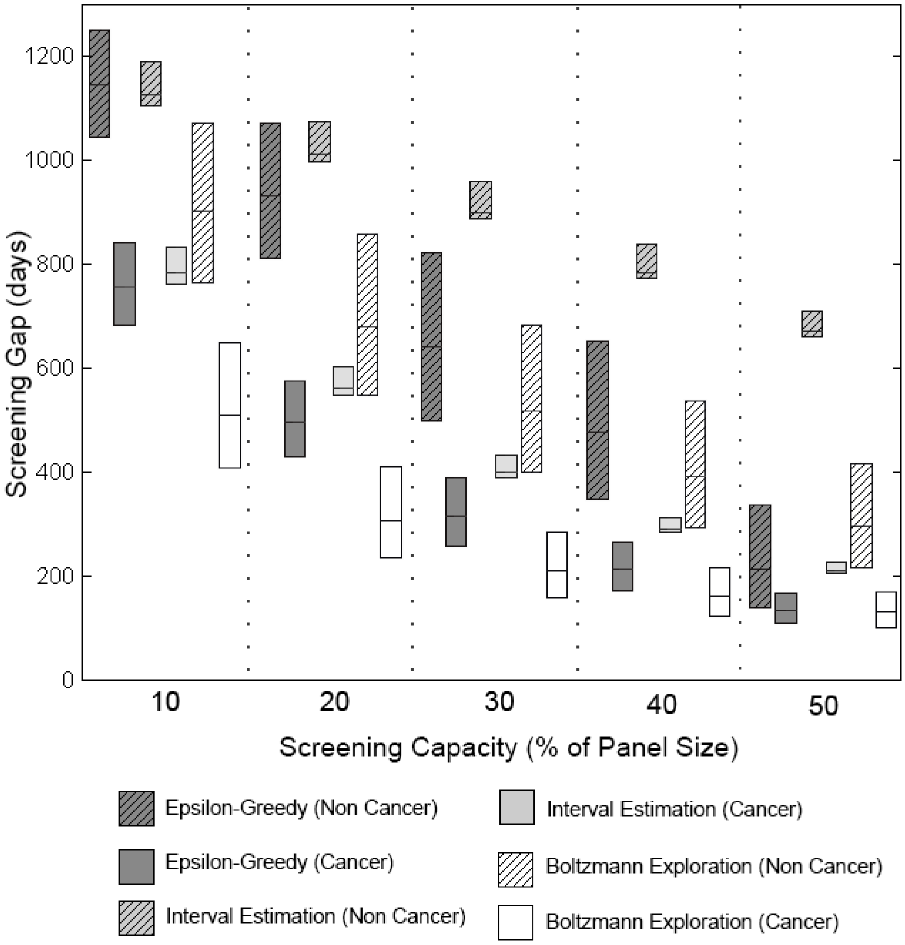


Figure 3.11: 25th, 50th, and 75th percentiles of screening gaps, averaged across cancer and non-cancer patients separately.

From this figure we can glean what kinds of screening policies these reinforcement learning methods require a specific patient to undergo. There is a distinct gap between the screening frequencies experienced by patients who do and do not develop cancer. This holds for every type of policy, at every resource constraint level.

Under current practice, the screening capacity is 50%, and both cancer and non cancer patients alike can expect to be screened once every 180 days. In the same setting, the average cancer patient being screened according to the optimal Boltzmann exploration policy can expect to have a median screening gap of 130 days. Similarly, the average patient who does not develop cancer will have a median screening gap of 296 days, an improved health outcome by means of avoiding unnecessary time, costs, and mental distress associated with screening visits. These same effects occur for all three policies.

We would advise a clinician looking to choose which of the three policies to adopt to choose based upon the findings in Figure 3.11. If patient equity is a priority, epsilon greedy strategies seem to achieve the most similar screening gaps between cancer and non-cancer patients. If identification and treatment of cancer patients is a priority, then Boltzmann exploration policies are the best choice, as they give the most frequent screenings to cancer patients. On the other hand, if the avoidance of costs and hassle of non-cancer patients with unnecessary screenings is a priority, we would advise the clinician to adopt interval estimation policies, as they demonstrate a large advantage in infrequently screening non-cancer patients.

3.8 Discussion

In this work, we searched a large, but by no means exhaustive, class of reinforcement learning algorithms to evaluate the benefits that can be gained by reallocating the existing screening resources. We believe that this approach of learning-based decision rules with a simulation built purely upon historical observations provides a highly accurate picture of the potential gains that can be made.

From this study, we induced several conclusions. The first is that current practice is roughly equal in performance to distributing resources randomly, thus creating the incentive to search for smarter behavior. Next, we saw that the optimal balance between explorative and exploitive behavior shifts towards the latter as resources become more scarce. We then estimated that switching from fixed-interval, equitable allocation of screening resources to a learning-based policy which utilizes sequentially gathered biological information will result in either 8.6% increased performance or 16.5% cost savings. We have also noted that these benefits of switching to a learning-based policy increase further as resources become more constrained. Lastly, we discussed the disadvantages of not utilizing learning based policies by showing myopic behavior to be vastly suboptimal.

We opted for the method of a historical simulation because of its potential for increased acceptance by the clinical community. This is due to two features of this method: we utilized a widely recognized clinical trials in the area of hepatocellular carcinoma, thereby creating results directly comparable to its well understood outcomes. Moreover, we avoided the usage of parametric assumptions on patient progression which are not recognized by the clinical community. These two features both establish strong rationale for a clinician to believe our simulation accurately replicated their situation.

Nevertheless, we would like to note that our method of historical simulation is not without its shortcomings. The simulation may suffer from censoring which is inherent to the data from which we drew patient progressions. The simulation may also fail to accurately reflect a typical cohort of American patients with Hepatitis C and advanced fibrosis. These concerns are mitigated by the particularly robust nature of the HALT-C dataset. With patients remaining under surveillance for an average of 5.3 years, the impact of censoring is far less than that associated with a typical observational study. And with over 1,000 patients enrolled at 12 different hospitals from all regions of the United States, the HALT-C trial can be viably accepted as an accurate depiction of the American population living with chronic Hepatitis C and advanced fibrosis.

In our model, recall that a positive detection is simulated only if the tumor is determined to be both greater than 1 cm in diameter, as well as being beyond the date of detection in the original data. These detection rules were chosen to increase acceptance of our analysis by the medical community. However, to evaluate the robustness of our results, a separate analysis re-sampled the best performing policies (which had been found based upon the original tumor detection assumptions) 5000 times each under an alternate set of assumptions where positive detections depended solely upon the size of the tumor.

We found that although the performances of the policies were 15% higher (averaged across all policies and scenarios) under these new detection rules, our major findings continued to hold. That is, current practice could either gain 8.6% increase in early stage detections, or a 16.4% cost-savings, by switching to a reinforcement learning based policy for patients at risk for HCC. While this provides some evidence of the robustness of our results, it may be worthwhile in future work to derive the optimal reinforcement learning policies according to this alternate detection rule.

It is interesting to note that Boltzmann exploration outperforms epsilon-greedy strategies and interval estimation, which is certainly not necessarily the case in other applications of reinforcement learning. We give two possible explanations for this result. Firstly, interval estimation performs exploration by encouraging screening of patients with high standard deviation of observed AFP readings. This is based upon the underlying assumption that more screenings results in a smaller standard deviation. This in turn is based upon the assumption that AFP readings are drawn from a normal distribution with known, fixed parameters, therefore eventually causing the sample variance to converge. For this reason, we believe interval estimation would be more successful in an application where patients biomarkers were distributed from normal distributions, and the problem of identifying their risk was equivalent to identifying the different means of those normal distributions.

Secondly, the weakness of an Epsilon-Greedy policy is forced exploration in long term scenarios. Consider a screening clinic which has been running for a very long time, relative to a disease which progresses very slowly. This clinic will have accumulated a large amount of information about all of its patients, and know each of its patients level of risk high certainty. The epsilon-greedy policy would still insist on allocating resources towards exploration in this scenario. Boltzmann exploration, on the other hand, adjusts its level of exploration according the strength of current knowledge. Therefore, we postulate that epsilon-greedy policies would be more successful if (1) we ran our case study over shorter time horizons, and/or (2) had a slower progressing underlying disease.

Further work may use stochastic simulation, where patient characteristics and their disease progressions are drawn parametrically. As an example, additive noise terms could be added to the AFP reading module to more realistically simulate the unpredictability of the AFP levels. Should the necessary parametrizations become established and available in the medical literature, this approach could potentially validate the results seen here, as well as provide further insight into the nature of efficient allocation of screenings.

These analyses could also be re-done with alternative objective functions to reflect other concerns of the screening clinic. Although we maximized the number of early-stage detections, it may be worthwhile consider rewards that are a function of the tumor size. Current staging definitions for HCC tumors use a threshold of 5 cm to distinguish between early stage and late stage tumors. While this may be a convenient definition for clinical classification, a patient's probability of survival may be better correlated with the tumor size at the time of his/her detection. This type of alternative objective function may identify a better policy more directly related to patient survival.

It might also be worthwhile to re-establish a measure of risk which turns all static risk factors (such as baseline age, smoking history, and baseline blood platelet count) into dynamic risk factors (current age, total years as a smoker, and current baseline blood platelet count). If this new measure of risk were to be established, we suspect it would only strengthen the decision making performed here. Our analysis could also benefit further from including the visual assessment of ultrasound images by doctors as an additional risk factor. Often a negative ultrasound will not contain enough features to warrant a diagnosis, yet it will still provide the doctor with some insight into the health of the liver organ. The main disadvantage of this usage of visual ultrasound images is that it is highly subjective between doctors, and it is difficult to quantify for a mathematical decision making framework.

Finally, other avenues for future work include complications that occur during real-life screening, such as false negative outcomes in the imaging process, panel size variability, penalties associated with screenings, and imperfect patient adherence.

We conclude with clinical recommendations derived from this work. Outside of direct policy adoption, a clinic can still utilize Equation 3.1 in isolation to gauge their patient's estimated current level of risk. Furthermore, our results can be used for capacity planning purposes to gain a better understanding of the potential marginal benefits of increasing their current screening resources. Lastly, we would advise doctors to recognize the importance of balancing the exploration and exploitation of information when allocating resources.

CHAPTER IV

Restless Bandit Based Policies

In this chapter, we seek an efficient way to screen a population of patients at-risk for hepatocellular carcinoma when (1) each patient's disease evolves stochastically, and (2) there are limited screening resources shared by the population. We model the problem as a family of restless bandits, with each patient's disease progression evolving as a partially observable markov decision process. We derive an optimal policy for this problem and discuss managerial insights into what characterizes more effective screening. To provide numerical evidence, we use two independent datasets of over 800 patients each, one to train the optimal policy, and the other to build a computer simulation to act as a testbed for said policy. We are able to show that our policy detects 22% more early stage cancers than current practice, while using the same amount of resource expenditure. We provide insights into the structure underlying our policy, and discuss the implications of our findings.

4.1 Introduction

The cost of modern American healthcare is projected to continue rising without impedance (*Keehan et al.* (2015)), and more policymakers are seeking ways to alleviate this exaggerated overspending on healthcare. Simultaneously, the aging population continues to impose increasing burdens upon existing healthcare infrastructure

(*Strunk et al.* (2006)), introducing new challenges to operate under limited capacities.

In the vast history of the mathematical optimization of medical decision making, the overwhelming majority of work has sought to optimize the health outcomes of a single patient. In light of the new realities of American healthcare, attempting to simultaneously execute policies which are optimal for every single patient will be both inordinately expensive, and possibly infeasible due to operational constraints. In this Chapter, we consider the novel setting of making medical decisions for multiple patients whose disease evolves simultaneously while sharing limited resources. As a proof of concept, we will look at liver cancer screening, a problem which is characterized by simultaneously evolving disease and a limited number of shared resources.

4.1.1 Hepatocellular Carcinoma Screening

Hepatocellular carcinoma (HCC) is the third leading cause of cancer-related deaths worldwide (*Altekruse et al.* (2009)). With over six million Americans at risk for HCC (*Wilkins et al.* (2010)), the size of the population at risk for HCC far outweighs the available infrastructure needed to properly screen them. Screening for HCC is critical because over 50% of patients whose cancer is diagnosed in early-stage are expected to be disease free after 5 years. By contrast, less than 10% of patients whose cancer is diagnosed in late-stage are expected to still be alive after 5 years (*Curley et al.* (2015)).

When screened for HCC, each patient undergoes multiple procedures. Firstly, an ultrasound of the liver is taken. If any new features of tumor growth are indicted in this ultrasound, the doctor will order a follow-up image via CT scan or MRI to confirm or deny these features. These tests are costly and limited in a hospital; ultrasounds are typically a shared resource between all departments of a major hospital, and in practice, are booked at over 100% capacity. CT scans and MRIs are expensive tests which should be used as sparingly as possible.

In addition, the patient's blood is measured for the alphafetoprotein (AFP), a biomarker for HCC. Recent medical literature has shown a relationship between a patient's AFP and his/her lifetime risk for HCC (*Lee et al.* (2012a), *Wong et al.* (2014), *Chaiteerakij et al.* (2013)). AFP readings are inherently noisy, and must be observed over time to properly assess a patient's risk of HCC.

The current recommended screening protocol in the United States proposes a fixed interval of six months for all patients at risk for HCC (*Bruix and Sherman* (2005)), regardless of any previous knowledge of the patient's risk. In contrast, we develop a model to determine which subset of patients to screen, based upon their observed risk, while ensuring that we only use as many resources as current practice. The current "one-size fits all" strategy will set the baseline performance against which our policy can be compared.

We model the problem of screening a population for HCC under limited resources as a restless bandit problem. Each patient's disease progression is modeled as a partially observable Markov process, and this accounts for how the decision maker will learn each individual patient's risk over time. We derive an optimal policy for this problem when the objective of the policy is to maximize early stage cancer detections over a finite planning horizon, and then show how the structural and computational complexity of this policy can be reduced.

We then provide a case study of how our policy would have performed in real life. The model is calibrated using data from a clinical trial, and the corresponding optimal policy is then tested against a simulation built upon historical patient data collected from a large university hospital. Through a robust set of scenarios, we show that our policy outperforms current practice.

The contributions of this work are three-fold: (1) Our work considers the optimization of patient-centered health outcomes, while constraining for resource expenditure at the population level. This is in contrast to most literature, which addresses the problem of how to optimally screen a single patient, with no regards to total system costs or resource expenditure. (2) We add new structural insights to a new class of restless bandit problems. In contrast, the analysis of restless bandits has traditionally relied upon approximative methods due to their complexity. (3) We interpret the policies derived to extract managerial insights for clinicians to screen more effectively; we provide strong numerical evidence (based on clinical trial and hospital data) that there is an opportunity for improvement in current screening practices that requires no additional cost.

The remainder of this Chapter is organized as follows. In §4.2 we survey the literature relevant to this problem. The problem is modeled and an optimal policy is derived in §4.3, followed by a reduction of that policy's structural and computational complexity. In §4.4, we present the simulation used to test our policy, and develop a heuristic for the implementation of our policy. We conclude with a discussion of the challenges of our findings, along with avenues for future work in §4.5.

4.2 Related Literature

Recently, there has been a growing concern for resource scarcity in medical decision making problems. For example, *Khademi et al.* (2015) considered policies to treat a population when drug supplies are severely limited. *Deo and Sohoni* (2015) studied the distribution of scarce diagnostic devices across a large population. In a similar vein, we consider how to allocate limited resources in liver cancer screening. We will survey the existing literature concerning (1) cancer screening, and (2) multi-armed bandits, a framework well-suited to handle resource allocation.

4.2.1 Optimization of Cancer Screening

The literature concerning the mathematical optimization of screening is vast, and we refer the reader to four surveys which together make a comprehensive overview of this problem's landscape: *Stevenson* (1995), *Alagoz et al.* (2011), *Pierskalla and Voelker* (1976) , *Pierskalla and Brailer* (1994).

Each type of cancer introduces new challenges in policymaking. In breast cancer, we must account for the possibility of overscreening, as well as regression (*Chhatwal et al.* (2010), *Maillart et al.* (2008), *Ayer et al.* (2012)). Colorectal cancer has particularly complex disease progression, with patients progressing in both risk and tumor stage (*Erenay et al.* (2014)). In some cancers, such as prostate cancer, we have additional biological information in the form of biomarkers to aid in screening decisions (*Zhang et al.* (2012), *Underwood et al.* (2012)). Despite the multitude of approaches to this problem, they all hold the common characteristic of seeking the optimal method to screen a single patient. While optimal treatment of a single patient to maximize his/her health outcomes is an important problem, this paper addresses the shortcomings of using a single patient perspective by developing a screening policy executed at a population wide level. We choose the methodology of multi-armed bandits for its ability to optimize rewards accrued over simultaneous stochastic processes, the key challenge in our problem.

4.2.2 Multi-armed Bandits

In a multi-armed bandit problem, a decision maker chooses one of several bandit arms for an immediate reward. The reward depends on the arm's current state, which evolves stochastically each time it is activated. The goal is to maximize the total discounted reward. The primary results of multi-armed bandit problems were established in the seminal work of *Gittins* (1979).

Our problem is of the restless bandit variation explored by *Whittle* (1988), where each bandit arm evolves stochastically regardless of whether or not it is activated. *Papadimitriou and Tsitsiklis* (1999) showed the restless bandit problem is P-SPACE hard, and thus an optimal policy can usually only be approximated. *Krishnamurthy* *and Evans* (2001) gave a more explicit solution to the restless bandit problem when the bandit arms are partially observable Markov decision process (POMDP), but their results are only valid for a small class of transition probability matrices that does not include our problem. In our paper, we take advantage of the structural characteristics of cancer screening to derive an explicit solution to this class of problems.

Ahuja and Birge (2016) showed how clinical trials can be viewed as multi-armed bandit problems, and how changing treatment decisions during the course of the trial can achieve better health outcomes for its participants. They demonstrated how a bandit problem can successfully be applied to population level medical decisions. However, they did not incorporate disease progression over time. *Negoescu et al.* (2014) modeled the treatment decisions of patients with Multiple Sclerosis as a continuous-time, multi-armed bandit problem. Their clinical outcomes of interest were singular events (i.e. relapses, flare-ups). In contrast, our outcomes of interest, early-stage cancer, is a state with time duration, which requires a significantly different approach.

Closest to our work is *Deo et al.* (2013), who modeled the treatment of chronic diseases at a community level using a restless bandit model. Their paper also takes a population-level approach to a medical decision. However, we assume our population is heterogenous in their underlying disease progression, whereas their population is stochastically homogenous. Also, our model specifically addresses liver cancer, the structure of which ultimately enables us to gain deeper insights into the structure of the optimal policy for this specific problem.

4.3 The Modeling Framework

We model this problem as a discrete-time, finite-horizon simple family of multi-armed restless bandits where each bandit is a partially observable Markov decision process. The sequence of events within a single decision epoch are depicted in Figure

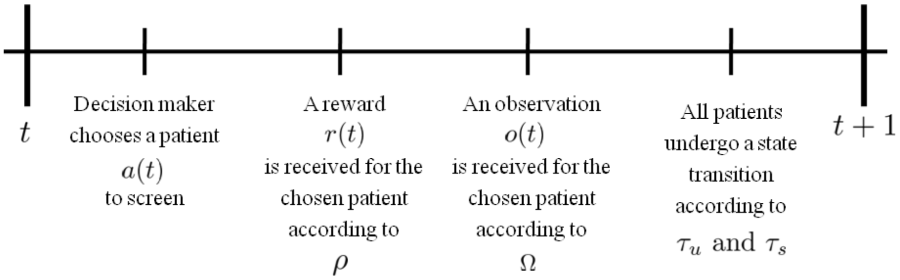

t
Decision maker chooses a patient $a(t)$ to screen
A reward $r(t)$ is received for the chosen patient according to ρ
An observation $o(t)$ is received for the chosen patient according to Ω
All patients undergo a state transition according to τ_u and τ_s
$t+1$

Figure 4.1: The sequence of events within a single decision epoch.

4.1.

A single decision maker seeks to maximize the number of early cancers detected over a finite horizon, with no terminal reward. The assumption of a finite horizon is chosen for its ease of interpretation amongst clinicians and policymakers. At the beginning of each decision epoch t, the decision maker must choose a single patient $a(t)$ to undergo screening. In the case study, we relax the assumption of screening only a single patient in each period. The decision maker will then receive a reward $r(t)$, and an observation $o(t)$, both of which depend on the current state of the patient chosen to be screened. After this decision, all patients (both screened and unscreened) will transition according to the underlying Markov chain, in the same vein as much of the literature on cancer screening optimization.

If a patient is in early-stage or late-stage cancer, we assume their state is perfectly observed, as in practice, diagnosis is made by ultrasound followed by CT or MRI. This combination has been reported to have sensitivity of 97% (*Arif-Tiwari et al.* (2014)) and specificity of 96% (*Nam et al.* (2011)). Future work may relax this assumption. However, if a patient does not have cancer, his/her risk is imperfectly observed through a noisy AFP reading. We assume perfect patient adherence to scheduled screenings, and that all decisions and events happen instantaneously within a decision epoch, two assumptions which may be relaxed in future work. The components of our model are as follows:

- t: Decision epochs, $t = 1, ..., T$; $T < \infty$.

- i: Patients, $i = 1, ..., N$. In our model, we assume the screening clinic will maintain a panel size of N patients. If a patient is diagnosed with cancer, or dies from non-cancer causes, that patient leaves the clinic and a new patient will be found to occupy the available capacity.

- j: Health state space, where $j = 1, ..., m, E, L$. The first $1, ..., m$ states are the possible cancer-free risk types. State E represents early-state cancer, and L represents a combination of health states in which a patient would leave the system, whether through late-stage cancer, cancer-caused mortality, or non-cancer caused mortality. These three outcomes are indistinguishable from the perspective of the decision maker, given the objective of maximizing early-stage cancers detected, and hence can be modeled as a single state.

- Y_{it}: The true health state of patient i at time t, $Y_{it} \in \{1, ..., .m, E, L\}$, which is unknown to the decision maker and can only be estimated.

- τ_u, τ_s: Transition probability matrices, for unscreened and screened patients, respectively. These transitions are depicted along with their respective matrix representations in Figures 4.2 and 4.3. Patients in cancer-free states $j = 1, ..., m$ remain cancer-free with probability P_{jj}, transition to early-stage cancer with probability P_{jE}, or die from non-cancer related mortality with probability δ. Patients in early-stage cancer remain in early-stage cancer with probability P_{EE}, or transition to late-stage cancer with probability P_{EL}. Patients in late-stage cancer/death (L) remain in that state with probability 1. The key difference between τ_u and τ_s is that when a patient is screened and found to be in early-stage or late-stage cancer, they leave the system and a new cancer-free patient is assumed to arrive to the clinic. This replacement is accomplished by letting the Markov chain transition to one of the $j = 1, ..., m$ cancer-free states, each

$$\mathcal{T}_u = \begin{array}{c|ccc|cc} & 1 & \cdots & m & E & L \\ \hline 1 & P_{11} & & & P_{1E} & \delta \\ \vdots & & \ddots & & \vdots & \vdots \\ m & & & P_{mm} & P_{mE} & \delta \\ \hline E & & & & P_{EE} & P_{EL} \\ L & & & & & 1 \end{array}$$

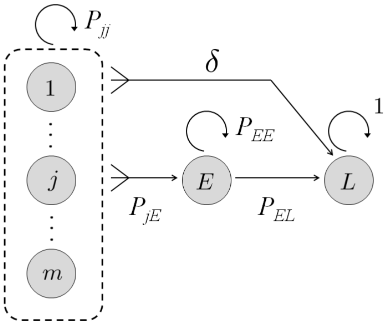


Figure 4.2: Transition matrix and diagram for an unscreened patient.

$$\mathcal{T}_s = \begin{array}{c|ccc|cc} & 1 & \cdots & m & E & L \\ \hline 1 & P_{11} & & & P_{1E} & \delta \\ \vdots & & \ddots & & \vdots & \vdots \\ m & & & P_{mm} & P_{mE} & \delta \\ \hline E & q_1 & \cdots & q_m & & \\ L & q_1 & \cdots & q_m & & \end{array}$$

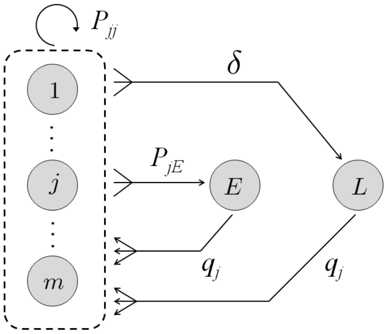


Figure 4.3: Transition matrix and diagram for a screened patient.

with probability q_j. While we acknowledge this to be simplification of the true system, it enables us to maintain a constant panel size which has been roughly observed in our partnering clinic. Further extensions of this model may incorporate changing panel sizes.

- Ω: Observation matrix, shown in Figure 4.4. When a cancer-free patient of type $j = 1, ..., m$ is screened, a discretized AFP observation is received of type $k = 1, ..., p$ with probability o_{jk}. $\bar{E}$ and $\bar{L}$ are used to denote perfect observation of cancer states. We will also frequently use the notation Ω_k, Ω_E, and Ω_L. These are diagonal matrices of dimension $(p + 2) \times (p + 2)$, whose entries consist of the rows k, E, and L of the matrix Ω.

- ρ: Reward vector. Because the goal of the decision maker is to maximize the

$$\Omega = \begin{array}{c} \\ 1 \\ \vdots \\ m \\ E \\ L \end{array} \begin{array}{ccc:cc} 1 & \cdots & p & \bar{E} & \bar{L} \\ \ddots & & \cdot^{\cdot^{\cdot}} & & \\ & o_{jk} & & & \\ \cdot^{\cdot^{\cdot}} & & \ddots & & \\ \hdashline & & & 1 & \\ & & & & 1 \end{array}$$

Figure 4.4: Observation probability matrix.

number of screenings in early-stage cancer, there is a reward of 1 for a patient screened in state E and 0 for all other states. Therefore rewards are dictated by the vector $\rho = \begin{array}{c} 1 \quad ... \quad m \quad E \quad L \\ \begin{pmatrix} 0 & ... & 0 & 1 & 0 \end{pmatrix} \end{array}$. A table summarizing the notation of this paper is provided in Table 4.1.

Indices	
$i = 1, ..n$	Patients
$j = 1, ..., m$	Risk types
$k = 1, ..., p$	AFP observations
$t = 1, ..., T$	Time
Primitive Data	
P_{jj}, P_{jE}	Transition probability from state j to state j, or E, etc.
τ_u, τ_s	Transition matrices, for unscreened and screened patients, respectively
O_{jk}	Observation probability that a patient of type j gives observation k
Ω	Observation matrix
$\Omega_k, \Omega_E, \Omega_L$	The kth row of Ω diagonalized into a matrix
ρ	Reward vector
Model Notation	
X_{ijt}	The belief that patient i is in state j at the beginning of time t
Y_{it}	The true health state of patient i at the beginning of time t
$a(t)$	The patient chosen to be screened at time t
$o(t)$	The observation received at time t
$r(t)$	The reward received at time t

Table 4.1: Table of notation

4.3.1 Belief States

We will use X_{ijt} to represent the belief that patient i is in state j at time t. More explicitly,

$$X_{ijt} := Pr(Y_{it} = j) \quad \forall i, j, t \tag{4.1}$$

X_t will represent the collection of all beliefs about all n patients at time t. X_{it} will represent all beliefs about a particular patient i at time t. The relationship between these forms of the belief state are demonstrated in Equation (4.2).

$$X_t = \begin{bmatrix} \text{---} & X_{1t} & \text{---} \\ & \vdots & \\ \text{---} & X_{it} & \text{---} \\ & \vdots & \\ \text{---} & X_{nt} & \text{---} \end{bmatrix} = \begin{bmatrix} X_{11t}, & \dots & X_{1jt}, & \dots & X_{1mt}, & X_{1Et}, & X_{1Lt} \\ & & \vdots & & & & \\ X_{i1t}, & \dots & X_{ijt}, & \dots & X_{imt}, & X_{iEt}, & X_{iLt} \\ & & \vdots & & & & \\ X_{n1t}, & \dots & X_{njt}, & \dots & X_{nmt}, & X_{nEt}, & X_{nLt} \end{bmatrix} \tag{4.2}$$

Given the order of events depicted in Figure 4.1, we can now compute how the beliefs X_{ijt} evolve from one period to the next. $X_{ij,t+1}$ will depend on two things: (1) whether or not patient i was the patient $a(t)$ chosen to be screened, and (2) the observation $o(t)$ received upon screening. Standard applications of Bayes' rule can be applied to give the formulas for the subsequent belief state $X_{i,t+1}$ in Equation (4.3).

$$X_{i,t+1} = \begin{cases} \overline{X_{it}\Omega_k}\tau_s & \text{if } a(t) = i \text{ and } o(t) = 1, ..., p \\ \overline{X_{it}\Omega_E}\tau_s & \text{if } a(t) = i \text{ and } o(t) = \bar{E} \\ \overline{X_{it}\Omega_L}\tau_s & \text{if } a(t) = i \text{ and } o(t) = \bar{L} \\ X_{it}\tau_u & \text{if } a(t) \neq i \end{cases} \tag{4.3}$$

Where we have used the convention $\overline{v}$ to represent the vector v, normalized to condense notation.

To derive $X_{i,t+1}$, given any previous belief state X_{it}, any action $a(t)$ and any screening outcome $o(t)$, we need $X_{ij,t+1} = Pr(Y_{i,t+1} = j|a(t), o(t), X_{it})$, $\forall j$. We will divide this section into three cases, depending upon the action $a(t)$ and the stochastic observation $o(t)$. Because in our model, observations occur before transitions in each decision epoch, we denote t' to denote the belief immediately after an observation is received, but before transition occurs.

Case 1. $a(t) = i$ **and** $k = 1, ..., p$ **(Screened patient and non-cancer reading)**

We wish to calculate

$$X_{ij,t+1} = Pr(Y_{i,t+1} = j|a(t) = i, o(t) = k, X_{it}) \tag{4.4}$$

Expanding by conditioning on the belief at the intermediate time t', we get:

$$X_{ij,t+1} = \sum_{j'} Pr(Y_{i,t+1} = j|Y_{it'} = j', a(t) = i, o(t) = k, X_{it}) \\ \cdot Pr(Y_{it'} = j'|a(t) = i, o(t) = k, X_{it}) \tag{4.5}$$

Recognizing that the transition from t' to $t + 1$ is independent of the action and observations, we get:

$$X_{ij,t+1} = \sum_{j'} Pr(Y_{i,t+1} = j|Y_{it'} = j') \cdot Pr(Y_{it'} = j'|a(t) = i, o(t) = k, X_{it}) \tag{4.6}$$

Substituting for primitive data where applicable

$$X_{ij,t+1} = \sum_{j'} P_{j'j} \cdot Pr(Y_{it'} = j'|a(t) = i, o(t) = k, X_{it}) \tag{4.7}$$

The remainder of this equation will be different, depending on 1 of 3 cases:

Case 1a. $j = 1, ..., m$ **(Probability of being cancer-free)**

Equation (4.7) can be simplified because $P_{j'j} = 0$ if $j \neq j'$

$$X_{ij,t+1} = P_{jj} \cdot Pr(Y_{it'} = j | a(t) = i, o(t) = k, X_{it})$$

Applying Bayes' Law, we get:

$$X_{ij,t+1} = P_{jj} \cdot \frac{Pr(o(t) = k | Y_{it'} = j, a(t) = i, X_{it}) Pr(Y_{it'} = j | a(t) = i, X_{it})}{\sum_{\hat{j}} Pr(o(t) = k | Y_{it'} = \hat{j}, a(t) = i, X_{it}) Pr(Y_{it'} = \hat{j} | a(t) = i, X_{it})}$$

Substituting for any known quantities:

$$X_{ij,t+1} = P_{jj} \cdot \frac{O_{jk} X_{ijt}}{\sum_{\hat{j}} O_{\hat{j}k} X_{i\hat{j}t}} \tag{4.8}$$

Case 1b. $j = E$ **(Probability of being in early-stage cancer)**

Containing from Equation (4.7):

$$X_{iE,t+1} = \sum_{j'} P_{j'E} \cdot Pr(Y_{it'} = j' | a(t) = i, o(t) = k, X_{it}) \tag{4.9}$$

The summands $j' = E, L$ can be removed from 4.9 because the state at t' cannot be E or L if an observation k was received

$$X_{iE,t+1} = \sum_{j'}^{m} P_{j'E} \cdot Pr(Y_{it'} = j' | a(t) = i, o(t) = k, X_{it}) \tag{4.10}$$

Apply Bayes' Law to get:

$$X_{iE,t+1} = \sum_{j'}^{m} P_{j'E} \frac{Pr(o(t) = k | X_{ijt'} = j', a(t) = i, X_{it}) Pr(Y_{it'} = j' | a(t) = i, X_{it})}{\sum_{\hat{j}} Pr(o(t) = k | X_{ijt'} = \hat{j}, a(t) = i, X_{it}) Pr(Y_{it'} = \hat{j} | a(t) = i, X_{it})} \tag{4.11}$$

Substituting for any known quantities:

$$X_{iE,t+1} = \frac{\sum_{j'}^{m} P_{j'E} O_{j'k} X_{ijt}}{\sum_{\hat{j}} O_{\hat{j}k} X_{i\hat{j}t}} \tag{4.12}$$

Case 1c. $j = L$ **(Probability of being in late-stage cancer)**

Continuing from Equation (4.7):

$$X_{iL,t+1} = \sum_{j'} P_{j'L} \cdot Pr(Y_{it'} = j' | a(t) = i, o(t) = k, X_{it}) \tag{4.13}$$

The summands $j' = E, L$ can be removed from 4.13 because the state at t' cannot be E or L if an observation k was received

$$X_{iL,t+1} = \sum_{j'}^{m} P_{j'L} \cdot Pr(Y_{it'} = j' | a(t) = i, o(t) = k, X_{it}) \tag{4.14}$$

Applying Bayes' Law, we get:

$$X_{iL,t+1} = \sum_{j'}^{m} P_{j'L} \frac{Pr(o(t) = k | X_{ijt'} = j', a(t) = i, X_{it}) Pr(Y_{it'} = j' | a(t) = i, X_{it})}{\sum_{\hat{j}} Pr(o(t) = k | X_{ijt'} = \hat{j}, a(t) = i, X_{it}) Pr(Y_{it'} = \hat{j} | a(t) = i, X_{it})} \tag{4.15}$$

Substituting for any known quantities:

$$X_{iL,t+1} = \frac{\sum_{j'}^{m} \delta O_{j'k} X_{ijt}}{\sum_{\hat{j}} O_{\hat{j}k} X_{i\hat{j}t}} \tag{4.16}$$

Therefore, Equations (4.8), (4.12), and (4.16) can be summarized in vector form as follows. When $a(t) = i$ observation $o(t) = 1, 2, ..., p$:

$$(X_{i,t+1}|a(t) = i, o(t) = k) = \begin{bmatrix} \left. \begin{matrix} X_{i1t+1} \\ \vdots \\ X_{ijt+1} \\ \vdots \\ X_{imt+1} \\ X_{iEt+1} \\ X_{iLt+1} \end{matrix} \right| a(t) = i, o(t) = k \end{bmatrix}^T \tag{4.17}$$

$$= \begin{bmatrix} \frac{P_{11}O_{1k}X_{i1t}}{\sum_{j=1}^{m} O_{jk}X_{ijt}} \\ \vdots \\ \frac{P_{jj}O_{jk}X_{ijt}}{\sum_{j=1}^{m} O_{jk}X_{ijt}} \\ \vdots \\ \frac{P_{mm}O_{mk}X_{imt}}{\sum_{j=1}^{m} O_{jk}X_{ijt}} \\ \frac{\sum_{j=1}^{m} P_{jE}O_{jk}X_{ijt}}{\sum_{j=1}^{m} O_{jk}X_{ijt}} \\ \frac{\sum_{j=1}^{m} \delta O_{jk}X_{ijt}}{\sum_{j=1}^{m} O_{jk}X_{ijt}} \end{bmatrix}^T = \overline{X_{it}\Omega_k}\tau_S \forall k = 1, ..., p$$

Case 2. $a(t) = i$ **and** $k = \bar{E}, \bar{L}$ **(Screened patient and cancer reading)**

The belief state update when $a(t) = i$ and the observation $o(t) = \bar{E}$ or $o(t) = \bar{L}$ is far simpler:

$$X_{ij,t+1} = Pr(Y_{i,t+1} = j|a(t) = i, o(t) = \bar{E}, X_{it}) \tag{4.18}$$

Expanding by conditioning on the intermediate state at time t' we get:

$$X_{ij,t+1} = \sum_{j'} Pr(Y_{i,t+1} = j|Y_{it'} = j', a(t) = i, o(t) = \bar{E}, X_{it}) \\ \cdot Pr(Y_{it'} = j'|a(t) = i, o(t) = \bar{E}, X_{it}) \tag{4.19}$$

But when we see an observation $\bar{E}$, we know that the state at time t' is E with probability 1, so all but one of the summands are zero

$$X_{ij,t+1} = Pr(X_{i,t+1} = j | Y_{it'} = \bar{E}, a(t) = i, o(t) = \bar{E}, X_{it}) \tag{4.20}$$

Lastly, we know from τ_S that a patient who is found to be in state E will replaced by new patient with probabilities q_j

$$X_{ij,t+1} = q_{j^*} \tag{4.21}$$

Equation (4.21) can be summarized in matrix form in Equation (4.22):

$$(X_{i,t+1} | a(t) = i, o(t) = k) = \left[\begin{array}{c} X_{i1t+1} \\ \vdots \\ X_{ijt+1} \\ \vdots \\ X_{imt+1} \\ X_{iEt+1} \\ X_{iLt+1} \end{array} \middle| a(t) = i, o(t) = k \right]^T \tag{4.22}$$

$$= \begin{bmatrix} q_1 \\ \vdots \\ q_j \\ \vdots \\ q_m \\ 0 \\ 0 \end{bmatrix}^T = \begin{cases} \overline{X_{it}\Omega_E}\tau_S & \text{if } k = \bar{E} \\ \overline{X_{it}\Omega_L}\tau_S & \text{if } k = \bar{L} \end{cases} \tag{4.23}$$

Case 3. $a(t) \neq i$ **(Non-screened patient)**

On the other hand, the subsequent belief state of the unscreened patient, $a(t) \neq i$ evolves regardless of the observation $o(t)$, therefore:

$$(X_{i,t+1}|a(t) \neq i, o(t) = k) = \left[\begin{array}{c|c} X_{i1t+1} & \\ \vdots & \\ X_{ijt+1} & \\ \vdots & a(t) \neq i, o(t) = k \\ X_{imt+1} & \\ X_{iEt+1} & \\ X_{iLt+1} & \end{array}\right]^T \tag{4.24}$$

$$= \begin{bmatrix} P_{11}X_{1t} \\ \vdots \\ P_{jj}X_{ijt} \\ \vdots \\ P_{mm}X_{mt} \\ \sum_{j=1}^{m} P_{jE}X_{jt} + P_{EE}X_{Et} \\ \sum_{j=1}^{m} \delta X_{jt} + P_{EL}X_{Et} + 1X_{Lt} \end{bmatrix}^T = X_{it}\tau_U \tag{4.25}$$

We can now combine Equations (4.17), (4.22) and (4.24) into Equation (4.26) to tell us the subsequent belief state $X_{ij,t+1}$, given any previous belief state X_{it}, any action $a(t)$ and any screening outcome $o(t)$

$$X_{i,t+1} = \begin{cases} \overline{X_{it}\Omega_k}\tau_s & \text{if } a(t) = i \text{ and } o(t) = 1, ..., p \\ \overline{X_{it}\Omega_E}\tau_s & \text{if } a(t) = i \text{ and } o(t) = \bar{E} \\ \overline{X_{it}\Omega_L}\tau_s & \text{if } a(t) = i \text{ and } o(t) = \bar{L} \\ X_{it}\tau_u & \text{if } a(t) \neq i \end{cases} \tag{4.26}$$

The probability of any observation in terms of the belief state X_{ijt} is given in Equation (4.30), can be found as follows:

We wish to know the probability of any observation, given the current belief state X_t and a patient choice $a(t) = i$:

$$Pr(o(t) = k|a(t) = i) \tag{4.27}$$

Conditioning on the underlying state of the patient i, we get:

$$Pr(o(t) = k|a(t) = i) = \sum_{j} Pr(o(t) = k|a(t) = i, X_{it} = j) Pr(X_{it} = j|a(t) = i) \tag{4.28}$$

Substituting for any known quantities:

$$Pr(o(t) = k|a(t) = i) = \begin{cases} \sum_{j=1}^{m} o_{jk} X_{ijt} & \text{for } k = 1, ..., p \\ X_{iEt} & \text{for } k = \bar{E} \\ X_{iLt} & \text{for } k = \bar{L} \end{cases} \tag{4.29}$$

$$Pr(o(t) = k \mid a(t) = i, X_t) = \begin{cases} X_{a(t)} \Omega_k \vec{1} & \text{for } k = 1, ..., p \\ X_{a(t)} \Omega_E \vec{1} & \text{for } k = \bar{E} \\ X_{a(t)} \Omega_L \vec{1} & \text{for } k = \bar{L} \end{cases} \tag{4.30}$$

Also, the reward received $r(t)$ is dictated by ρ, and depends upon the observation alone:

$$r(t) = \begin{cases} 0, & \text{if } o(t) = 1, ..., p \\ 1, & \text{if } o(t) = \bar{E} \\ 0, & \text{if } o(t) = \bar{L} \end{cases} \tag{4.31}$$

Equation (4.26) tells us how belief states will evolve given any action and any observation. Equation (4.30) tells us the probability of these stochastic events given any action. Lastly, Equation (4.31) tells us how the system will accrue rewards given any stochastic event. These are all the elements needed to write the corresponding dynamic program for this model.

4.3.2 Optimality Equation

Let $V_t(X_t)$ be the maximum expected reward from periods t through T, given a belief state X_t at time t. We use Bellman's optimality principle to expand $V_t(X_t)$. The value of any state is the maximum of the immediate reward, given an action, plus the value of the next state resulting from that action. All time subscripts will be omitted from the belief state in this section whenever the distinction is unnecessary.

$$V_t(X_t) = \max_{a(t)=1,\ldots,n} \{(r_t|a(t)=i) + V_{t+1}(X_{t+1}|a(t)=i)\} \tag{4.32}$$

We now condition on the stochastic observation $o(t)$:

$$V_t(X_t) = \max_{a(t)=1,\ldots,n} \Bigg\{ \sum_{k=1,\ldots,p,\bar{E},\bar{L}} Pr(o(t)=k) \\ \Big[(r_t|a(t)=i, o(t)=k) + V_{t+1}\left(X_{t+1}|a(t)=i, o(t)=k\right)\Big]\Bigg\} \tag{4.33}$$

After expanding the summation over the three possible types of observations, the optimality equation can now be expressed in terms of the primitive data, made explicit in the Equations (4.26), (4.30), and (4.31). Substituting for each appropriate expression, we get:

$$V_t(X_t) = \max_{a(t)=1,...,n} \left\{ \begin{array}{l} \displaystyle\sum_{k=1}^{p} X_{a(t)}\Omega_k\vec{1} \cdot \left[0 + V_{t+1}\left(X_1\tau_u, ..., \overline{X_{a(t)}\Omega_k}\tau_s, ..., X_n\tau_u\right)\right] \\ +X_{a(t)}\Omega_E\vec{1} \cdot \left[1 + V_{t+1}\left(X_1\tau_u, ..., \overline{X_{a(t)}\Omega_E}\tau_s, ..., X_n\tau_u\right)\right] \\ +X_{a(t)}\Omega_E\vec{1} \cdot \left[0 + V_{t+1}\left(X_1\tau_u, ..., \overline{X_{a(t)}\Omega_L}\tau_s, ..., X_n\tau_u\right)\right] \end{array} \right\} \tag{4.34}$$

First, notice that $\overline{X\Omega_E}\tau_S = [q_1, ..., q_m, 0, 0] = \overline{X\Omega_L}\tau_S$, which can be used to simplify this expression into:

$$V_t(X_t) = \max_{a(t)=1,...,n} \left\{ \begin{array}{l} \displaystyle\sum_{k=1}^{p} X_{a(t)}\Omega_k\vec{1} \cdot \left[0 + V_{t+1}\left(X_1\tau_u, ..., \overline{X_{a(t)}\Omega_k}\tau_s, ..., X_n\tau_u\right)\right] \\ +X_{a(t)}\Omega_E\vec{1} \cdot \left[1 + V_{t+1}\left(X_1\tau_u, ..., \overline{X_{a(t)}\Omega_E}\tau_s, ..., X_n\tau_u\right)\right] \\ +X_{a(t)}\Omega_L\vec{1} \cdot \left[0 + V_{t+1}\left(X_1\tau_u, ..., \overline{X_{a(t)}\Omega_E}\tau_s, ..., X_n\tau_u\right)\right] \end{array} \right\} \tag{4.35}$$

If we distribute the terms outside of the hard brackets, we get:

$$V_t(X_t) = \max_{a(t)=1,...,n} \left\{ \begin{array}{c} \displaystyle\sum_{k=1}^{p} X_{a(t)}\Omega_k\vec{1} \cdot V_{t+1}\left(X_1\tau_u, ..., \overline{X_{a(t)}\Omega_k}\tau_s, ..., X_n\tau_u\right) \\ +X_{a(t)}\Omega_E\vec{1} \\ +(X_{a(t)}\Omega_E\vec{1} + X_{a(t)}\Omega_L\vec{1}) \cdot V_{t+1}\left(X_1\tau_u, ..., \overline{X_{a(t)}\Omega_E}\tau_s, ..., X_n\tau_u\right) \end{array} \right\} \tag{4.36}$$

We replace $X_{a(T)}\Omega_E\vec{1} = X_{a(t)}\rho$.

$$V_t(X_t) = \max_{a(t)=1,\ldots,n} \left\{ \begin{array}{c} \displaystyle\sum_{k=1}^{p} X_{a(t)}\Omega_k\vec{1} \cdot V_{t+1}\left(X_1\tau_u, ..., \overline{X_{a(t)}\Omega_k}\tau_s, ..., X_n\tau_u\right) \\ +X_{a(t)}\rho \\ +(X_{a(t)}\Omega_E\vec{1} + X_{a(t)}\Omega_L\vec{1}) \cdot V_{t+1}\left(X_1\tau_u, ..., \overline{X_{a(t)}\Omega_E}\tau_s, ..., X_n\tau_u\right) \end{array} \right\} \tag{4.37}$$

And finally notice that $X_{a(t)}\Omega_E\vec{1} = X_{a(t),E}$ and $X_{a(t)}\Omega_L\vec{1} = X_{a(t),L}$

$$V_t(X_t) = \max_{a(t)=1,\ldots,n} \left\{ \begin{array}{c} \displaystyle\sum_{k=1}^{p} X_{a(t)}\Omega_k\vec{1} \cdot V_{t+1}\left(X_1\tau_u, ..., \overline{X_{a(t)}\Omega_k}\tau_s, ..., X_n\tau_u\right) \\ +X_{a(t)}\rho \\ +(X_{a(t),E} + X_{a(t),L}) \cdot V_{t+1}\left(X_1\tau_u, ..., \overline{X_{a(t)}\Omega_E}\tau_s, ..., X_n\tau_u\right) \end{array} \right\} \tag{4.38}$$

Furthermore, our clinical collaborators felt the ending health states of any patients were of negligible consequence in comparison to the length of the planning horizon. Therefore, we impose no terminal reward for our problem:

$$V_{T+1}(X_{T+1}) = 0 \quad \forall X_{T+1} \tag{4.39}$$

4.3.3 An Optimal Screening Policy

The optimality Equation (4.38) is written in recursive form. We wish to derive a non-recursive form of the optimality equation for any time t, along with its corresponding optimal policy. We will proceed to do this through backwards induction, but first we require some helpful definitions.

Definition 1. The condition $C(r)$, for any $r > 0$, is said to be satisfied by the belief state X_t at time t if the following holds true:

For any $b \in \{1, ..., n\}$, there exists a corresponding $a \in \{1, ..., n\}, a \neq b$ such that

$$X_{at}\tau_u^r\rho \;\geq\; \overline{X_{bt}\Omega_k}\tau_s\tau_u^{r-1}\rho$$

For every possible $k \in \{1, ..., p, E, L\}$.

Intuitively, the condition $C(r)$ can be understood in the following way. The left hand side of the inequality represents the expected value of a patient a's belief state, after having gone unscreened for r consecutive periods. The right hand side of the inequality represents the expected value of patient b's belief state after being screened once and obtaining the observation k, then going unscreened for the next $r-1$ periods. Therefore, the condition $C(r)$ holds if for every possible patient screened and any observation outcome, there exists another corresponding patient with higher probability of being in early cancer stage cancer after r periods of being unscreened. The strength of this condition depends on the problem-specific values of Ω and τ_U, and will be investigated in the case-study portion of this paper. These conditions $C(r)$ are the sufficient conditions needed to write a non-recursive form of the optimality equation in Theorem IV.1.

Theorem IV.1. *At every time $t = 1, 2, ..., T$, let $r = T - t$. If $C(1)$, $C(2)$,..., $C(r)$ hold at time t, then the optimality equation is*

$$V_t(X_t) = \max_{a(t)=1,\ldots,n} \left\{ X_{a(t)}\rho + \max_{\substack{a(t+1)=1,\ldots,n \\ a(t+1)\neq a(t)}} \left\{ X_{a(t+1)}\tau_u\rho + \ldots \right.\right.$$
$$\left.\left. \ldots + \max_{\substack{a(T)=1,\ldots,n \\ a(T)\neq a(T-1),a(T-2),\ldots,a(t)}} \left\{ X_{a(T)}\tau_u^r\rho \right\} \right\} \right\} \qquad (4.40)$$

Furthermore, it is optimal to screen patient $a(t)$ at time t.

Theorem IV.1 tells us that the optimal patient to screen at any time t can be found by solving a new separate problem in Equation (4.40). Although Theorem IV.1 may be cumbersome, it is necessary to establish before a simpler policy is derived in Section 4.3.4.

Proof. By plugging the terminal condition in Equation (4.39) above into the recursive optimality Equation (4.38), we get

$$V_T(X_T) = \max_{a(T)=1,\ldots,n} \left\{ \begin{array}{c} \displaystyle\sum_{k=1}^{p} X_{a(T)}\Omega_k \vec{1} \cdot (0) \\ +X_{a(T)}\rho \\ +(X_{a(t),E} + X_{a(T),L}) \cdot (0) \end{array} \right\} \tag{4.41}$$

Which simplifies to:

$$V_T(X_T) = \max_{a(T)=1,\ldots,n} \left\{ X_{a(T)}\rho \right\} \tag{4.42}$$

This agrees with Theorem IV.1, and agrees with our intuition as well. The optimal action in the final period should be to act greedily, i.e. screen the patient with the highest belief of being in early-stage cancer. We now iterate this process of backwards recursively deriving optimality equations. If we apply Equation (4.38) for $t = T - 1$, we get:

$$= \max_{a(T-1)=1,\ldots,n} \left\{ \begin{array}{c} \displaystyle\sum_{k=1}^{p} X_{a(T-1)}\Omega_k \vec{1} \cdot V_T\left(X_1\tau_u, \ldots, \overline{X_{a(T-1)}\Omega_k}\tau_s, \ldots, X_n\tau_u\right) \\ +X_{a(T-1)}\rho \\ +(X_{a(T-1),E} + X_{a(T-1),L}) \cdot V_T\left(X_1\tau_u, \ldots, \overline{X_{a(T-1)}\Omega_E}\tau_s, \ldots, X_n\tau_u\right) \end{array} \right\} \tag{4.43}$$

Now use Equation (4.42) to substitute for V_T where applicable:

$$V_{T-1}(X_{T-1}) = \max_{a(T-1)=1,\dots,n} \left\{ \begin{array}{c} \sum_{k=1}^{p} X_{a(T-1)}\Omega_k \vec{1} \\ \cdot \max\limits_{a(T)=1,\dots,n} \left\{ X_1 \tau_u \rho, \dots, \overline{X_{a(T-1)}\Omega_k} \tau_s \rho, \dots, X_n \tau_u \rho \right\} \\ + X_{a_{(T-1)}}\rho \\ + (X_{a(T-1),E} + X_{a(T-1),L}) \\ \cdot \max\limits_{a(T)=1,\dots,n} \left\{ X_1 \tau_u \rho, \dots, \overline{X_{a(T-1)}\Omega_E} \tau_s \rho, \dots, X_n \tau_u \rho \right\} \end{array} \right\} \tag{4.44}$$

The inner nested maximizations can be re-written. Notice that it is a maximization of n summands, $n-1$ of which have the same form, and the $a(T-1)^{th}$ summand is different from the others.

$$V_{T-1}(X_{T-1}) = \max_{a(T-1)=1,\dots,n} \left\{ \begin{array}{c} \sum_{k=1}^{p} X_{a(T-1)}\Omega_k \vec{1} \\ \cdot \max\limits_{\substack{a(T)=1,\dots,n \\ a(T)\neq a(T-1)}} \left\{ X_{a(T)} \tau_u \rho, \ \overline{X_{a(T-1)}\Omega_k} \tau_s \rho \right\} \\ + X_{a(T-1)}\rho \\ + (X_{a(T-1),E} + X_{a(T-1),L}) \\ \cdot \max\limits_{\substack{a(T)=1,\dots,n \\ a(T)\neq a(T-1)}} \left\{ X_{a(T)} \tau_u \rho, \ \overline{X_{a(T-1)}\Omega_E} \tau_s \rho \right\} \end{array} \right\} \tag{4.45}$$

The reasons for the conditions $C(r)$ will now become apparent. The condition guarantees that the maximum of the inner nested maximizations will be achieved by one of the $n-1$ summands of similar form.

Lemma 1. *Suppose $C(1)$ holds at time $T-1$. Then it follows that*

$$\max_{\substack{a(T)=1,\dots,n \\ a(T)\neq a(T-1)}} \left\{ X_{a(T)} \tau_u \rho, \ \overline{X_{a(T-1)}\Omega_k} \tau_s \rho \right\} = \max_{\substack{a(T)=1,\dots,n \\ a(T)\neq a(T-1)}} \left\{ X_{a(T)} \tau_u \rho \right\}$$

and

$$\max_{\substack{a(T)=1,\ldots,n\\a(T)\neq a(T-1)}} \left\{X_{a(T)}\tau_u\rho,\ \overline{X_{a(T-1)}\Omega_E}\tau_s\rho\right\} = \max_{\substack{a(T)=1,\ldots,n\\a(T)\neq a(T-1)}} \left\{X_{a(T)}\tau_u\rho\right\}$$

Proof. Both of these statements apply as a direct result of the definition of $C(r)$ with $r = 1$, $a = a(T-1)$ and $b = a(T)$. □

Therefore if $C(1)$ holds at time $T-1$, the optimality equation for V_{T-1} can be simplified further via Lemma 1 into:

$$V_{T-1}(X_{T-1}) = \max_{a(T-1)=1,\ldots,n} \left\{ \begin{array}{c} \sum_{k=1}^{p} X_{a(T-1)}\Omega_k\vec{1}\cdot \max\limits_{\substack{a(T)=1,\ldots,n\\a(T)\neq a(T-1)}} \left\{X_{a(T)}\tau_u\rho\right\} \\ +X_{a(T-1)}\rho \\ +(X_{a(T-1),E}+X_{a(T-1),L})\cdot \max\limits_{\substack{a(T)=1,\ldots,n\\a(T)\neq a(T-1)}} \left\{X_{a(T)}\tau_u\rho\right\} \end{array} \right\} \tag{4.46}$$

The two inner maximizations are the same, so their coefficients can be added:

$$\begin{gathered} V_{T-1}(X_{T-1}) = \max_{a(T-1)=1,\ldots,n} \Bigg\{ X_{a(T-1)}\rho + \\ \left((\sum_{k=1}^{p} X_{i,T-1}\Omega_k\vec{1}) + (X_{a(T-1),E}+X_{a(T-1),L}) \right) \max_{\substack{a(T)=1,\ldots,n\\a(T)\neq a(T-1)}} \left\{X_{a(T)}\tau_u\rho\right\} \Bigg\} \end{gathered} \tag{4.47}$$

But it is easy to verify that $(\sum_{k=1}^{p} X_{a(T-1)}\Omega_k\vec{1}) + (X_{a(T-1),E}+X_{a(T-1),L}) = 1$, therefore we are left with the following optimality equation for $T-1$, given that $C(1)$ holds at time $T-1$:

$$V_{T-1}(X_{T-1}) = \max_{a(T-1)=1,\ldots,n} \left\{ X_{a(T-1)}\rho + \max_{\substack{a(T)=1,\ldots,n\\a(T)\neq a(T-1)}} \left\{X_{a(T)}\tau_u\rho\right\} \right\} \tag{4.48}$$

Again, this agrees with Theorem IV.1. We iterate this process one additional time to derive V_{T-2}. Applying Equation (4.38) at $t = T-2$, we get:

$$= \max_{a(T-2)=1,\ldots,n} \left\{ \begin{array}{c} \displaystyle\sum_{k=1}^{p} X_{a(T-2)}\Omega_k\vec{1} \cdot V_{T-1}\left(X_1\tau_u, \ldots, \overline{X_{a(T-2)}\Omega_k}\tau_s, \ldots, X_n\tau_u\right) \\ +X_{a(T-2)}\rho \\ +(X_{a(T-2),E} + X_{a(T-2),L}) \cdot V_{T-1}\left(X_1\tau_u, \ldots, \overline{X_{a(T-2)}\Omega_E}\tau_s, \ldots, X_n\tau_u\right) \end{array} \right\} \tag{4.49}$$

After substituting Equation (4.48) for V_{T-1}, we get the following for V_{T-2}

$$= \max_{a(T-2)=1,\ldots,n} \left\{ \begin{array}{c} \displaystyle\sum_{k=1}^{p} X_{a(T-2)}\Omega_k\vec{1}\cdot \\ \displaystyle\max_{\substack{a(T-1)=1,\ldots,n \\ a(T-1)\neq a(T-2)}} \left\{ \begin{array}{c} X_{a(T-1)}\tau_u\rho + \displaystyle\max_{\substack{a(T)=1,\ldots,n \\ a(T)\neq a(T-2) \\ a(T)\neq a(T-1)}} \{X_T\tau_u^2\rho, \overline{X_{a(T-2)}\Omega_k}\tau_s\tau_u\rho\}, \\ \overline{X_{a(T-1)}\Omega_k}\tau_s + \displaystyle\max_{\substack{a(T)=1,\ldots,n \\ a(T)\neq a(T-2)}} \{X_{a(T)}\tau_u^2\rho\} \end{array} \right\} \\ +X_{a(T-2)}\rho + \\ \left(X_{a(T-2),E} + X_{a(T-2),L}\right)\cdot \\ \displaystyle\max_{\substack{a(T-1)=1,\ldots,n \\ a(T-1)\neq a(T-2)}} \left\{ \begin{array}{c} X_{a(T-1)}\tau_u\rho + \displaystyle\max_{\substack{a(T)=1,\ldots,n \\ a(T)\neq a(T-2) \\ a(T)\neq a(T-1)}} \{X_{a(T)}\tau_u^2\rho, \overline{X_{a(T)}\Omega_E}\tau_s\tau_u\rho\}, \\ \overline{X_{a(T-1)}\Omega_E}\tau_s + \displaystyle\max_{\substack{a(T)=1,\ldots,n \\ a(T)\neq a(T-2)}} \{X_{a(T)}\tau_u^2\rho\} \end{array} \right\} \end{array} \right\} \tag{4.50}$$

Similar to the way that condition $C(1)$ resolved the nested max in the derivation of V_{T-1}, conditions $C(2)$ will resolve the nested maxes of V_{T-2}

Lemma 2. *Suppose $C(2)$ holds at time T. Then it follows that*

$$\max_{\substack{a(T)=1,\ldots,n \\ a(T)\neq a(T-2),a(T-1)}} \left\{X_T\tau_u^2\rho,\ \overline{X_{a(T-2)}\Omega_k}\tau_s\tau_u\rho\right\} = \max_{\substack{a(T)=1,\ldots,n \\ a(T)\neq a(T-2),a(T-1)}} \left\{X_T\tau_u^2\rho\right\}$$

and

$$\max_{\substack{a(T)-1=1,\dots,n\\ a(T)-1\neq a(T-2)}} \left\{X_{T-1}\tau_u^2\rho,\ \ \overline{X_{a(T-1)}\Omega_E}\tau_s\tau_u\rho\right\} = \max_{\substack{a(T-1)=1,\dots,n\\ a(T-1)\neq a(T-2)}} \left\{X_{T-1}\tau_u^2\rho\right\}$$

Proof. Both of these statements apply as a direct result of the definition of $C(r)$ with $r = 2$, $i = a(T-2)$ and $j = a(T-1)$. □

.

Therefore if $C(2)$ holds at time $T-1$, the inner-most nested maximzations simplify to give:

$$V_{T-2}(X_{T-2}) = \\ \max_{a(T-2)=1,\dots,n} \left\{ \begin{array}{c} \displaystyle\sum_{k=1}^{p} X_{a(T-2)}\Omega_k\vec{1}\cdot \max_{\substack{a(T-1)=1,\dots,n\\ a(T-1)\neq a(T-2)}} \left\{ \begin{array}{c} X_{a(T-1)}\tau_u\rho+ \\ \displaystyle\max_{\substack{a(T)=1,\dots,n\\ a(T)\neq a(T-2)\\ a(T)\neq a(T-1)}} \{X_T\tau_u^2\rho\}, \\ \overline{X_{a(T-2)}\Omega_k}\tau_s+ \\ \displaystyle\max_{\substack{a(T)=1,\dots,n\\ a(T)\neq a(T-2)}} \{X_{a(T)}\tau_u^2\rho\} \end{array} \right\} \\ +X_{a(T-2)}\rho \\ +\left(X_{a(T-2),E}+X_{a(T-2),L}\right)\cdot \max_{\substack{a(T-1)=1,\dots,n\\ a(T-1)\neq a(T-2)}} \left\{ \begin{array}{c} X_{a(T-1)}\tau_u\rho+ \\ \displaystyle\max_{\substack{a(T)=1,\dots,n\\ a(T)\neq a(T-2)\\ a(T)\neq a(T-1)}} \{X_T\tau_u^2\rho\}, \\ \overline{X_{a(T-2)}\Omega_E}\tau_s+ \\ \displaystyle\max_{\substack{a(T)=1,\dots,n\\ a(T)\neq a(T-2)}} \{X_{a(T)}\tau_u^2\rho\} \end{array} \right\} \end{array} \right\} \tag{4.51}$$

Lastly, if $C(1)$ also hold at $T-2$, the second inner-most nested maximizations become resolved to give:

$$V_{T-2}(X_{T-2}) = \max_{a(T-2)=1,\ldots,n} \left\{ \begin{array}{c} \displaystyle\sum_{k=1}^{p} X_{a(T-2)}\Omega_k\vec{1}\cdot \\ \displaystyle\max_{\substack{a(T-1)=1,\ldots,n\\ a(T-1)\neq a(T-2)}} \left\{ X_{a(T-1)}\tau_u\rho + \max_{\substack{a(T)=1,\ldots,n\\ a(T)\neq a(T-2)\\ a(T)\neq a(T-1)}} \{X_T\tau_u^2\rho\} \right\} \\ +X_{a(T-2)}\rho \\ +\left(X_{a(T-2),E}+X_{a(T-2),L}\right)\cdot \\ \displaystyle\max_{\substack{a(T-1)=1,\ldots,n\\ a(T-1)\neq a(T-2)}} \left\{ X_{a(T-1)}\tau_u\rho + \max_{\substack{a(T)=1,\ldots,n\\ a(T)\neq a(T-2)\\ a(T)\neq a(T-1)}} \{X_T\tau_u^2\rho\} \right\} \end{array} \right\} \tag{4.52}$$

And in a similar vein to the time $T-1$, the inner nested maximizations are now identical, allowing us to sum their coefficients $(\sum_{k=1}^{p} X_{a(T-2)}\Omega_k\vec{1}) + (X_{a(T-2),E} + X_{a(T-2),L}) = 1$. The final optimality equation for time $T-2$ is the following:

$$\begin{aligned} V_{T-2}(X_{T-2}) = \max_{a(T-2)=1,\ldots,n} \Bigg\{ & X_{a(T-2)}\rho + \\ & \max_{\substack{a(T-1)=1,\ldots,n\\ a(T-1)\neq a(T-2)}} \left\{ X_{a(T-1)}\tau_u\rho + \max_{\substack{a(T)=1,\ldots,n\\ a(T)\neq a(T-1)\\ a(T)\neq a(T-2)}} \left\{ X_{a(T)}\tau_u^2\rho \right\} \right\} \Bigg\} \end{aligned} \tag{4.53}$$

Again this agrees with Theorem IV.1. At this point, the method of recursively deriving V_{T-3}, V_{T-4}, and so forth will follow the same steps as used for V_T, V_{T-1}, V_{T-2}, and the structure of Equations (4.42), (4.48), and (4.53) will continue to follow the structure proposed in Theorem IV.1. □

We will spend the remainder of the analytical portion of this paper addressing three challenges in implementing this theorem: (1) The number of sufficient conditions to be checked is equal to the number of stages remaining, which can be very large. (2) The value function is currently in the unintuitive form of a sequence of nested maximizations, and (3) The number of nested maximizations increases by 1 for each prior stage, causing the problem to grow in complexity. We will hence present an alternative optimal policy which addresses all three issues. Table 4.2 organizes the method by which we accomplish this.

Optimality Equation	**Modification**	**Weakness**
Equation (4.9)		Recursive
Theorem IV.1	Non-recursive	Several conditions
Theorem IV.2	Single condition	Long cumbersome structure
Theorem IV.3	Simple truncated structure	Single patient per period
Equation (4.31)	Multiple patients per period	

Table 4.2: A table organizing the theoretical development of our screening policy.

4.3.4 Reduction of the Optimal Policy

We will first seek to reduce the number of sufficient conditions necessary for our Theorem IV.1 by exploiting properties of coefficients of the condition $C(r)$, $\tau_u^r\rho$ and $\tau_s\tau_u^r\rho$. In particular, consider the j-th component as a sequence in r.

Definition 2.

$$\alpha_j(r) := [\tau_u^r\rho]_j \tag{4.54}$$

$$\beta_j(r) := [\tau_u^{r-1}\tau_s\rho]_j \tag{4.55}$$

The definitions of these four indices are shown graphically in Figure 4.5. Notice that $\alpha_j(r)$ is non-decreasing until δ_j, and decreases beyond its first term at λ_j (or $\beta_j(r)$, ϵ_j, and μ_j, respectively.)

Lemma 3. *For every $j = 1, ..., m$, there exists indices $\delta_j, \epsilon_j, \lambda_j, \mu_j$ such that*

the sequences $\alpha_j(r)$ and $\beta_j(r)$ are

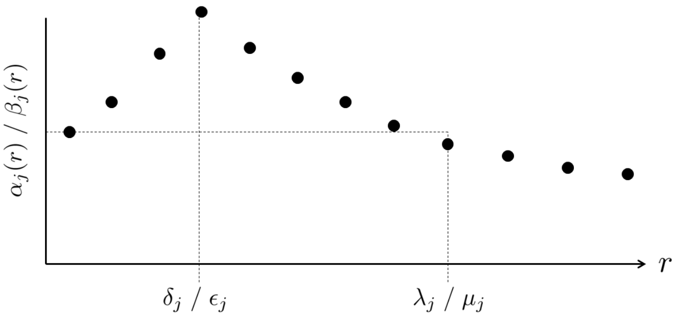


Figure 4.5: A depiction of the relationship between $\alpha_j(r)$, δ_j, and λ_j. (or $\beta_j(r)$, ϵ_j, and μ_j, respectively.)

1. *non-decreasing from* $r = 1, ..., \delta_j$ *and* $r = 1, ..., \epsilon_j$*, respectively, and non-increasing thereafter.*

2. *guaranteed to decrease beyond the first term in the sequence at indices* λ_j *and* μ_j*, respectively.*

The indices δ_j, ϵ_j are necessary for the definition of an alternative set of conditions, which are easier to check than the original set of conditions. The purpose of the indices λ_j, μ_j are to create a less computational expensive version of our optimal policy.

Proof. The coefficients $\alpha_j(r)$ can be written in terms of the primitive data. Simple matrix multiplication gives

$$\alpha_j(r) = [\tau_u^r \rho]_j = \sum_{s=0}^{n-1} P_{jj}^s P_{jE} P_{EE}^{n-s-1} \tag{4.56}$$

The results of this proposition will be apparent after rewriting this coefficient sequence in a recursive form. After pulling the $s = 0$ term out of the summation, we get

$$= P_{jE} P_{EE}^{n-1} + \sum_{s=1}^{n-1} P_{jj}^s P_{jE} P_{EE}^{n-s-1} \tag{4.57}$$

We then use a change of variables in the summation:

$$= P_{jE}P_{EE}^{n-1} + \sum_{s=0}^{n-2} P_{jj}^{s+1} P_{jE} P_{EE}^{n-s-2} \tag{4.58}$$

Then factor out a single common term from the summation

$$= P_{jE}P_{EE}^{n-1} + P_{jj}\left(\sum_{s=0}^{n-2} P_{jj}^{s} P_{jE} P_{EE}^{n-s-2}\right) \tag{4.59}$$

We can now re-apply the original formula for the coefficient sequence

$$\alpha_j(r) = P_{jE}P_{EE}^{n-1} + P_{jj}\left(\alpha_j(r-1)\right) \tag{4.60}$$

A similar method gives a recursive formula for $\beta_j(r)$:

$$\beta_j(r) = P_{jj}P_{jE}P_{EE}^{r-2} + P_{jj}\left(\beta_j(r-1)\right) \tag{4.61}$$

The recursive form of these sequences reveals that each subsequent term can be obtained by multiplying by the same factor which is strictly less than 1, (P_{jj}), and then the addition of a term which tends to zero $(P_{jj}P_{jE}P_{EE}^{r-2})$. This guarantees that the sequences will first be non-decreasing, then non-increasing thereafter, thus proving the existence and uniqueness of indices δ_j, ϵ_j. We can also apply limits across Equation (4.60) to get

$$\lim_{n\to\infty} \alpha_j(r) = \lim_{n\to\infty}\left(P_{jE}P_{EE}^{n-1} + P_{jj}\alpha_j(r-1)\right) \tag{4.62}$$

Then because limits distribute across sums

$$= \lim_{n\to\infty} P_{jE}P_{EE}^{n-1} + \lim_{n\to\infty} P_{jj}\alpha_j(r-1) \tag{4.63}$$

The first additive term becomes arbitrarily small in the limit because $P_{EE} < 1$

therefore

$$= \lim_{n \to \infty} P_{jj} \alpha_j(r-1) \tag{4.64}$$

Which is equal to

$$= P_{jj} \lim_{n \to \infty} \alpha_j(r-1) \tag{4.65}$$

However we know the sequences on the left-hand and right-hand side must have the same limit, because they are the same sequence, only shifted in index. In general, the equation $x = P_{jj}x$ has only two solutions: either $P_{jj} = 1$ or $x = 0$. The first is a contradiction of our model's assumptions, so therefore we have the unique solution:

$$\lim_{n \to \infty} \alpha_j(r-1) = 0 \tag{4.66}$$

Therefore λ_j is guaranteed to exist and to be unique. A similar proof gives the existence and uniqueness of μ_j. □

Definition 3.

$$\alpha_j^*(r) = \begin{cases} \alpha_j(\lambda_j) & \text{if } r \leq \delta_j \\ \alpha_j(r) & \text{if } r > \delta_j \end{cases} \qquad \beta_j^*(r) = \begin{cases} \beta_j(r) & \text{if } r \leq \epsilon_j \\ \beta_j(\epsilon_j) & \text{if } r > \epsilon_j \end{cases} \tag{4.67}$$

α^* is the sequence α, modified to be non-increasing. Similarly, β^* is the sequence β modified to be non-decreasing. Now consider $C^*(r)$, which looks exactly the same in form as $C(r)$, except that it uses α^* and β^*, instead of α and β. These new sequences are shown graphically in Figures 4.6 and 4.7.

Definition 4. The condition $C^*(r)$ is said to be satisfied by the belief state X_t at time t if the following holds true:

For every $b \in \{1, ..., n\}$, there exists a corresponding $a \in \{1, ..., n\}, b \neq a$ such that

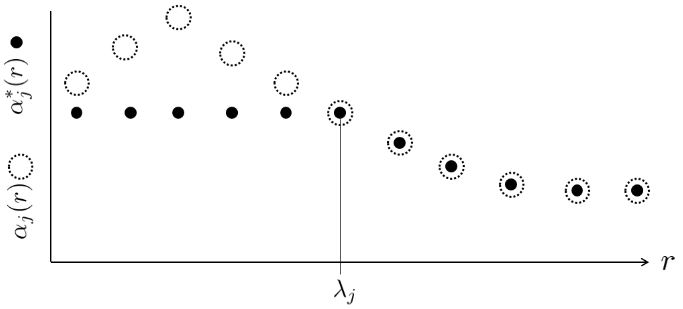


Figure 4.6: A depiction of the relationship between α^* and α.

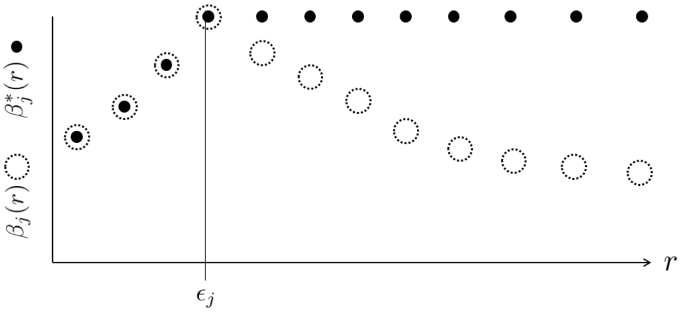


Figure 4.7: A depiction of the relationship between β^* and β.

$$X_a\alpha^*(r) \geq \overline{X_b\Omega_k}\beta^*(r)$$

for every possible $k \in \{1, ..., p\}$.

$C^*(r)$ has the advantage of the following two properties.

Lemma 4. *Suppose $C^*(r)$ holds at time t. Then $C(r)$ also holds at time t.*

Lemma 5. *Suppose $C^*(r)$ holds at time t. Then $C^*(r-1)$ holds at time t.*

Proof. Suppose $C^*(r)$ holds at time t. Then by definition, $\forall b \in \{1, ..., n\}, \exists a \in \{1, ..., n\}, b \neq a$ such that

$$X_a\alpha^* \geq \overline{X_b\Omega_k}\beta^* \tag{4.68}$$

But we know by construction that $\alpha(r) \geq \alpha^*(r)$, $\forall r$ and $\beta^*(r) \geq \beta(r)$, $\forall r$ Therefore, it is a direct consequence that

$$X_a\alpha(r) \geq X_a\alpha^*(r) \geq \overline{X_b\Omega_k}\beta^*(r) \geq \overline{X_b\Omega_k}\beta(r) \tag{4.69}$$

Therefore $C(r)$ holds. □

Proof. Suppose $C^*(r)$ holds at time t. Then by definition, $\forall b \in \{1, ..., n\}, \exists a \in \{1, ..., n\}, j \neq i$ such that

$$X_b\alpha^* \geq \overline{X_a\Omega_k}\beta^* \tag{4.70}$$

We know that α^* is non-increasing, and β^* is non-decreasing. Therefore $\alpha^*(r-1) \geq \alpha^*(r)$ and $\beta^*(r) \geq \beta^*(r-1)$. Therefore, we can write the following chain of inequalities:

$$X_a\alpha^*(r-1) \geq X_a\alpha^*(r) \geq \overline{X_b\Omega_k}\beta^*(r) \geq \overline{X_b\Omega_k}\beta^*(r-1) \tag{4.71}$$

Therefore $C^*(r)$ holds. □

We can now combine the two properties of $C^*(r)$ to reduce the many sufficient conditions of Theorem IV.1 from $C(1), C(2), ..., C(r)$ to just the single sufficient condition $C^*(r)$.

Theorem IV.2. *At every time $t = 1, 2, ..., T$, let $r = T - t$. If $C^*(r)$ holds at time t then the optimality equation is*

$$V_t(X_t) = \max_{a(t)=1,...,n} \Bigg\{ X_{a(t)}\rho + \max_{\substack{a(t+1)=1,...,n \\ a(t+1)\neq a(t)}} \Big\{ X_{a(t+1)}\tau_u\rho + ... \max_{\substack{a(T)=1,...,n \\ a(T)\neq a(T-1),a(T-2),...,a(t)}} \big\{ X_{a(T)}\tau_u^r\rho \big\} \Big\} \Bigg\} \tag{4.72}$$

$$\begin{tikzcd}
C^*(r) \arrow[r, Rightarrow] \arrow[d, Rightarrow] & C^*(r-1) \arrow[r, Rightarrow] \arrow[d, Rightarrow] & \dots \arrow[r, Rightarrow] & C^*(1) \arrow[d, Rightarrow] \\
C(r) & C(r-1) & \dots & C(1)
\end{tikzcd}$$

Figure 4.8: The structure of the proof of Theorem 2.

Furthermore, it is optimal to screen patient $a(t)$ at time t.

Proof. This comes as an application of the combination of Lemmas 4 and 5 to Theorem IV.1. Suppose $C^*(r)$ holds at time $T - r$. Then by repeated applications of Lemma 5, $C^*(r-1)$, $C^*(r-2)$, ... , $C^*(1)$ also hold. Consequently, r applications of Lemma 4 guarantees that $C(r-1)$, $C(r-2)$,..., $C(1)$ also hold. Therefore the conditions of Theorem IV.1 are met.

The logical deduction of this proof is depicted in Figure 4.8.

□

Theorem IV.2 only requires a single condition, and therefore is a reduced form of Theorem IV.1 which requires many conditions. We now seek to further reduce the complexity of the policy by investigating the expression needed to be maximized.

Lemma 6. *Let $c_1, c_2, ...c_r \in \mathbb{R}^{(1\times n)}$ and $X_1, X_2, ..., X_n \in \mathbb{R}^{(n\times 1)}$ Then*

$$\max_{a(1)\in\{1,...,n\}} \left\{ c_1 X_{a(1)} + \max_{\substack{a(2)\in\{1,...,n\} \\ a(2)\neq a(1)}} \left\{ c_2 X_{a(2)} + ... \max_{\substack{a(r)\in\{1,...,n\} \\ a(1)\neq a(1),a(2),...,a(r-1)}} c_r X_{a(r)} \right\} \right\}$$

$$=$$

$$\max_{\{a(1),...,a(r)\}\subset\{1,...,n\}} \left\{ c_1 X_{a(1)} + ... + c_r X_{a(r)} \right\}$$

Furthermore, the arguments which maximize the left-hand side expression also maximize the right-hand side expression.

When applied with $c_i = \tau_u^{i-1}\rho$, lemma 6 allows us to reduce Equation 4.72 into a simpler combinatorial maximization. The latter expression is easier to compute and simpler to understand.

Proof. We prove this statement for any fixed n, and by induction on the number of summands r.

Clearly this statement is true for the case of a single summand, because they equate the exactly the same expression.

Now suppose that this statement holds true for r summands.

$$\max_{a(r)\in\{1,...,n\}} \left\{ c_r X_{a(r)} + \max_{\substack{a(r-1)\in\{1,...,n\}\\ a(r-1)\neq a(r)}} \left\{ c_{r-1} X_{a(r-1)} + ... \max_{\substack{a(1)\in\{1,...,n\}\\ a(1)\neq a(2),a(3),...,a(r)}} c_1 X_{a(1)} \right\}\right\}$$
$$=$$
$$\max_{\{a(1),...,a(r)\}\subset\{1,...,n\}} \left\{ c_r X_{a(r)} + ... + c_1 X_{a(1)} \right\} \tag{4.73}$$

We can use this inductive hypothesis to simplify the claim for $r+1$ summands.

$$\max_{a(r+1)\in\{1,...,n\}} \left\{ c_{r+1} X_{a(r+1)} + \max_{\substack{a(r)\in\{1,...,n\}\\ a(r)\neq a(r+1)}} \left\{ c_r X_{a(r)} + \max_{\substack{a(r-1)\in\{1,...,n\}\\ a(r-1)\neq a(r),a(r+1)}} \left\{ c_{r-1} X_{a(r-1)} + \right.\right.\right.$$
$$\left.\left.\left. ... + \max_{\substack{a(1)\in\{1,...,n\}\\ a(1)\neq a(2),a(3),...,a(r),a(r+1)}} c_1 X_{a(1)} \right\}\right\}\right\}$$
$$=$$
$$\max_{a(r+1)\in\{1,...,n\}} \left\{ c_{r+1} X_{a(r+1)} + \left[\max_{\{a(1),...,a(r)\}\subset\{1,...,n\}\backslash\{a(r+1)\}} \left\{ c_r X_{a(r)} + ... + c_1 X_{a(1)} \right\} \right] \right\} \tag{4.74}$$

We can move the term $c_{r+1} X_{a(r+1)}$ inside the inner max because it is not a function of the arguments being maximized.

$$= \max_{a(r+1)\in\{1,...,n\}} \left\{ \max_{\{a(1),...,a(r)\}\subset\{1,...,n\}\backslash\{a(r+1)\}} \left\{ c_{r+1} X_{a(r+1)} + c_r X_{a(r)} + ... + c_1 X_{a(1)} \right\} \right\} \tag{4.75}$$

We are now maximizing a single expression first over two sets of arguments. However the arguments are guaranteed to have no intersection, so it is possible to evaluate both simultaneously.

$$= \max_{\substack{a(r+1)\in\{1,...,n\} \\ \{a(1),...,a(r)\}\subset\{1,...,n\}\setminus\{a(r+1)\}}} \left\{c_{r+1}X_{a(r+1)} + c_rX_{a(r)} + ... + c_1X_{a(1)}\right\} \tag{4.76}$$

Which is equivalent to:

$$= \max_{\{a(1),...,a(r),a(r+1)\}\subset\{1,...,n\}} \left\{c_{r+1}X_{a(r+1)} + c_rX_{a(r)} + ... + c_1X_{a(1)}\right\} \tag{4.77}$$

Therefore maximizing the two expressions are one in the same. □

Lemma 7. *Let $c_1, c_2, ...c_r \in \mathbb{R}^{(1\times n)}$ and $X_1, X_2, ..., X_n \in \mathbb{R}^{(n\times 1)}$ and Suppose $\exists\lambda$ such that every vector $c_1, ..., c_{\lambda-1}$ strictly dominates every vector $c_\lambda, ..., c_r$ in every component. Then the first $a(1), ..., a(\lambda-1)$ arguments which maximize*

$$\max_{\{a(1),...,a(r)\}\subset\{1,...,n\}} \left\{c_1X_{a(1)} + ... + c_rX_{a(r)}\right\} \tag{4.78}$$

will also maximize

$$\max_{\{a(1),...,a(\lambda-1)\}\subset\{1,...,n\}} \left\{c_1X_{a(1)} + ... + c_{\lambda-1}X_{a(\lambda-1)}\right\} \tag{4.79}$$

This lemma states that in solving the large combinatorial maximization of lemma 6, it will be sufficient to consider a truncated expression if the coefficients c_r eventually become strictly smaller than the first coefficient c_1.

Proof. Notice that if every vector $c_1, ..., c_{\lambda-1}$ strictly dominates every vector $c_\lambda, ..., c_r$ in every component, then this problem can be solved in two discrete stages without any loss of optimality. That is the first $1, ..., \lambda-1$ terms can be maximized without regard to the final $\lambda, ..., r$ terms. □

We now combine Lemmas 6 and 3 to further reduce Theorem IV.2.

Theorem IV.3. *At every time $t = 1, 2, ..., T$, let $r = T - t$. If $C^*(r)$ holds at time t, then there exists λ such that, to find the optimal patient to screen, it is sufficient to consider*

$$\max_{\{a(t),...,a(t+\lambda)\}\subset\{1,...,n\}} \left\{X_{a(t)}\rho + X_{a(t+1)}\tau_u\rho + ...X_{a(t+\lambda)}\tau_u^{\lambda-1}\rho\right\} \tag{4.80}$$

Furthermore, it is optimal to screen patient $a(t)$ at time t.

Proof. We know from Theorem IV.2 that the optimality equation at time t is

$$V_t(X_t) = \max_{a(t)=1,...,n} \left\{ X_{a(t)}\rho + \max_{\substack{a(t+1)=1,...,n \\ a(t+1)\neq a(t)}} \left\{ X_{a(t+1)}\tau_u\rho \right.\right.$$
$$\left.\left. + \ldots \max_{\substack{a(T)=1,...,n \\ a(T)\neq a(T-1),a(T-2),...,a(t)}} \left\{X_{a(T)}\tau_u^r\rho\right\} \right\}\right\} \tag{4.81}$$

However we can apply Lemma 6 with $c_r := \tau_u^r\rho$ to turn this optimality equation from a sequence of nested maxes into a combinatorial maximization.

$$V_t(X_t) = \max_{\{a(t),...,a(t+r)\}\subset\{1,...,n\}} \left\{X_{a(t)}\rho + X_{a(t+1)}\tau_u\rho + ...X_{a(t+r)}\tau_u^r\rho\right\} \tag{4.82}$$

Furthermore, we know from Lemma 3 that there exists λ_j such that $[\tau_u^r]_j$ is strictly less than $[\tau_u^1]_j$, $\forall r > \lambda_j$. So define $\lambda := \max_j \lambda_j$. Since our coefficients satisfy the strict domination requirement of Lemma 7, and it can now be applied with $c_r := \tau_u^r\rho$. Therefore to find the optimal patient to screen now, it is sufficient to consider the problem

$$\max_{\{a(t),...,a(t+\lambda)\}\subset\{1,...,n\}} \left\{X_{a(t)}\rho + X_{a(t+1)}\tau_u\rho + ...X_{a(t+\lambda)}\tau_u^{\lambda-1}\rho\right\} \tag{4.83}$$

□

4.3.5 Managerial Insights

The optimal policy dictated by Theorem IV.3 gives us several insights into the nature of effective screening. The most immediate lesson is that it is not necessarily optimal to screen the patient most likely to currently have early-stage cancer. One can construct a set of belief states where myopic behavior will not be optimal. This is due to the fact that the future value of learning may outweigh the current value of detecting cancer. This may come as a surprise to many clinicians.

In fact, the dependance of our policy upon the transition probability matrix τ_U tells us how disease progression will influence optimal behavior. A more aggressively evolving disease will encourage the screening of patients with current higher risk of early-stage cancer, whereas a slower evolving disease will shift optimal screening decisions towards patients for whom less is known, placing more benefit upon learning. The knowledge of a disease's evolution and natural history can guide clinicians on how to trade-off between exploration and exploitation.

It is surprising that the observation probability matrix Ω is absent from Equation (4.80). Ω was used to update belief states upon past observations, however it should not influence future behavior. The relative value of screening a patient derives solely from their predicted disease progression given their current belief state, and not on any potential belief states that could result from future learning.

We can gain further insight into efficacious screening behavior by looking at the structure of the policy. The expression in Equation (4.80) is precisely the expected value of screening patient a_t now, a_{t+1} in the next period, then a_{t+2} and so forth for the next λ periods. A clinic only need to consider a time horizon of λ periods into the future in order to behave optimally in the current period. Interestingly, λ is guaranteed in Theorem IV.3 to only be a function of τ_U which captures the disease dynamics. In particular, λ is the same, regardless of the distribution of belief states X_t, the number of time periods remaining t, and even the panel size n. Therefore, as long as Theorem IV.3's conditions continue to hold, the same truncated planning horizon can be used for all clinics screening the same disease, regardless of size and situation.

4.4 Case Study

The analytical results derived in the first part of this paper will now be evaluated in a numerical case study. We develop a computer simulation of a panel of patients at risk for HCC who are screened according to both our policy and current practice. Individual patient disease progressions, as well as outcomes of hypothetical screening events, are drawn from historical patient data. IRB approval was obtained for the collection and usage of the data for this study (HUM00088566).

4.4.1 Sources of Data

Longitudinal data on disease progression for patients at risk for HCC were acquired from two independent sources: the Hepatitis C Antiviral Long-term Treatment against Cirrhosis clinical trial (HALT-C), and the University of Michigan Health System Hospital's records (UM). The characteristics of the two datasets are given in Table 1.

The HALT-C dataset followed 1050 patients for an average of 5.3 years. The level of alpha-fetoprotein (AFP) was measured every 3 months for the first 3.5 years, then every 6 months thereafter on a voluntary basis. Patients received an ultrasound imaging every 6-12 months. 946 patients remained after our exclusion criteria of having fewer than 2 AFP readings, and/or any AFP reading more than 5 standard deviations above the mean.

For the UM dataset, the authors manually collected individual patient charts of all patients enrolled in the hospital's HCC screening program from the dates of May 1, 2004 to May 1, 2014. From these charts, we extracted all relevant demographics

Characteristic	**HALT-C** (N=946)	**UM** (N=820)
Age at Baseline (years)	50.2 ± 7.2	55.9 ± 11.0
Ever a Smoker	75.3%	57.7%
Length of Follow-up (years)	5.3 ± 1.8	3.3 ± 1.2
Time Between Screenings (days)	126.3 ± 29.7	205.0 ± 77.8
% Developed Cancer	7.7%	5.0%

Table 4.3: Summary statistics of the two independent datasets used. Statistics are reported as mean ± standard deviation.

and screening results related to disease progression. In total, 820 patient charts met the inclusion criteria of (1) having at least 2 AFP readings on record, and (2) having at least 1 abdominal imaging on record.

The HALT-C dataset was used to parameterize the analytical model. Clinical trials benefit from strict inclusion criteria, as well as the more regularly administered screenings, thus giving a better estimate of the underlying disease progression. We then applied the parameterized model to a simulation built upon the UM dataset, for a better estimate of how our policy would perform in practice with complications, such as non-adherence and co-morbidities.

4.4.2 Model Parametrization

To implement our screening policy, the decision maker needs three components: (1) τ, the transition probability matrix which reflects the decision maker's beliefs about how the disease progresses (2) Ω, the observation probability matrix, which will be used to update the decision maker's beliefs upon observing screening results, and (3) the function $\beta(\cdot)$, which translates baseline information of newly entering patients into initial belief states.

Figure 4.9 provides a graphical depiction of how the HALT-C dataset was used to obtain these three necessary components. The HALT-C dataset contains two types of patient records: (1) the date and value of each patient's AFP readings over time, and (2) risk factors for HCC for each patient, measured at baseline enrollment. This

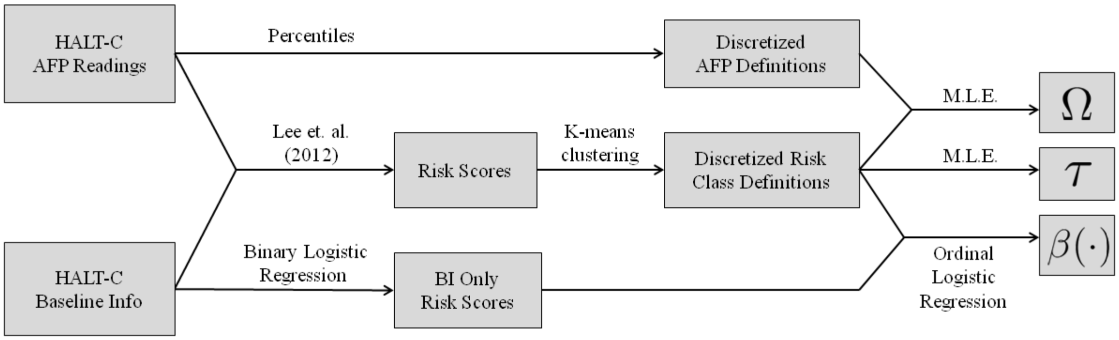


Figure 4.9: The use of the HALT-C dataset to parameterize the model.

baseline information includes age, race, smoking status, blood platelet count, alkaline phosphatase level, and presence/absence of esophageal varices.

We now provide the numerical values determined from parameterizing our model to the HALT-C dataset. To determine the discretized definitions for AFP observations, we used the 33rd and 67th percentile of all AFP readings observed throughout the history of HALT-C. We provide the thresholds to divide a patient's risk score into discretized ranges in Table 4.4

Low Risk	0 - 0.039
Medium Risk	0.040 - 0.069
High Risk	>0.070

Table 4.4: Definition of continuous risk scores as discretized risk classes.

These are the transition probabilities between those risk classes as a result of those definitions in Equation 4.84.

$$\tau_U = \begin{array}{l} \\ \text{Low Risk} \\ \text{Med Risk} \\ \text{High Risk} \\ \text{Early Cancer} \\ \text{Death/Late} \end{array} \begin{array}{c} \begin{array}{ccccc} \text{Low Risk} & \text{Med Risk} & \text{High Risk} & \text{Early Cancer} & \text{Death/Late} \end{array} \\ \left(\begin{array}{ccccc} 0.976 & 0 & 0 & 0.018 & 0.006 \\ 0 & 0.973 & 0 & 0.021 & 0.006 \\ 0 & 0 & 0.966 & 0.028 & 0.006 \\ 0 & 0 & 0 & 0.872 & 0.128 \\ 0 & 0 & 0 & 0 & 1 \end{array}\right) \end{array} \tag{4.84}$$

We then provide the thresholds to divide AFP into discretized ranges in Table 4.5, followed by the probability of observing each range as a function of risk class in Equation 4.85. Together, these compose the parameters τ_U and Ω used in our case study.

Low AFP Reading	0 - 2.39
Medium AFP Reading	2.40 - 4.39
High AFP Reading	>4.40

Table 4.5: Definition of continuous AFP readings as discretized observations.

$$\Omega = \begin{array}{l} \\ \text{Low Risk} \\ \text{Med Risk} \\ \text{High Risk} \end{array} \begin{array}{c} \begin{array}{ccc} \text{Low AFP} & \text{Med AFP} & \text{High AFP} \end{array} \\ \left(\begin{array}{ccc} 0.40 & 0.35 & 0.24 \\ 0.22 & 0.32 & 0.46 \\ 0.05 & 0.28 & 0.67 \end{array}\right) \end{array} \tag{4.85}$$

To determine the discretized risk classes, we performed a logistic regression on baseline risk factors and AFP readings against the development of HCC, similar to the methods in *Lee et al.* (2012a). This logistic model resulted in a risk value associated with each patient reflecting his/her predicted probability of developing HCC, given AFP information. We then performed a k-means cluster analysis of these scores. A choice of 3 clusters was decided for this model, because it achieved the minimum Akaike Information Criterion (AIC) of 7.09.

Once the definitions for discretized risk classes and AFP observations were determined, model parameterization was performed in a frequentist approach. To obtain the desired observation probabilities, we counted the number of times each discrete screening outcome was observed for all patients of a particular discrete risk class, and then divided by the total number of times patients in that discrete risk class were screened. To obtain the transition probabilities, we counted the number of early stage cancers developed, then divided it by the number of 30 day periods patients of a particular risk class did not develop cancer. Under the assumption that patients in the dataset evolved according to the model proposed, this method provides the maximum likelihood estimate for the desired transition and observation probabilities.

Lastly, the decision maker requires the function $\beta(\cdot)$ in order to initialize the beliefs over new patients entering the simulation. Presumably, the decision maker would have complete knowledge of the baseline risk information of the patient (i.e. age, race, smoking status) but none of the future AFP-related risk information. We parameterized a logistic regression without any future AFP information, in order to produce a Baseline Information Only (BI) risk score.

We then parameterized an ordinal logistic regression on the BI risk score to estimate how the BI only risk score predicted a patient's true discretized risk class. This analysis provided us with the function $\beta(\cdot)$ which translates a patient's BI risk score into probabilities of being a particular discretized risk class, C. The results of this analysis are shown in Equation (4.86).

$$Pr(C \leq j) = \frac{1}{1 + e^{-(-\alpha_j + \beta BI)}} \quad \text{for } j = \text{Low, Medium, High} \tag{4.86}$$

$$\beta = -22.59,\ \alpha_{Low} = 4.10,\ \alpha_{Medium} = 8.09$$

Once the model is parameterized, we tested its performance through a computer simulation built upon the UM dataset.

4.4.3 The Screening Simulation

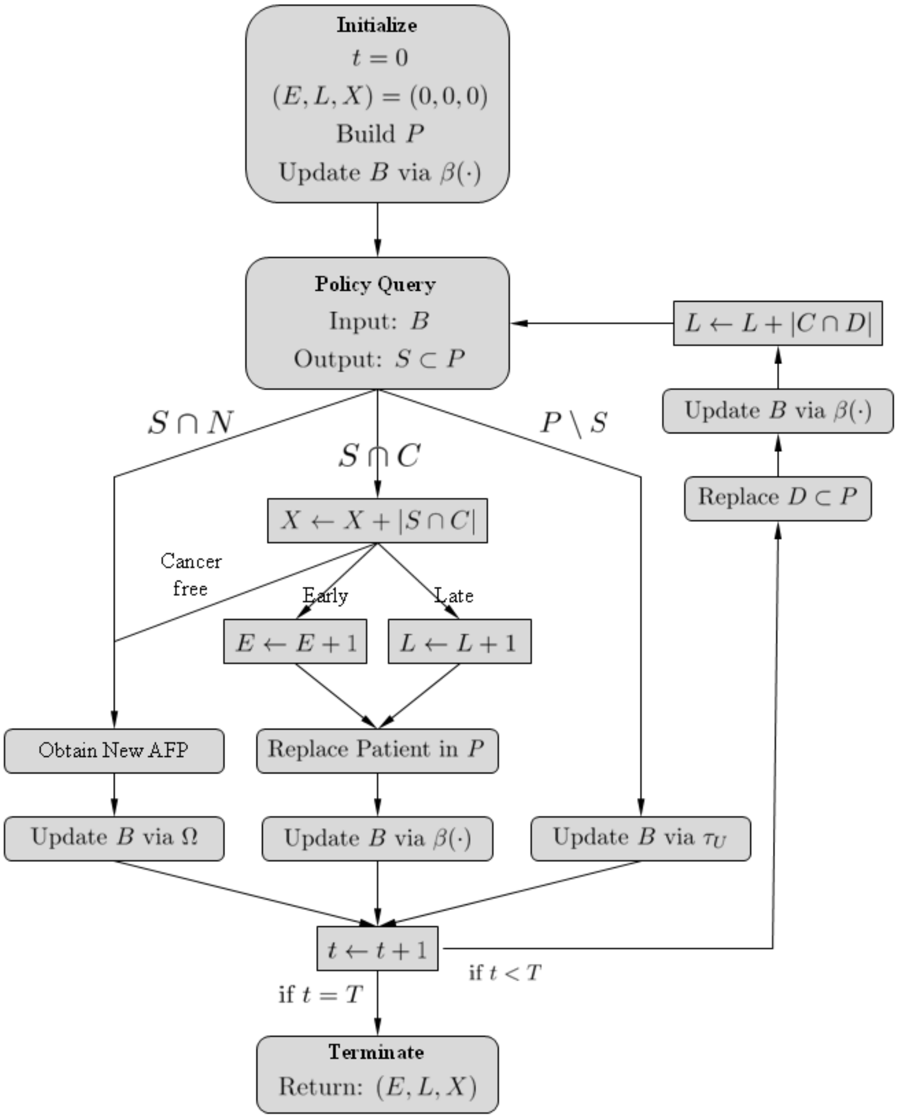


Figure 4.10: Discrete event simulation event logic.

The discrete event simulation event logic is depicted graphically in Figure 4.10, similar to that of Chapter III. The simulation keeps track of three intermediate statistics: E, early stage cancers detected, L, late stage cancers detected, and X, screenings spent on patients who eventually develop cancer.

The simulation begins by building an initial patient panel P. The simulation chooses a historical patient from the UM dataset, with replacement, to fill each of the $|P|$ panel slots. From this point on, there are two separate sets of knowledge:

1. P, the true disease history of each patient according to HALT-C, including all AFP readings, and if/when each patient first develops cancer. P is made of two subsets: C patients who will eventually develop cancer, and N the patients who will not. Note that the policy is unaware of which patients will and will not develop cancer.

2. B, the decision maker's belief state of the patients in P. B is initialized by using the function $\beta(\cdot)$, which translates baseline information about a patient into an initial belief state.

Once the simulation is initialized, the main loop begins. Each decision epoch begins by querying the screening policy for which patients $S \subset P$ to screen, given a current belief B. We screen multiple patients per period (see Section 4.4.4) because operational procedures often require patients to be scheduled in batches.

For all patients who are chosen to be screened who will never develop cancer, $S \cap N$, we find the two AFP readings in UM which are closest in date to the current simulation date t, and generate a new reading for this patient via linear interpolation. This will then be used to update the decision maker's belief state of this patient in B through the usage of Ω according to the Equation (4.26).

For all patients who are chosen to be screened who will eventually develop cancer, $S \cap C$, we increment the statistic X by $|S \cap C|$ to record that we have efficiently used a screening. Next, we check the patient's true disease state by comparing the current simulation date t to each patient's date of cancer development in UM.

If a chosen patient is cancer-free, as for patients in $S \cap N$, we perform the same process of generating a new AFP reading and updating the decision maker's belief about this patient in B.

If a chosen patent is currently in early-stage cancer, we assume it is detected with perfect accuracy, increment E by 1, then replace this patient in P with a new patient drawn from UM. B is also updated via $\beta(\cdot)$, according to the newly incoming patient's baseline information. Similarly, if the chosen patient is currently in late-stage cancer, we increment L by 1, and replace this patient in P in the same manner, we update B accordingly for this new patient.

For all unscreened patients $P \setminus S$, no new information will be observed. However the decision maker still knows that their underlying disease will progress, and therefore their beliefs are updated through τ_U according to Equation (4.24).

Lastly, before the beginning of the next decision epoch, we search for patients who leave the screening program, D, before the next time period. The patients in UM may have left voluntarily, due to cancer-related death, or due to non-cancer related death. These patients are replaced in P and their corresponding belief state in B is updated according to their chosen replacement. We increment the penalty statistic L by $|C \cap D|$ for any cancer patients who expired from the system, due to the failure to detect their cancer.

4.4.4 Implementation of Our Policy in Practice

The policy developed in Theorem IV.3 is not conveniently implementable within most common hospital operations. Most hospitals and screening clinics set their schedules and appointment in batches, and it would not be convenient to execute a screening policy which must alternate between screening a single patient, observing their result, then setting an appointment with the next single patient. Therefore, we adapt our policy to screen multiple patients in each decision epoch in order to create a policy which is implementable within the operational capabilities of a real clinic.

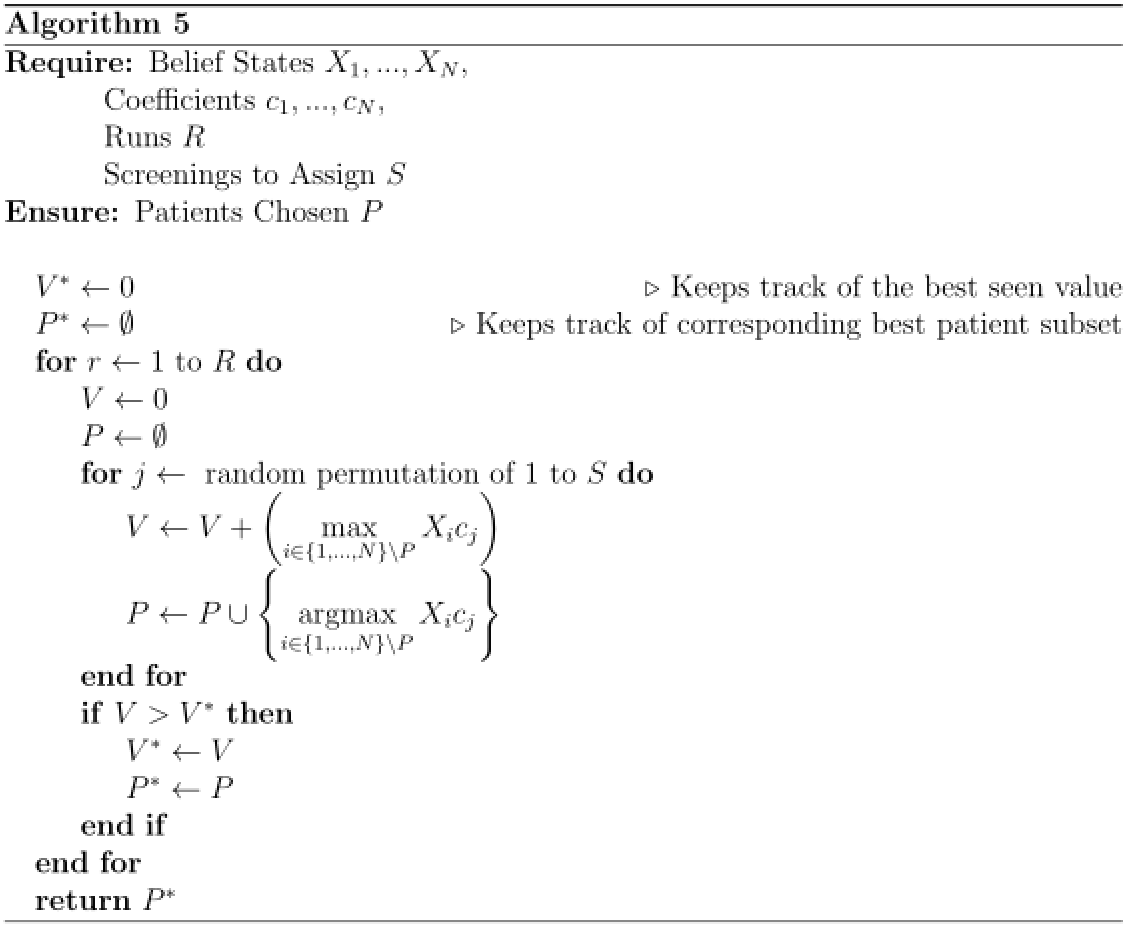
For this reason, in our case study, the decision maker will use the following modification of the derived optimal policy: in each decision epoch, given a set of current belief states $\{X_1, X_2, ..., X_n\}$, solve the problem:

$$\max_{\substack{\{a(1),...,a(s)\}\subset \\ \{1,...,.n\}}} \left\{X_{a(1)}R + X_{a(2)}\tau R + X_{a(3)}\tau^2 R + ... + X_{a(s-1)}\tau^{s-2}R + X_{a(s)}\tau^{s-1}R\right\} \tag{4.87}$$

where s is the number of patients needed to be screened every 30 days to achieve an equivalent rate of screening as current practice. The modification of the optimal policy will be to screen all patients $X_{a(1)}, ..., X_{a(s)}$, instead of just the single patient $X_{a(1)}$. Notice that this policy has no guarantee of optimality for the problem where multiple patients are screened each period. Nevertheless, we will demonstrate that this policy performs very well in practice. It should be noted that solving multi-armed bandits with multiple plays have been shown to be non-trivial (see *Pandelis and Teneketzis* (1999)).

The combinatorial optimization problem in Equation (4.87) has computational complexity $\mathcal{O}(\binom{n}{s}m^3s^2)$, when there are n patients, s screenings per period, and there are m risk types. Exact solution to this problem would be too time consuming to implement realistically at our partnering hospital, even for modest problem sizes. We sought to develop a computational heuristic which estimates Equation (4.87) with high accuracy and significantly less time than an exact solution.

The key idea behind our algorithm is that while Equation (4.87) is difficult to maximize combinatorially, it is simple to maximize each summand individually. Therefore, we propose to randomize the order of the summands, and to choose a belief state, without replacement, to maximize each summand. This construction of the desired expression may be suboptimal. To alleviate this, the randomized process is repeated for R runs, constructing a new expression in each run. Upon termination, the algorithm outputs the k belief states which achieved the highest value across all runs.



```
Algorithm 5
Require: Belief States $X_1, ..., X_N$,
        Coefficients $c_1, ..., c_N$,
        Runs $R$
        Screenings to Assign $S$
Ensure: Patients Chosen $P$

  $V^* \leftarrow 0$                                  ▷ Keeps track of the best seen value
  $P^* \leftarrow \emptyset$                          ▷ Keeps track of corresponding best patient subset
  for $r \leftarrow 1$ to $R$ do
      $V \leftarrow 0$
      $P \leftarrow \emptyset$
      for $j \leftarrow$ random permutation of 1 to $S$ do
          $V \leftarrow V + \left( \max_{i \in \{1,...,N\} \setminus P} X_i c_j \right)$
          $P \leftarrow P \cup \left\{ \operatorname{argmax}_{i \in \{1,...,N\} \setminus P} X_i c_j \right\}$
      end for
      if $V > V^*$ then
          $V^* \leftarrow V$
          $P^* \leftarrow P$
      end if
  end for
  return $P^*$
```

Figure 4.11: Pseudocode for the heuristic solution algorithm.

Note that this algorithm has computational complexity $\mathcal{O}(nm^2R)$, where the runs parameter, R, can be chosen to trade-off between accuracy and speed.

Our heuristic was tested against an exact solution of Equation 4.87 on a small tractable problem. For 1000 iterations, 10 patient belief states were chosen from the pool of initial patient belief states used within the case study simulation. Within each iteration, both the heuristic and the exact solution were queried for which patients to screen if 5 screenings were available in each period. This test was executed at 5 separate levels of heuristic strength $R = 100, 200, 300, 400, 500$. The results are shown in Table 4.6, where * indicates a comparison to the exact solution value or exact solution run time.

Heuristic Used	100 runs	200 runs	300 runs	400 runs	500 runs
Matched Exact Solution	58%	83%	93%	96%	99%
Average Solution Value*	99.91%	99.98%	>99.99%	>99.99%	>99.99%
Average Run Time*	5.79%	11.59%	17.38%	23.11%	29.89%

Table 4.6: Performance of the combinatorial optimization heuristic at various strengths, compared to an exact solution.

The two methods were compared on three performance metrics (1) the percentage of the 1000 iterations where the heuristic solution matched the exact solution, (2) the average value of the heuristic solution, compared to the value of the exact solution, and (3) the average run time of the heuristic, compared to the run time of the exact solution. The heuristic is extremely efficient at finding good solutions to the combinatorial optimization problem. Even at the most expensive strength of 500 runs, the heuristic requires less than a third of the time as the exact solution, while finding the same solution in 99% of the scenarios.

4.4.5 Case Study Results

The simulation of current practice and our policy were tested on 6 different panel sizes of $50, 100, 150, 200, 250, 300$ to imitate clinics of various sizes, each for 100 iterations. Each time the simulation queried our policy for a decision, the heuristic was employed at a strength of 500 runs. All simulations were run in MATLAB v.8.5(R2015a). For each initial patient panel, we tested whether the condition $C^{*}(r)$ in Theorem IV.3 which dictated our policy was violated. In 100% of the tested scenarios, the sufficient condition was satisfied. These initial tests indicate that our problem does warrant the usage of Theorem IV.3 in our case study parameterized to liver cancer.

Figure 4.12 shows that our policy is able to detect, on average, 22.2% more early cancers per year. This increase in performance comes at no additional cost to the decision maker, as our policy used the same number of screenings every 30 days as current practice, yet only differed in the choice of patients screened.

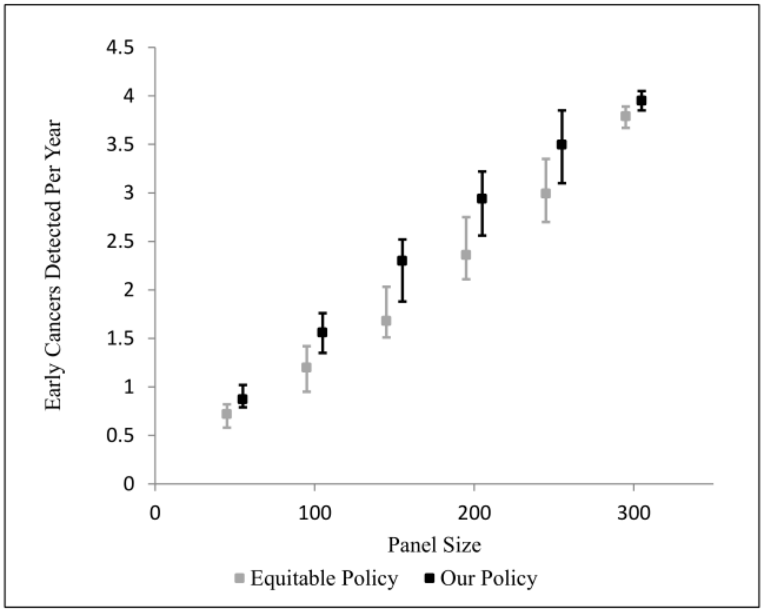


Figure 4.12: Detection rates of our policy vs current practice, with 95th percentiles across all iterations.

While early stage cancer detection is the primary objective of any screening policy, we can measure the auxiliary benefits of a smarter screening policy through other metrics. Figure 4.13 displays the percentage of screenings spent on patients who eventually developed cancer. On average, our policy spends 30.5% more of its resources on cancer developing patients than current practice. This benefit is two-fold: more resources are spent on those patients who need it, and unnecessary time, costs, and psychological distress are saved for patients who ultimately do not develop the disease.

Additionally, we compare the performance of the two screening policies on a more focused scope. Figure 4.14 shows that, if we only consider patients who developed cancer, 19.6% more of these patients would be detected while still in early stage. This proves that our policy translates into better health outcomes for these critical patients.

While the parameterizations of our models was done in the most intuitive way

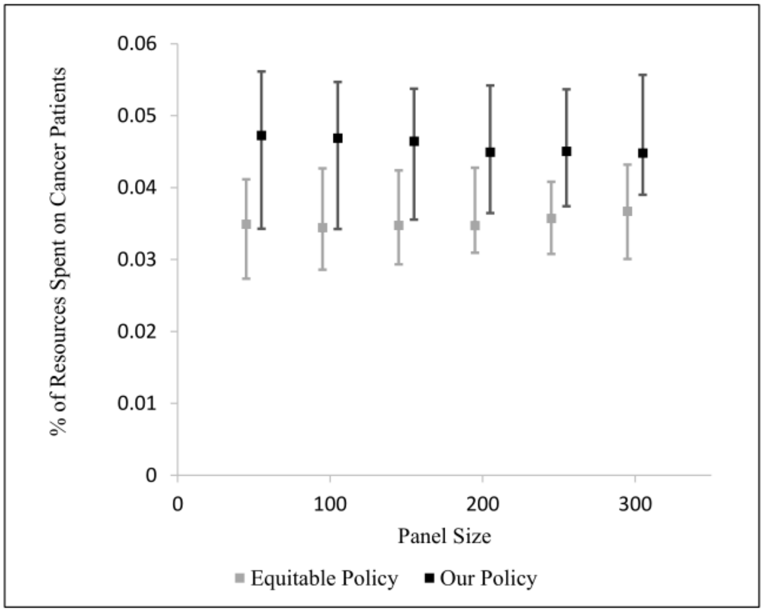


Figure 4.13: Resource usage of our policy vs current practice, with 95th percentiles across all iterations.

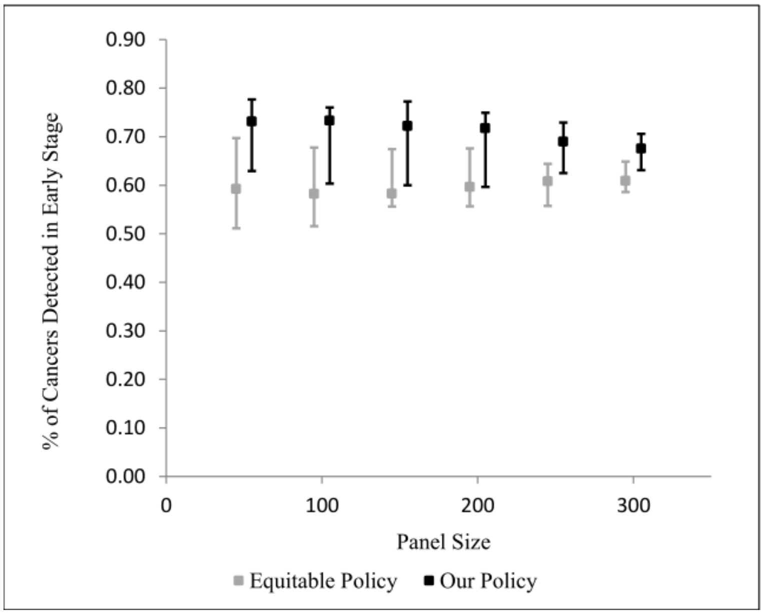


Figure 4.14: Health outcomes of our policy vs current practice, with 95th percentiles across all iterations.

possible, and in a way which is consistent with concurrent medical screening optimization literature, it is still fair to question whether the estimates of beliefs about patient risk are accurate. This concern is alleviated by the performance of our policy which depends upon these beliefs for decision-making. Our policies' numerical success serves as high level validation of our underlying disease and belief models.

In summary, Figure 4.12 demonstrates that our approach outperforms current practice in the primary clinical outcome of interest. Figure 4.13 provides deeper evidence that our policy is highly effective at identifying patients of higher risk, and allocating screenings to them accordingly. Lastly, Figure 4.14 demonstrates that our policy not only screens the correct patients, but at the critical times needed to improve health outcomes.

4.5 Discussion

In this paper, we considered a novel framework to screen a population for liver cancer while simultaneously considering disease progression and resource availability. We modeled the problem as a restless bandit problem, and derived an optimal policy and investigated its structural properties. We addressed its computational complexity by reducing the number of sufficient conditions, as well as simplifying its structure, so that it could be implementable within a clinical setting. Lastly, we parameterized this model with clinical trial data, and demonstrated its numerical efficacy in a computer simulation built upon historical patient data.

The policy derived in this paper has an easily interpreted structure which provides several intuitive suggestions for screening. Namely, it advocates a trade-off between screening to learn patient risk and screening to detect cancer. Our policy advises against myopic behavior which may seem like the intuitive choice to many clinicians. Our research also proves that screening clinics can make decisions using a truncated planning horizon without any loss of optimality, an observation which simplifies planning decisions for clinicians. Lastly, the policy remains the same, regardless of panel size and patient health.

We acknowledge that there are several complications which are not captured in our model. Patient adherence to screening requests are assumed to be perfect. The impact of imperfect adherence could easily be estimated in future modifications to the case study simulation by adding a fixed probability that a chosen patient does not receive a screening. Furthermore, the time delays involved with requesting a screening, performing the screening, and receiving the analyzed results of screening tests are all assumed to be negligible. Future formulations of our model could start by adding a fixed delay between the stage when a patient is chosen for screening and the stage when their belief state is updated. This modification would enable us to study the impact of time delays upon the structure of the optimal policy. Lastly, early-stage and late-stage cancer are assumed to be observed with perfect accuracy. (Note that while the ultrasound is not reliable for tumor diagnosis, it is standard practice to follow-up any suspicious features in the ultrasound with a CT scan or MRI, both of which have very high sensitivity and specificity.) Imperfect diagnoses could be studied in future work both analytically and numerically in a similar vein to *Ayer et al.* (2012). While these extensions are possible, we choose to leave direct incorporation of these features to future work because we believe our work provides a starting point for understanding how to allocate a limited capacity amongst patients whose risk is simultaneously evolving.

To study this problem from an economic perspective, one could incorporate the cost of screenings, and QALYs (quality-adjusted life years) gained by an early cancer detection. To add in an ethical perspective, one could also place a constraint on the policy's maximum permitted time between any screenings for each patient to guarantee a level of treatment equity. In both cases, our work provides bounds on the maximum number of early stage cancers detected per year, without these additional considerations. Furthermore, while our model was tested in the area of liver cancer, it could be adapted to analogous problems that require detecting a critical state amongst simultaneously evolving processes with a limited number of resources. For example, this model could be parameterized to other cancers and machine maintenance problems.

The use of bandit problems to healthcare will continue to become more relevant as expenditure continues to rise and existing infrastructure is further stretched thin. We hope that this modeling framework, as well as its successful validation with real-world data, opens possibilities for this methodology to be explored in new applications to further enhance healthcare delivery.

CHAPTER V

Conclusion and Future Work

This research applies operations research methods to address problems of how to simultaneously manage a population of chronically ill patients. We began by characterizing disease progression so that we could exploit this knowledge for better decisions. We turned this knowledge over to reinforcement learning based policies, and found that this knowledge can be learned over time to the advantage of a decision maker. Lastly, we modeled this problem as a restless bandit, and showed that this learning could be optimized.

We now turn our attention to issues which would arise from implementing this policy. We first discuss three areas where we could see these centrally planned policies being used in the future, as well as for other chronic diseases. We discuss how we could improve the strength of our base risk models. We then provide two alternative policies which address potential extensions of our work. We then study how changes in fixed problem parameters (such as patient panel size and biomarker accuracy) affect overall performance, and how planners could take advantage of this relationship for capacity planning. Finally, we address recent developments surrounding liver cancer screening and provide some concluding remarks.

5.1 Uses of Centrally Planned Healthcare in the Future

We discuss four sample areas in which we could see a centrally planned healthcare policy being utilized in the near future.

5.1.1 In the United States

The first setting is in the changing landscape of American healthcare. We have already established that American policymakers have long recognized the need to curb overall healthcare expenditure (*Orszag* (2009)), while accounting for the aging and increasingly ill population. Our approach is motivated by these circumstances, and the United States remains the area where we believe it to be most appropriate.

According to the National Center for Biotechnology Information, the United States is the last remaining developed country in the world without government provided universal healthcare for its citizens (*Vladeck* (2003)). Recent political developments, such as the Patient Protection and Affordable Care Act (more commonly known as"Obamacare"), point towards centrally distributed healthcare becoming more and more of a reality in this country. However, we do acknowledge that public approval for these new government provision of healthcare has not yet been widely accepted by the general public (*Conway* (2013)).

The largest obstacle we foresee to implementation of centrally planned healthcare in the United States is the need to change the public mindset on rationing healthcare spending. The general mindset of American healthcare has been to spare no expense for optimal health outcomes of every patient, and to allow each individual's personal preferences and financial resources to determine the level of care they receive (*Singer* (2009)). Our approach does not dictate a patient's treatment by their preferences or financial resources, but only by their observed health status, and therefore the general public may not be receptive to this lack of control over individual treatment. However, as healthcare spending continues to soar and burden individuals, we believe our models could provide financial relief to the population and government. Further evidence for the potential success of centrally planned healthcare is demonstrated by the far greater satisfaction with healthcare in similarly developed countries which provide national health. 44% of Americans are "very dissatisfied" with the availability of affordable healthcare, as opposed to 17% and 25% in Canada and Great Britain, respectively (*Blizzard* (2003)).

5.1.2 In Closed Healthcare Systems

Thinking beyond a centralized system of healthcare for the general public, one may consider closed systems that already exist in the United States. An example of such a system is the Department of Veteran Affairs (VA). The VA is an especially attractive place for our models to be implemented for two reasons: (1) Empirical studies have shown the VA to have a long history of maintaining accurate and robust clinical data on their patients (*Kashner* (1998)), a requirement for our policy to make proper decisions. (2) With over 150 medical centers nationwide, the VA is the national's single largest integrated health network, and its wide coverage has lead it to be described as "a national resource for clinical research" (*Fisher and Welch* (1995)). Historically, the VA has been a pioneer for many new medical technologies and health delivery systems. We believe an implementation of our screening policies would fit into the VA's long tradition translating innovative research into standard medical practice.

The VA is not, however, without its own unique potential challenges for implementation of our policies. Recent studies have brought forward issues of patient equity in the VA. These disparities have been recognized across gender (*Hoff and Rosenheck* (1998)), race (*Trivedi et al.* (2011)), and socioeconomic status (*Trivedi and Grebla* (2011)). The VA has even established a Center for Health Equity to address the importance of these issues. Patient equity would need to be addressed before implementation, and we discuss potential ways to modify our policies later in this chapter.

5.1.3 In Humanitarian Healthcare

Moving outside of the United States, our work may prove to be of use in humanitarian efforts in developing countries. This is the setting with the most disproportionately exaggerated difference between population size and available resources, and thus our approach might be very appropriate. More often than not, the sole objective in that setting is to improve health outcomes at a population level (*Blanchet and Roberts* (2013)), and thus this setting is very much aligned with the presumed objectives of our model.

In this setting, the healthcare provider distributes resources and services at presumably no cost to the patients. Some policymakers may object to the usage of our methods in humanitarian efforts in developing countries because of the lower relevance of chronic disease in that setting. Typically, screening for infectious diseases could represent a more pressing matter than screening for chronic diseases. However, death from chronic conditions in developing countries continues to outweigh deaths from infection and injury combined (*Nugent* (2008)). Therefore we argue that improved management of chronic diseases needs to be considered.

Lastly, consider the following potential advantage of our findings. Many highly burdened screenings clinics, such as those in humanitarian efforts, are limited by ultrasound availability, not AFP blood test availability (*Kurjak and Breyer* (1986)). This is for two reasons: firstly, drawing a patient's blood and analyzing it for AFP costs far less than an ultrasound imaging. Secondly, AFP blood tests are quick in administration and evaluation, whereas an ultrasound image requires an appointment with a trained ultrasound technician, as well as analysis by a medical doctor for specific image features. Therefore we could imagine highly burdened screening clinics being able to simultaneously administer two modifications of our policy: AFP blood draws for a large proportion of our population, and ultrasounds for a select few. By essentially separating the problems of exploration and exploitation into separated problems, a highly burdened humanitarian screening clinic may maximize the efficiency of their available resources.

A second point to consider in the implementation of our work in developing countries is the lack of stable and sustainable healthcare infrastructure (*Perry* (2007)). Because our model leads to a policy which is administered over time, several aspects of our problem are assumed to remain constant throughout the planning horizon, such as the availability of medical professionals, durable equipment, and disposable medical resources. The modelling of changing environments and resources could represent a meaningful extension of the models presented in this thesis to be studied in the future.

5.1.4 In Other Chronic Diseases

We would like to add that our models could be applied to other chronic disease besides HCC. Although our case studies were parameterized for the setting of liver cancer screening, the models themselves fit a large number of chronic diseases. For example, the problems of screening for prostate cancer are very similar to that of screening for liver cancer. Both problems seek to screen patients during a critical point in their disease progression, and both problems have a blood biomarker which serves as a signal of underlying patient risk (*Catalona et al.* (1993)).

5.2 Re-Parameterizations of Models

The success of the policies developed in Chapters III and IV depend on the strength of the base risk model developed in Chapter II. Our policies performed well numerically, despite the fact that the AFP is considered to be relatively weak signal of risk compared to other biomarkers of chronic disease. Therefore it would be natural to investigate how much the policies performance could improve, given a stronger base model of patient risk.

A stronger base model of risk essentially amounts to a stronger fitting logistic regression model between individual factors and HCC development. This could happen in the future in one of two ways: (1) If more robust surveillance data become available, the prediction of HCC development could be more accurately associated to individual factors. We could seek out further datasets from hospital charts and other clinical trials to supplement the HALT-C dataset to accomplish this. (2) Secondly, medical literature continues to discover new biomarkers for chronic diseases. We could adapt our models to reflect any future findings in the medical literature to improve our underlying risk models. As our understanding of what factors predict tumor development improves, we postulate that our policies based upon these models would improve as well.

Having discussed potential areas of implementation for our models, the potential for usage with other diseases, and the prospect of re-parameterizing our model with stronger base models, we now proceed to discuss three extensions of our work: (1) A modified policy which limits the maximum time a patient can go without being seen, (2) A modified policy which balances the competing objective of patient equity, and (3) A capacity planning tool for new screening clinics.

5.3 Extensions of Our Work: The Maximum Delay Policy

An unfortunate outcome of maximizing population wide health metrics is that any individual patient may be neglected by the system for long periods of time. The possibility of this event, however unlikely, could cause our policy to draw criticism, and so we propose a method to curb these detriments.

We could study a modification of our proposed policy, called the Maximum Delay

Policy. For a decision maker with n available screenings, the choice of who to screen could be made in two stages: First, any patient who has not been screened within the last p periods must be included in the current decision epoch's screening cohort (or the next possible period while maintaining feasibility). This ensures a maximum gap between any two screenings for every single patient, thus providing a minimal level of guaranteed treatment across the population. Secondly, any remaining screenings would be distributed according to our existing method. Further modifications may be required to ensure feasibility when the fixed capacity is insufficient.

Recall Equation 4.3 of Chapter IV, which summarizes the beliefs of all patients $i = 1, .., n$, for any state $j = 1, .., m$, at any time $t = 1, .., T$.

$$X_{i,t+1} = \begin{cases} \overline{X_{it}\Omega_k}\tau_s & \text{if } a(t) = i \text{ and } o(t) = 1, ..., p \\ \overline{X_{it}\Omega_E}\tau_s & \text{if } a(t) = i \text{ and } o(t) = \bar{E} \\ \overline{X_{it}\Omega_L}\tau_s & \text{if } a(t) = i \text{ and } o(t) = \bar{L} \\ X_{it}\tau_u & \text{if } a(t) \neq i \end{cases} \tag{5.1}$$

We would append the state space of the problem with an additional variable, $Z_{i,t}$ which tells the decision maker how long it has been since the patient was last screened. $Z_{i,t}$ evolves deterministically in the following way: If patient i was screened at time t, this counter is reset to 1. If patent i was not screened at time t, then this counter would increment by 1.

$$Z_{i,t+1} = \begin{cases} Z_{i,t} + 1 & \text{if } a(t) \neq i \\ 1 & \text{if } a(t) = i \end{cases} \tag{5.2}$$

Another key difference from the formulation in Chapter IV is that $a(t)$ was the single patient $i \in \{1, ..., N\}$ that was chosen to be screened at time t. Instead, we now let

$e(t)$ be defined as patients who must be screened to achieve minimal equity.

$$\begin{aligned} e(t) &= \text{ patients with } p \text{ periods since their last screening} \qquad (5.3) \\ &= \{i \in \{1,..,N\} | Z_{i,t} = p\} \qquad (5.4) \end{aligned}$$

$$a(t) \;=\; \text{patients to chosen to be screened with the remaining available capacity} \qquad (5.5)$$

The last new notation we need is to describe all the choices of combinations of patients who can be screened. Let $R(t)$ be defined as patients who do not necessarily have to be screened for equity purposes.

$$R_t = \{i \in \{1,..,N\} | Z_{i,t} < p\} \qquad (5.6)$$

Then the power set of R describes all combinations of patients that the decision maker can choose to screen at time t:

$$C_t = a(t) \in 2^R || (a(t)| = N - |e(t)| \qquad (5.7)$$

With this new notation, we can now formulate the dynamic program for the Minimally Equitable Policy variation. The value of the current state is the sum of the immediate rewards from patients who must be screened $e(t)$ and the patients we choose to screen $a(t)$, plus the value of the next state resulting from these choices.

$$V_t(X_t, Z_t) = \max_{a(t) \in C_t} \left\{ \sum_{i \in e(t)} r_{i,t} + \sum_{i \in a(t)} r_{i,t} + V_{t+1}(X_{t+1}, Z_{t+1} | e(t), a(t)) \right\} \qquad (5.8)$$

For the second portion of studying the maximum delay policy, we could write the most intuitive analog of our policy adapted to this new two-stage decision process, and measure its efficiency numerically. Recall that our policy decides which n patients to screen according to the following equation:

$$\max_{\substack{\{a(1),\ldots,a(s)\}\subset \\ \{1,\ldots,.n\}}} \left\{X_{a(1)}R + X_{a(2)}\tau R + X_{a(3)}\tau^2 R + ... + X_{a(s-1)}\tau^{s-2}R + X_{a(s)}\tau^{s-1}R\right\} \tag{5.9}$$

We could test the numerical performance of this policy in simulation, and analyze the impact of the maximum time gap p upon the performance of the policy. We would reasonably expect that (1) our original problem (which is essentially our new problem with $p = \infty$ to be an upper bound on policy performance for all other policies, and (2) the policy performance to be a non-decreasing function of p. We believe that these two statements would be confirmed numerically, but would be more challenging to prove analytically.

5.4 Extensions of Our Work: The Equitable Policy

We now consider a second extension to our models: equitable treatment of patients. As in the previous section, the maximization of population-wide health metrics pays no regard to the treatment of each individual patient, and as a result, there may be a perceived disparity in medical attention. To alleviate the potential disparities in treatment, we could explicitly account for patient equity in the decision maker's objective.

We propose a second variation of our policy which more directly addresses the issue of patient equity. We would do so by adding a weighted term to the objective function which measures the equity of the patient chosen to be screened. We would create a measure of equity which encourages screening patients who have not been seen in a long time, and penalizes seeing patient who have been seen relatively recently. Consider the following quantity:

$$\max_i Z_{i,t} - Z_{a(t),t} \tag{5.10}$$

This quantity represents the difference between the time between screenings of the patient chosen to be screened $Z_{a(t),t}$ and the patient who has not been screened least recently $\max\limits_i Z_{i,t}$. By adding this term as a weighted penalty to objective function, we could penalize inequitable choices to any degree that we choose.

To add an equity consideration to our decision maker's objective, recall the previous reward function in our dynamic program:

$$(r(t)|a(t) = i, o(t) = p) = \begin{cases} 0, & \text{if } o(t) = 1, ..., p \\ 1, & \text{if } o(t) = \bar{E} \\ 0, & \text{if } o(t) = \bar{L} \end{cases} \tag{5.11}$$

The new reward function would add on our measure of equity with a weighted penalty, where k is the adjustable weight:

$$(r(t)|a(t) = i, o(t) = p) = \begin{cases} 0 + k \cdot \left(\max\limits_i Z_{i,t} - Z_{a(t),t}\right), & \text{if } o(t) = 1, ..., p \\ 1 + k \cdot \left(\max\limits_i Z_{i,t} - Z_{a(t),t}\right), & \text{if } o(t) = \bar{E} \\ 0 + k \cdot \left(\max\limits_i Z_{i,t} - Z_{a(t),t}\right), & \text{if } o(t) = \bar{L} \end{cases} \tag{5.12}$$

Let us consider the impact of this change upon the optimality equation. We begin with the same generic optimality equation conditioned upon the action and observations:

$$V_t(X_t) = \max_{a(t)=1,...,n} \Bigg\{ \sum_{k=1,...,p,\bar{E},\bar{L}} Pr(o(t) = k) \\ \Big[(r_t|a(t) = i, o(t) = k) + V_{t+1}\left(X_{t+1}|a(t) = i, o(t) = k\right)\Big] \Bigg\} \tag{5.13}$$

If we follow the same basic steps of substitution and simplification followed in Chapter

IV, we would arrive at the following optimality equation for any generic time t:

$$V_t(X_t) = \max_{a(t)=1,...,n} \left\{ \begin{array}{c} \displaystyle\sum_{k=1}^{p} X_{a(t)}\Omega_k\vec{1} \cdot V_{t+1}\left(X_1\tau_u, ..., \overline{X_{a(t)}\Omega_k}\tau_s, ..., X_n\tau_u\right) \\ +X_{a(t)}\rho \\ +k \cdot \left(\max_i Z_{i,t} - Z_{a(t),t}\right) \\ +(X_{a(t),E} + X_{a(t),L}) \cdot V_{t+1}\left(X_1\tau_u, ..., \overline{X_{a(t)}\Omega_E}\tau_s, ..., X_n\tau_u\right) \end{array} \right\} \tag{5.14}$$

All of our previous results should set a numerical upper bound on performance for any equitable policy (since it corresponds to the $k = 0$ case).

5.5 Capacity Planning for a New Clinic

One final way to expand upon the ideas of Chapter IV is to see our models as a component of a larger hospital planning problem. When building a new screening clinic, it is difficult to balance the costs and benefits of the various clinic sizes. Increasing the size of a clinic translates into hiring more staff and doctors, purchasing more durable medical equipment, and planning for more resource usage. Our model could provide a way for a new screening clinic to decide these operational parameters.

Recall that the panel size N, and the number of screenings available per period s, were fixed parameters in Chapter IV. In that scenario, the total number of early cancers detected could be written in terms of the value function as follows: $V_0(X_0)$, given some initial belief state X_0. If we expanded this problem to make panel size N and screening capacity s to be decision variables instead of parameters, we could write the total expected number of early-stage cancer detections to be

$$V_0(X_0|N, s) \tag{5.15}$$

To complete our analysis, we would require three new parameters:

1. B the total budget available for operating this screening clinic
2. c the marginal cost of single screening
3. q the value of an early stage cancer detection

Now consider the problem of maximizing the total expected number of early stage detections gained by a clinic of size N which screens s patients per month for a time horizon of T months, given an operating budget of B. This could be written as follows:

$$\max q \cdot V_0(X_0|N, s) \tag{5.16}$$

subject to: $s \cdot c \cdot N \cdot T \leq B \; s, N \in \mathbb{Z}^+$.

Intutitively, we are seeking to maximize the value gained by a screening clinic using our policy over the entire time horizon $q \cdot V_0(X_0|N, s)$ without the total cost of implementing that screening policy going over budget $s \cdot c \cdot N \cdot T \leq B$.

This is a new optimization problem in the variables N and s. Fortunately, the decision variables s and N are limited to the positive integers, and the constraint $s \cdot c \cdot N \cdot T$ is linear in the two variables, which simplifies this problem. The difficulty of the analysis would lie in the objective function, $V_0(X_0|N, s)$, which itself an entire dynamic program.

A natural first approach might be to compute $V_0(X_0|N, s)$ for all possible values of N and s satisfying $s \cdot c \cdot N \cdot T \leq B$, with brute force. However, this approach is wasteful, and with some insight, can be reduced. It is reasonable to assume that $V_0(X_0|N, s)$ is non-decreasing in s. That is, for any two clinics with the same panel size N, but with differing screening capacities $s_1 < s_2$, we should expect that $V_0(X_0|N, s_1) \leq V_0(X_0|N, s_2)$. This conclusion is only through an intuitive interpreation of the problem, and would require formal proof. However, if proven, we could avoid computing $V_0(X_0|N, s)$ for many strictly dominated portions of the feasible region.

5.6 Recent Developments

In the past year, the FDA has approved Zepatier, which has been shown to cure Hepatitis C in over 95% of patients (*Food and Administration* (2016)). This does not diminish the value of this research for two reasons: (1) The cost of a full drug regimen required to completely eradicate the Hepatitis C virus costs approximately 100,000 US dollars (*Wapner* (2014)). The drug Zepatier, while fast and effective, is not yet a financially viable option for most of the general US population. (2) The drug has only been shown to be effective on certain genotypes of the disease (*Zeuzem et al.* (2015)), therefore a large portion of people with Hepatitis C remain incurable. While future pharmaceutical developments seem promising for Hepatitis C, we believe more efficient screening still holds value for many years to come.

5.7 Concluding Remarks

In this thesis, we have given a new operations approach to the surveillance of populations with chronic diseases. We have contributed to the academic community by modeling screening problems in more realistic and complex settings. We have also contributed to the medical community by providing implementable methods of screening which we demonstrated thoroughly to be superior to current practice. We believe this is a rich area of research, holding both interest to the academic community and benefits for society in the future.

BIBLIOGRAPHY

BIBLIOGRAPHY

Ahuja, V., and J. R. Birge (2016), Response-adaptive designs for clinical trials: Simultaneous learning from multiple patients, *European Journal of Operational Research*, *248*(2), 619 – 633.

Alagoz, O., T. Ayer, and F. S. Erenay (2011), Operations research models for cancer screening, *Wiley Encyclopedia of Operations Research and Management Science.*

Altekruse, S. F., K. A. McGlynn, and M. E. Reichman (2009), Hepatocellular carcinoma incidence, mortality, and survival trends in the United States from 1975 to 2005, *Journal of Clinical Oncology*, *27*(9), 1485–1491.

Arif-Tiwari, H., B. Kalb, S. Chundru, P. Sharma, J. Costello, R. W. Guessner, and D. R. Martin (2014), MRI of hepatocellular carcinoma: an update of current practices, *Diagnostic Interventional Radiology*, *20*, 209–221.

Ayer, T., O. Alagoz, and N. Stout (2009), A mathematical model to optimize breast cancer screening policy, in *Proceedings of the 31st Annual Meeting of the Society for Medical Decision Making Abstract.*

Ayer, T., O. Alagoz, and N. K. Stout (2012), OR Forum-A POMDP approach to personalize mammography screening decisions, *Operations Research*, *60*(5), 1019–1034.

Ayvaci, M. U., O. Alagoz, and E. S. Burnside (2012), The effect of budgetary restrictions on breast cancer diagnostic decisions, *Manufacturing & Service Operations Management*, *14*(4), 600–617.

Bechhofer, R. E. (1954), A single-sample multiple decision procedure for ranking means of normal populations with known variances, *The Annals of Mathematical Statistics*, pp. 16–39.

Blanchet, K., and B. Roberts (2013), An evidence review of research on health interventions in humanitarian crises, *London: London School of Hygiene & Tropical Medicine.*

Blizzard, R. (2003), Healthcare System Ratings: U.S., Great Britain, Canada, *Gallup.*

Bodenheimer, T., E. Chen, and H. D. Bennett (2009), Confronting the growing burden of chronic disease: can the US health care workforce do the job?, *Health Affairs*, *28*(1), 64–74.

Bruix, J., and M. Sherman (2005), Management of hepatocellular carcinoma, *Hepatology*, *42*(5), 1208–1236.

Bruix, J., et al. (2001), Clinical management of hepatocellular carcinoma. conclusions of the barcelona-2000 easl conference, *Journal of Hepatology*, *35*(3), 421–430.

Catalona, W. J., D. S. Smith, T. L. Ratliff, and J. W. Basler (1993), Detection of organ-confined prostate cancer is increased through prostate-specific antigenbased screening, *Jama*, *270*(8), 948–954.

Centers for Disease Control (2010), Hepatocellular Carcinoma — United States, 2001–2006, *Morbidity and Mortality Weekly Report*, *59*, 517–520.

Centers for Disease Control (2016), Deaths and mortality, *FastStats*.

Chaiteerakij, R., B. D. Addissie, and L. R. Roberts (2013), Update on biomarkers of hepatocellular carcinoma, *Clinical Gastroenterology and Hepatology*.

Chhatwal, J., O. Alagoz, and E. S. Burnside (2010), Optimal breast biopsy decision-making based on mammographic features and demographic factors, *Operations Research*, *58*(6), 1577–1591.

Clemen, R. T., and C. J. Lacke (2001), Analysis of colorectal cancer screening regimens, *Health Care Management Science*, *4*(4), 257–267.

Colli, A., M. Fraquelli, G. Casazza, S. Massironi, A. Colucci, D. Conte, and P. Duca (2006), Accuracy of ultrasonography, spiral ct, magnetic resonance, and alpha-fetoprotein in diagnosing hepatocellular carcinoma: a systematic review, *The American Journal of Gastroenterology*, *101*(3), 513–523.

Conway, B. A. (2013), Addressing the medical malady: second-level agenda setting and public approval of obamacare, *International Journal of Public Opinion Research*, *25*(4), 535–546.

Curley, S. A., C. C. Barnett Jr, E. K. Abdalla, K. K. Tanabe, and D. M. Savarese (2015), Staging and prognostic factors in hepatocellular carcinoma.

Davies, R., D. Crabbe, P. Roderick, J. R. Goddard, J. Raftery, and P. Patel (2002), A simulation to evaluate screening for helicobacter pylori infection in the prevention of peptic ulcers and gastric cancers, *Health Care Management Science*, *5*(4), 249–258.

Davis, G. L., M. J. Alter, H. El-Serag, T. Poynard, and L. W. Jennings (2010), Aging of hepatitis c virus (hcv)-infected persons in the united states: a multiple cohort model of hcv prevalence and disease progression, *Gastroenterology*, *138*(2), 513–521.

Deo, S., and M. Sohoni (2015), Optimal decentralization of early infant diagnosis of HIV in resource-limited settings, *Manufacturing & Service Operations Management*, *17*(2), 191–207.

Deo, S., S. Iravani, T. Jiang, K. Smilowitz, and S. Samuelson (2013), Improving health outcomes through better capacity allocation in a community-based chronic care model, *Operations Research*, *61*(6), 1277–1294.

Dhamodharan, A., and R. Proano (2012), Determining the optimal vaccine vial size in developing countries: a monte carlo simulation approach, *Health Care Management Science*, *15*(3), 188–196.

Dudewicz, E. J., and S. R. Dalal (1975), Allocation of observations in ranking and selection with unequal variances, *Sankhyā: The Indian Journal of Statistics, Series B*, pp. 28–78.

El-Serag, H. B., J. A. Davila, N. J. Petersen, and K. A. McGlynn (2003), The continuing increase in the incidence of hepatocellular carcinoma in the united states: an update, *Annals of Internal Medicine*, *139*(10), 817–823.

Erenay, F. S., O. Alagoz, and A. Said (2014), Optimizing colonoscopy screening for colorectal cancer prevention and surveillance, *Manufacturing & Service Operations Management*, *16*(3), 381–400.

Fisher, E. S., and H. G. Welch (1995), The future of the Department of Veterans Affairs health care system, *JAMA*, *273*(8), 651–655.

Food, and D. Administration (2016), Fda approves zepatier for treatment of chronic hepatitis c genotypes 1 and 4, *FDA News Release*.

Frazier, A. L., G. A. Colditz, C. S. Fuchs, and K. M. Kuntz (2000), Cost-effectiveness of screening for colorectal cancer in the general population, *JAMA: the Journal of the American Medical Association*, *284*(15), 1954–1961.

Gittins, J. C. (1979), Bandit processes and dynamic allocation indices, *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 148–177.

Globocan (2012), Liver cancer: Estimated incidence, mortality and prevalence worldwide in 2012, *World Health Organization*.

Goldie, S. J., J. J. Kim, and T. C. Wright (2004), Cost-effectiveness of human papillomavirus dna testing for cervical cancer screening in women aged 30 years or more, *Obstetrics & Gynecology*, *103*(4), 619–631.

Güneş, E. D., S. E. Chick, and O. Z. Akşin (2004), Breast cancer screening services: trade-offs in quality, capacity, outreach, and centralization, *Health Care Management Science*, *7*(4), 291–303.

Hanin, L., A. Tsodikov, and A. Y. Yakovlev (2001), Optimal schedules of cancer surveillance and tumor size at detection, *Mathematical and Computer Modelling*, *33*(12), 1419–1430.

Harper, P., and S. Jones (2005), Mathematical models for the early detection and treatment of colorectal cancer, *Health Care Management Science*, *8*(2), 101–109.

Helm, J. E., M. S. Lavieri, M. P. Van Oyen, J. D. Stein, and D. C. Musch (2015), Dynamic forecasting and control algorithms of glaucoma progression for clinician decision support, *Operations Research*, *63*(5), 979–999.

Hoff, R. A., and R. A. Rosenheck (1998), Female veterans' use of Department of Veterans Affairs health care services, *Medical Care*, *36*(7), 1114–1119.

Imaeda, T., and H. Doi (1992), A retrospective study on the patterns of sequential fluctuation of serum alpha-fetoprotein level during progression from liver cirrhosis to hepatocellular carcinoma, *Journal of Gastroenterology and Hepatology*, *7*(2), 132–135.

Johnson, P. J. (2001), The role of serum alpha-fetoprotein estimation in the diagnosis and management of hepatocellular carcinoma, *Clinics in liver disease*, *5*(1), 145–159.

Kaelbling, L. P. (1993), *Learning in Embedded Systems*, MIT Press.

Kashner, T. M. (1998), Agreement between administrative files and written medical records: a case of the department of veterans affairs, *Medical care*, *36*(9), 1324–1336.

Keehan, S. P., G. A. Cuckler, A. M. Sisko, A. J. Madison, S. D. Smith, D. A. Stone, J. A. Poisal, C. J. Wolfe, and J. M. Lizonitz (2015), National health expenditure projections, 2014–24: spending growth faster than recent trends, *Health Affairs*, *34*(8), 1407–1417.

Khademi, A., D. R. Saure, A. J. Schaefer, R. S. Braithwaite, and M. S. Roberts (2015), The price of nonabandonment: HIV in resource-limited settings, *Manufacturing & Service Operations Management*, *17*(4), 554–570.

Knudsen, A. B., P. M. McMahon, and G. S. Gazelle (2007), Use of modeling to evaluate the cost-effectiveness of cancer screening programs, *Journal of Clinical Oncology*, *25*(2), 203–208.

Krishnamurthy, V., and R. J. Evans (2001), Hidden markov model multiarm bandits: a methodology for beam scheduling in multitarget tracking, *IEEE Transactions on Signal Processing*, *49*(12), 2893–2908.

Kulasingam, S. L., S. Benard, R. V. Barnabas, N. Largeron, and E. R. Myers (2008), Cost effectiveness and resource, *Cost Effectiveness and Resource Allocation*, *6*, 4.

Kurjak, A., and B. Breyer (1986), The use of ultrasound in developing countries, *Ultrasound in medicine & biology*, *12*(8), 611–621.

Lee, E., S. Edward, A. G. Singal, M. S. Lavieri, and M. Volk (2012a), Improving screening for hepatocellular carcinoma by incorporating data on levels of α-fetoprotein over time, *Clinical Gastroenterology and Hepatology*.

Lee, E., M. Lavieri, and M. Volk (2012b), Evaluating hypothetical screening policies on historical patient data., *Proceedings of Data Mining and Health Informatics Workshop*.

Lee, E., M. Lavieri, and M. Volk (2016), Optimal screening for hepatocellular carcinoma under limited resources: A restless bandit model, *Submitted for Publication*.

Lee, S., and M. Zelen (2003), Modelling the early detection of breast cancer, *Annals of Oncology*, *14*(8), 1199–1202.

Lee, S. J., and M. Zelen (2008), Mortality modeling of early detection programs, *Biometrics*, *64*(2), 386–395.

Leshno, M., Z. Halpern, and N. Arber (2003), Cost-effectiveness of colorectal cancer screening in the average risk population, *Health Care Management Science*, *6*(3), 165–174.

Loeve, F., R. Boer, G. van Oortmarssen, M. van Ballegooijen, and J. Habbema (1999), The miscan-colon simulation model for the evaluation of colorectal cancer screening, *Computers and Biomedical Research*, *32*(1), 13 – 33.

Lok, A., et al. (2009), Incidence of hepatocellular carcinoma and associated risk factors in hepatitis c-related advanced liver disease, *Gastroenterology*, *136*(1), 138–148.

Luce, R. D. (1959), *Individual Choice Behavior a Theoretical Analysis*, John Wiley and sons.

Maillart, L. M., J. S. Ivy, S. Ransom, and K. Diehl (2008), Assessing dynamic breast cancer screening policies, *Operations Research*, *56*(6), 1411–1427.

Myers, E. R., D. C. McCrory, K. Nanda, L. Bastian, and D. B. Matchar (2000), Mathematical model for the natural history of human papillomavirus infection and cervical carcinogenesis, *American Journal of Epidemiology*, *151*(12), 1158–1171.

Nam, C. Y., V. Chaudhari, S. S. Raman, C. Lassman, M. J. Tong, R. W. Busuttil, and D. S. Lu (2011), CT and MRI improve detection of hepatocellular carcinoma, compared with ultrasound alone, in patients with cirrhosis, *Clinical Gastroenterology and Hepatology*, *9*(2), 161–167.

National Center for Chronic Disease Prevention and Health Promotion (2015), At a glance 2015.

National Center for Health Statistics (2016), Deaths and mortality, *FastStats*.

Negoescu, D. M., K. Bimpikis, M. L. Brandeau, D. A. Iancu, et al. (2014), Dynamic learning of patient response types: An application to treating chronic diseases, *Tech. rep.*

Nugent, R. (2008), Chronic diseases in developing countries, *Annals of the New York Academy of Sciences*, *1136*(1), 70–79.

Okada, S., et al. (1993), Follow-up examination schedule of postoperative hcc patients based on tumor volume doubling time., *Hepato-gastroenterology*, *40*(4), 311.

Orszag, P. (2009), The long-term budget outlook and options for slowing the growth of health care costs, *Tesimony before the Committee on Finance in the United States Senate.*

Pandelis, D. G., and D. Teneketzis (1999), On the optimality of the Gittins index rule for multi-armed bandits with multiple plays, *Mathematical Methods of Operations Research*, *50*(3), 449–461.

Papadimitriou, C. H., and J. N. Tsitsiklis (1999), The complexity of optimal queuing network control, *Mathematics of Operations Research*, *24*(2), 293–305.

Parmigiani, G., S. Skates, and M. Zelen (2002), Modeling and optimization in early detection programs with a single exam, *Biometrics*, *58*(1), 30–36.

Perry, M. (2007), Medical brain drain hindering aids battle; lack of medical infrastructure in developing countries, *Conference on HIV Pathogenesis, Treatment and Prevention*, pp. 22–25.

Pierskalla, W. P., and D. J. Brailer (1994), Applications of operations research in health care delivery, *Handbooks in OR & MS*, *6*, 4–7.

Pierskalla, W. P., and J. A. Voelker (1976), A survey of maintenance models: the control and surveillance of deteriorating systems, *Naval Research Logistics Quarterly*, *23*(3), 353–388.

Piette, J. D., J. B. Sussman, P. N. Pfeiffer, M. J. Silveira, S. Singh, and M. S. Lavieri (2013), Maximizing the value of mobile health monitoring by avoiding redundant patient reports: Prediction of depression-related symptoms and adherence problems in automated health assessment services, *Journal of medical internet research*, *15*(7), e118.

Preston, A. J., and W. Smith (2001), Disease screening designs: sensitivity and screening frequency, in *Proceedings of the Annual Meeting of the American Statistical Association*, pp. 5–9.

Rauner, M. S., W. J. Gutjahr, K. Heidenberger, J. Wagner, and J. Pasia (2010), Dynamic policy modeling for chronic diseases: metaheuristic-based identification of pareto-optimal screening strategies, *Operations Research*, *58*(5), 1269–1286.

Romero, H., N. Dellaert, S. Geer, M. Frunt, M. Jansen-Vullers, and G. Krekels (2013), Admission and capacity planning for the implementation of one-stop-shop in skin cancer treatment using simulation-based optimization, *Health Care Management Science*, *16*(1), 75–86.

Schell, G. J., M. S. Lavieri, J. E. Helm, X. Liu, D. C. Musch, M. P. Van Oyen, and J. D. Stein (2014), Using filtered forecasting techniques to determine personalized monitoring schedules for patients with open-angle glaucoma, *Ophthalmology*, *121*(8), 1539–1546.

Singal, A., M. Volk, A. Waljee, R. Salgia, P. Higgins, M. Rogers, and J. Marrero (2009), Meta-analysis: surveillance with ultrasound for early-stage hepatocellular carcinoma in patients with cirrhosis, *Alimentary pharmacology & therapeutics*, *30*(1), 37–47.

Singer, P. (2009), Why we must ration health care, *The New York Times*.

Stevenson, C. (1995), Statistical models for cancer screening, *Statistical Methods in Medical Research*, *4*(1), 18–32.

Strunk, B. C., P. B. Ginsburg, and M. I. Banker (2006), The effect of population aging on future hospital demand, *Health Affairs*, *25*(3), w141–w149.

Sutton, R. S., and A. G. Barto (1998), *Introduction to reinforcement learning*, MIT Press.

Tomasi Jr, T. B. (1977), Structure and function of alpha-fetoprotein, *Annual review of medicine*, *28*(1), 453–465.

Trivedi, A. N., and R. C. Grebla (2011), Quality and equity of care in the veterans affairs health-care system and in medicare advantage health plans, *Medical care*, *49*(6), 560–568.

Trivedi, A. N., R. C. Grebla, S. M. Wright, and D. L. Washington (2011), Despite improved quality of care in the Veterans Affairs health system, racial disparity persists for important clinical outcomes, *Health Affairs*, *30*(4), 707–715.

Tsodikov, A., A. Szabo, and J. Wegelin (2006), A population model of prostate cancer incidence, *Statistics in Medicine*, *25*(16), 2846–2866.

Underwood, D. J., J. Zhang, B. T. Denton, N. D. Shah, and B. A. Inman (2012), Simulation optimization of PSA-threshold based prostate cancer screening policies, *Health Care Management Science*, *15*(4), 293–309.

Urban, N., C. Drescher, R. Etzioni, and C. Colby (1997), Use of a stochastic simulation model to identify an efficient protocol for ovarian cancer screening, *Controlled Clinical Trials*, *18*(3), 251 – 270.

Vladeck, B. (2003), Universal health insurance in the United States: Reflections on the past, the present, and the future, *American Journal of Public Health*, *93*(1), 16–19.

Wapner, J. (2014), We now have the cure for hepatitis c, but can we afford it?, *Scientific American.*

Watkins, C. J. C. H. (1989), Learning from delayed rewards., Ph.D. thesis, University of Cambridge.

Whittle, P. (1988), Restless bandits: activity allocation in a changing world, *Journal of Applied Probability*, pp. 287–298.

Wilkins, T., J. K. Malcolm, D. Raina, and R. R. Schade (2010), Hepatitis C: diagnosis and treatment, *Am Fam Physician*, *81*(11), 1351–7.

Wong, G. L., H. L. Chan, Y.-K. Tse, H.-Y. Chan, C.-H. Tse, A. O. Lo, and V. W. Wong (2014), On-treatment alpha-fetoprotein is a specific tumor marker for hepatocellular carcinoma in patients with chronic hepatitis b receiving entecavir, *Hepatology*, *59*(3), 986–995.

Yaesoubi, R., and S. D. Roberts (2008), How much is a health insurer willing to pay for colorectal cancer screening tests?, in *Simulation Conference, 2008. WSC 2008. Winter*, pp. 1624–1631, IEEE.

Zeuzem, S., et al. (2015), Grazoprevir–elbasvir combination therapy for treatment-naive cirrhotic and noncirrhotic patients with chronic hepatitis c virus genotype 1, 4, or 6 infection: a randomized trial, *Annals of internal medicine*, *163*(1), 1–13.

Zhang, J., B. T. Denton, H. Balasubramanian, N. D. Shah, and B. A. Inman (2012), Optimization of prostate biopsy referral decisions, *Manufacturing & Service Operations Management*, *14*(4), 529–547.

Zhang, Y., and M. L. Puterman (2013), Developing an adaptive policy for long-term care capacity planning, *Health Care Management Science*, pp. 1–9.