ESTIMATING THE PRICE OF PRIVACY IN LIVER TRANSPLANTATION

by

Burhaneddin Sandıkçı

B.S., Marmara University, 1999

M.S. I.E., Bilkent University, 2001

M.S. O.R., University of North Carolina at Chapel Hill, 2003

Submitted to the Graduate Faculty of

the Swanson School of Engineering in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2008

UNIVERSITY OF PITTSBURGH

SWANSON SCHOOL OF ENGINEERING

This dissertation was presented

by

Burhaneddin Sandıkçı

It was defended on

June 23, 2008

and approved by

Andrew J. Schaefer, Associate Professor, Department of Industrial Engineering

Lisa M. Maillart, Assistant Professor, Department of Industrial Engineering

Mark S. Roberts, Professor, School of Medicine

Alan A. Scheller-Wolf, Associate Professor, Tepper School of Business, Carnegie-Mellon University

Oguzhan Alagoz, Assistant Professor, Department of Industrial and Systems Engineering, University of Wisconsin-Madison

Dissertation Co-Director: Andrew J. Schaefer, Associate Professor, Department of Industrial Engineering

Dissertation Co-Director: Lisa M. Maillart, Assistant Professor, Department of Industrial Engineering

ESTIMATING THE PRICE OF PRIVACY IN LIVER TRANSPLANTATION

Burhaneddin Sandıkçı, PhD

University of Pittsburgh, 2008

In the United States, patients with end-stage liver disease must join a waiting list to be eligible for cadaveric liver transplantation. However, the details of the composition of this waiting list are only partially available to the patients. Patients currently have the prerogative to reject any offered livers without any penalty. We study the problem of optimally deciding which offers to accept and which to reject. This decision is significantly affected by the patient's health status and progression as well as the composition of the waiting list, as it determines the chances a patient receives offers. We evaluate the value of obtaining the waiting list information through explicitly incorporating this information into the decision making process faced by these patients. We define the concept of the patient's price of privacy, namely the number of expected life days lost due to a lack of perfect waiting list information.

We develop Markov decision process models that examine this question. Our first model assumes perfect waiting list information and, when compared to an existing model from the literature, yields upper bounds on the true price of privacy. Our second model relaxes the perfect information assumption and, hence, provides an accurate representation of the partially observable waiting list as in current practice. Comparing the optimal policies associated with these two models provides more accurate estimates for the price of privacy. We derive structural properties of both models, including conditions that guarantee monotone value functions and control-limit policies, and solve both models using clinical data.

We also provide an extensive empirical study to test whether patients are actually making their accept/reject decisions so as to maximize their life expectancy, as this is assumed in our

previous models. For this purpose, we consider patients transplanted with living-donor livers only, as considering other patients implies a model with enormous data requirements, and compare their actual decisions to the decisions suggested by a nonstationary MDP model that extends an existing model from the literature.

Keywords: Markov decision processes, partially observable Markov decision processes, structured optimal policies, value of information, price of privacy, organ transplantation, medical decision making, health care treatment.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGMENTS

to my family

It's been long and not easy! But thank Allah, I am at the stage where I am at now. I thank Allah, because all I did, I did by His help and guidance. He gives the strength and wisdom to understand things right, He gives the patience when I need it the most, He gives the intelligence to comprehend. I thank Allah for all the things I couldn't count.

I came to Pittsburgh on May 30th, 2004, with very high hopes. I have fulfilled my expectations and a lot more. My advisor, Andrew Schaefer, has a prominent role in all of these. I did the best thing I could have ever done for my academic life by choosing to move to Pittsburgh to work with him. I appreciate his encouragement to write papers and to present my work at conferences. I express my sincere gratitude to him for his motivation, trust, encouragement, generous financial support, and for all the other things he has done for me throughout my PhD education. Lacking elegant writing skills, I know I will not be able to properly thank my co-advisor Lisa Maillart for all she has done to help me during the last two years of my PhD education. Her technical and editorial advice, and to-the-point comments were essential to the completion of this dissertation. I thank her for her patience and kindness. Since my first day at Pitt, I knew I was going to work on health care applications. I thank Mark Roberts for accepting me to collaborate with him on various research projects. It is extremely difficult to find a medical doctor who understands the language of mathematics to the level he does. I admire him for being such a great mentor. I also had the opportunity to work and write a paper with Oguzhan Alagoz. Beyond writing this paper, he has been an endless resource throughout my studies. I thank him for sharing his technical skills, ideas, and, most importantly, time for long discussions. I am also grateful to my committee member Alan Wolf for agreeing to read and review this dissertation and for his suggestions.

Most of the faculty and staff at Pitt have helped to make my life at Pitt pleasant. I would especially like to mention Professors Mainak Mazumdar and Matt Bailey, for being great

instructors, and for warmly welcoming and answering my random probability questions; Professor Bopaya Bidanda, for patiently listening to me during my times of confusion, and the wonderful staff of the department, Minerva Pilachowski, Richard Brown, and Jim Segneff, for their continuous technical support.

I have been fortunate to work and live alongside great friends in Pittsburgh. For their academic and moral support, I would like to thank İbrahim and Hanife Aydemir, Sakine Batun, Halil Bayrak, Mustafa Baz, Mehmet Demirci, Zeynep Erkin, Nuri Mehmet Gökhan, Ahmet and Latife Kaya, Nan Kong, Rob Koppenhaver, Murat Kurt, Ümit and Rukiye Ogras, Osman Yalın and Nur Özge Özaltın, Sepehr Nemati Proon, Görkem Saka, Steven Shechter, Andy Trapp, Fikret Korhan Turan, Natasa Vidic, Hakan Yıldız and Server Levent Yılmaz. I am also grateful to my longtime friends Berkant Çamdereli and Abdullah Karaman.

Finally, this work would never had been possible without my wonderful family: my parents Süheyla and Kemal, my grandmother Saadet, my brother Mustafa, and my sisters Saadet, Neslihan, Ayşe, Asiye, and Güler. They have been always there for me when I needed. Their endless love, encouragement, and unconditional support have always been an inspiration and strength to me. I have added another wonderful family to my life by marrying my wife: my parents-in-law Hanife and Yusuf Şevki, and my brothers-in-law Malik, Hasan, and Mehmet, to whom I owe a lot. Your prayers are all accepted. Lastly, the two people I owe more thanks to than all of the previous people combined: my enduring, caring, encouraging, and loving wife Zahide and my shining little daughter Melike Sena.

1.0 INTRODUCTION

Health care is the largest industry in the United States and continues to grow. The share of health care spending in the U.S. gross domestic product has increased from 5% in 1960 to 16% in 2004, and it is projected to increase to 18.7% in 2014 [35]. Furthermore, it is estimated that the population of Americans aged 65 or older will double by the year 2030, forming about 20% of the U.S. population [53]. This aging population will result in increased demand for health services and more economic challenges as the cost of providing health care for older adults is 3-5 times greater than that for someone younger than 65 [53].

Operations Research (OR) techniques have been applied to a wide variety of problems in health care. Early surveys of the OR applications in health care are provided in [54, 134] and a recent survey is given in [23]. Historically, most applications have considered logistical issues in health care including hospital capacity planning (particularly, planning bed requirements) [57, 107], personnel (particularly, nurse) scheduling [78, 80, 97, 161], operating room scheduling [162], staffing emergency departments [158], and locating ambulances [22, 46] and other emergency service facilities [77, 152].

Recently, OR literature has considered public policy questions related to health care as well as optimizing treatment plans for individual patients. Examples of those applications focusing on health care related public policy questions include drug policy evaluation [29, 30, 42, 153], pediatric immunization [129], vaccine selection [175], control of infectious diseases [14, 16, 86, 115], bioterrorism [81], and organ transplantation policy evaluation [40, 41, 69, 111, 116, 146, 177, 180, 181]. Finally, examples of those applications focusing on therapeutic optimization problems include optimally deciding which organs to transplant [1, 4, 6, 5, 39], drug infusion plans [72], treatment of ischemic heart disease [63], cancer treatment planning [84, 85, 96, 110, 118, 133], and evaluation of screening policies for disease detection [21, 94].

This dissertation focuses on liver transplantation. The first successful solid organ transplant was a kidney transplant between living twin brothers that took place in 1954, whereas the first successful liver transplant was performed from a cadaveric donor in 1967 [73, 105]. The advances in medical science and technology currently allows the transplantation of various kinds of solid organs such as liver, kidney, heart, lung, pancreas, and intestine as well as tissues, cells, and fluids such as stem cells, bone marrow, blood transfusion, blood vessels, heart valve, bone, and skin.

End-stage liver disease (ESLD), which includes diseases such as cirrhosis and hepatitis, is the 12th leading cause of death in the United States [98], in large part because transplantation is the only available therapy for ESLD patients. Medical science and technology are still years away from finding alternative therapies for ESLD. Despite this lack of alternative therapies, efforts to increase the number of available organs for transplantation have only modest effect on the widening gap between the demand and supply for such organs [101].

Fortunately, transplantation is found to be a highly effective treatment in improving the length of life as indicated by the patient survival rates in Figure 1.1. The 10-year survival rate from cadaveric donors approaches 60%. Survival rates from living donors are slightly higher than that of the cadaveric donors. Furthermore, it has been demonstrated that transplant recipients display statistically significant improvements in physical functioning, mental health, social functioning, and overall quality of life [43].

Currently, there are two sources for a liver transplant: living donors and cadaveric donors. Originally developed for pediatric patients in 1989 [38, 105], living-donor liver transplantation is currently being offered to adult patients as well. It involves taking advantage of the regenerative capability of healthy liver cells so that the entire lobe or part of a lobe of the healthy liver is removed from the donor and placed in the diseased recipient's body. In most cases, both livers grow back to their regular size and function normally in a few weeks [100].

More than 96% of over 87,000 liver transplants since 1988 are from cadaveric donors [106]. Moreover, liver transplants from living donors have decreased in the U.S. since 2001, when it reached an all time high of 522 [106], mainly due to post-operative complications to the donor [24], which is coupled with the highly publicized death of a living donor from New York. There is an estimated 0.2% mortality rate for adult living donors [114].

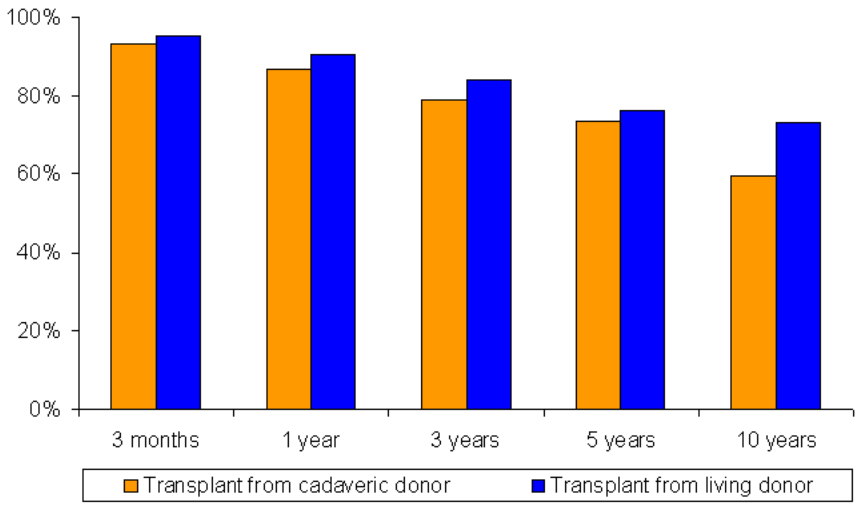


Figure 1.1: Patient survival rates from liver transplantation [66]

As a result, the vast majority of ESLD patients join a waiting list of patients who are eligible for transplantation from cadaveric livers. As seen in Figure 1.2, the disparity between the demand for and supply of cadaveric livers is large, which results in significant waiting times and large number of patient deaths while waiting. Indeed, there are currently about $17,000$ patients on the liver transplant waiting list and, over the last 12 years, this list has grown by approximately 1,000 patients per year [106]. The size of the waiting list seems to have stabilized due to new allocation policies that, in 2001, incorporated the MELD scoring system (see Section 1.1). Unfortunately, demand still far exceeds the supply. Furthermore, about 13.21% of the recovered livers from cadaveric donors are not transplanted [66][1] due to various reasons including excessive cold ischemia time (the duration during which the organ is left without any blood flow, which is the primary determinant of the viability of the organ; the shorter the cold ischemia time the better.) Approximately 1,800 patients die each year while waiting for a liver transplant and about 500 patients get removed from the waiting list

[1]The data and analyses reported in the 2007 Annual Report of the U.S. Organ Procurement and Transplantation Network and the Scientific Registry of Transplant Recipients have been supplied by UNOS and Arbor Research under contract with HHS. The authors alone are responsible for reporting and interpreting these data; the views expressed herein are those of the authors and not necessarily those of the U.S. Government.

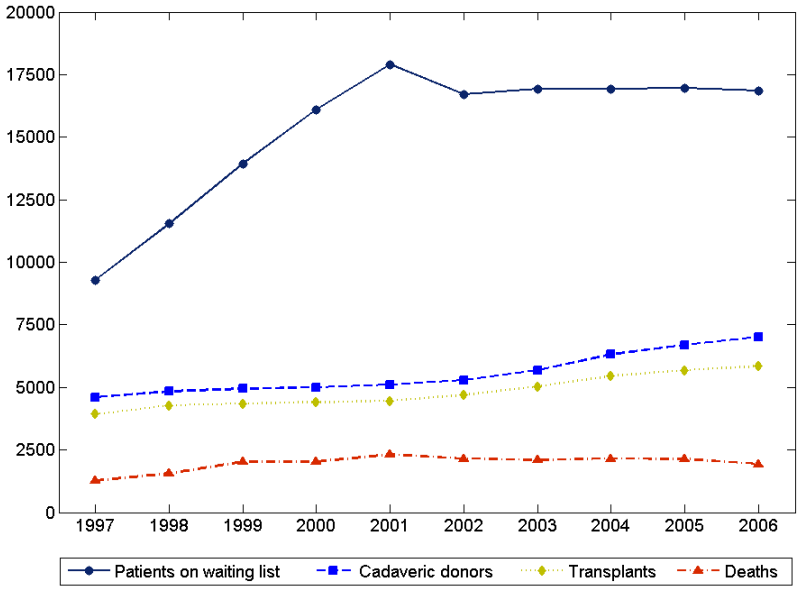


Figure 1.2: Recent trends in US liver transplantation (1997-2006) [66]

because they are considered too sick to be transplanted [66]. By the end of 2006, 64.1% of all active patients on the waiting list had been waiting for at least 1 year and 46.6% had been on the list for at least 2 years [66]. All of these facts motivate the need for research on better management of these scarce resources.

There have been several reports indicating that patients who are most likely to benefit should be given priority so that the use of the donated livers is maximized [101]. UNOS Ethics Committee [155] notes that *"Because donated organs are a severely limited resource the best potential recipients should be identified. The probability of a good outcome must be highly emphasized to achieve the maximum benefit for all transplants."* The problem of optimally using the donated organs have been analyzed from two perspectives. The societal perspective looks at the entire system and designs an allocation mechanism to optimize an objective or combination of objectives such as minimizing refusals and maximizing aggregate benefit to all patients. The patient's perspective, on the other hand, considers individual patients separately and optimizes the organs to accept and reject for each patient under a fixed allocation mechanism.

Looking at the problem from an individual patient's perspective, this dissertation assumes that the available patient prioritization policy is fixed and helps the patient make a selection among offered livers to maximize the benefit to herself. The waiting list characteristics has significant impact on this optimization problem, as the composition of the waiting list determines the priorities of the patients each time an organ becomes available for transplantation. This dissertation makes explicit use of the waiting list knowledge in optimizing the accept/reject decision problem from a patient's perspective.

1.1 LIVER ALLOCATION SYSTEM

Organ Procurement and Transplantation Network (OPTN) is an umbrella establishment charged by federal legislation to develop and implement an equitable system for organ allocation in the US and to compile data from US transplant centers. The United Network for Organ Sharing (UNOS) is the sole administrator of this network since its establishment.

The liver allocation system divides the country into 11 mutually exclusive geographic areas, called regions. These regions are further divided into smaller mutually exclusive subgroups, called the donation service areas of Organ Procurement Organizations (OPOs). Currently, there are about 60 OPOs serving unique areas of varying sizes, population densities, donation rates, and transplantation activities. UNOS handles all its activities through these OPOs.

UNOS uses slightly modified allocation policies for adult and pediatric liver patients. This dissertation focuses only on adult patients, therefore we provide a brief overview of the adult liver allocation algorithm (see [156] for more details of this algorithm).

When a cadaveric liver becomes available, UNOS uses the algorithm depicted in Figure 1.3 to prioritize the patients on the waiting list. This algorithm divides the patients into three categories in terms of geography: local patients, regional patients, and national patients. Local patients are those that are registered to the system from the same donation service area as the harvested organ[2]. Regional patients are those that are outside of the harvesting

[2]We will use 'harvesting OPO,' or shortly 'OPO,' throughout the rest of the dissertation to mean the

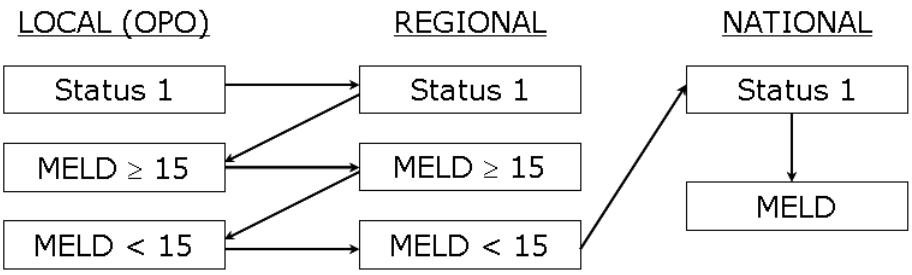


Figure 1.3: UNOS liver allocation policy

OPO but within the same region as the harvesting OPO. Finally, national patients are those remaining outside the harvesting region. This algorithm also divides the patients into two categories in terms of disease severity: Status 1 patients and MELD (Model for End-stage Liver Disease) patients. Status 1 patients are defined to be fulminant liver disease patients who have a life expectancy of less than a week without a transplant. MELD patients are those patients that are assigned a MELD score to measure the severity of the disease. UNOS restricts MELD scores to be integers ranging between 6 and 40, where a higher score indicates more severe disease. MELD scores are simply calculated by a regression equation that uses three lab test results (i.e., bilirubin, INR, and creatinine) from the patient. A patient's MELD score can go up or down over time depending on the status of her disease. We should also note that the MELD scoring system used by UNOS is slightly modified from the originally proposed formula [79, 174].

When a liver arrives, it is first offered to local-Status 1 patients. If no one is found in this category, the search is expanded to Status 1 patients in the region. If the liver is still not transplanted in this category, the algorithm goes back to the local area looking for MELD patients with a MELD score of at least 15. And the search continues as depicted by the arrows in Figure 1.3 until some patient accepts the offer or the liver is discarded. Within each MELD box in Figure 1.3, organs are offered to patients in decreasing MELD score, where ties are broken first by blood type compatibility and then by the cumulative waiting time of patients at their current or higher MELD scores.

donation service area of the OPO harvesting the organ.

UNOS considers the final decision of whether or not to use the offered liver to be "the prerogative of the transplant surgeon and/or the physician responsible for the care of the patient" [156]. Therefore, patients reject offers without any penalty. As a result, a rejecting patient has full access to future offers and her history of rejections does not affect her priority for the new offers. Indeed, despite the scarcity of donated organs, almost half of the liver offers are rejected by the first surgeon to whom the offer is made [70]. In another analysis, Alagoz [2] reports that 60% of all liver offers are rejected.

As Howard [70] reports, surgeons reject low quality organs for healthy patients in the hope that they may receive a better organ offer in the future. Several characteristics of the donor may affect the perceived quality of the donated organ such as length of intensive care unit stay and antecedents of hypertension [37]; existence and degree of steatosis [123]; race, height and involvement in a cerebravascular accident [50]; and age, race, blood type and gender [117].

1.2 PROBLEM STATEMENT AND CONTRIBUTIONS

The optimization of the accept/reject decision problem described in Section 1.1, within the confines of the current allocation system, is the focus of this dissertation. We assume that the decision making process is joint between the patient and an agent (such as the patient's physician and/or surgeon) who acts in the patient's best interest. However, for expositional simplicity, we refer the decision-maker as the patient throughout the rest of the dissertation.

In solving this optimization problem, we are particularly interested in the objective of maximizing the patient's expected total discounted life. We leave the consideration of other objectives for future research. Discounting future health outcomes is a standard practice in medical decision making. A detailed discussion of the issues related to discounting is provided in [2, 130] and references therein.

Several researchers consider the organ accept/reject problem. Section 2.3.2 provides more detailed discussion of this literature. Most of this literature makes unrealistic assumptions such as patient health does not change over time, each organ may be offered to at most

one patient, new patients do not arrive, listed patients do not die, all patients are homogeneous, organ quality does not deteriorate, and all patients have the same pre-transplant life expectancy. This dissertation relaxes each of these assumptions.

The solution to this optimization problem depends on three main factors: the patient's current and future health, the quality of the offered liver, and the prospects of receiving future offers. The prospects of receiving future offers are heavily influenced by the composition of the waiting list. This includes the characteristics of the specific patient under consideration as well as those of the other patients on the waiting list.

Historically, much of the relevant information about the patients on the waiting list has been hidden. While this practice ensures some level of confidentiality, it also forces patients to make accept/reject decisions with incomplete information. To alleviate this problem, UNOS now publishes coarse (yet still incomplete) descriptions of the waiting list on their website [157]. More detail about this coarse waiting list descriptions is given in Section 4.1.

A partial description of the waiting list, however, still results in some loss to the patient since this lack of knowledge may result in suboptimal decisions. We call this loss due to incomplete rank knowledge the *patient's price of privacy.* We define this quantity to be the number of life days gained by the patient when she acts optimally based on full knowledge of the waiting list, as opposed to her optimal actions under the current allocation system. This quantity may also be interpreted as the value of obtaining rank information. We leverage our model analysis to provide quantitative estimates of this price of privacy in terms of overall life expectancy. That is, we do not advocate any particular change in the privacy policy; rather, we wish to quantify the costs of this privacy to the patient.

In order to estimate a patient's price of privacy, we develop two Markov decision process models that explicitly use the waiting list information while optimizing the patient's accept/reject decision problem. The first of these models considers completely observable waiting list, whereas the second model considers partially observable waiting list as in the current system. We compare the results of these two models to obtain our estimates for a patient's price of privacy. Finally, to assess how MDP based optimization models perform, we conduct an extensive empirical study comparing the results of the model-suggested decision rules to the actual decisions made by real patients. In this empirical analysis, we restrict our

study to patients transplanted with living donors only. The main reason for this limitation is the unavailability of real data for patients transplanted with cadaveric donors that would reveal the decisions made by patients at various rank states.

1.3 OVERVIEW OF THE DISSERTATION

The remainder of this dissertation is organized as follows. Chapter 2 discusses the relevant literature, including key results for completely and partially observable Markov decision processes (MDPs) and their applications in medical decision making. Chapter 3 describes and presents the results of an MDP formulation for the accept/reject decision problem faced by liver patients under a completely observed waiting list, assuming that the liver allocation system is in equilibrium and this equilibrium will not change upon completely revealing waiting list information. We explore structural properties of this model and present detailed numerical study parameterized with clinical data along with our preliminary estimates for patients' price of privacy. Chapter 4 extends the MDP formulation of Chapter 3 to model the partial information availability in the current allocation system. We explore structural properties and solve this model using clinical data to obtain more precise estimates of a patient's price of privacy. Chapter 5 considers the issue of whether patients behave according to their optimal policies, as our price of privacy estimates in the previous two chapters assume. We introduce a nested MDP model for this purpose and present the results of an empirical study, in which the actual decisions of real patients are compared to the decisions suggested by the nested MDP model. Finally, the dissertation concludes in Chapter 6 by summarizing and discussing our results and limitations along with possible extensions for future research.

2.0 LITERATURE REVIEW

In this chapter, we briefly discuss the related literature to our study. Sections 2.1 and 2.2 provide an overview of Markov decision processes (MDPs) and partially observable MDPs, respectively. In Section 2.3, we review the MDP applications in medical decision making as well as other mathematical models applied to organ transplantation. Finally, in Section 2.4, we review other related literature.

2.1 MARKOV DECISION PROCESSES

In this section, we provide a brief overview of completely observable Markov decision processes (MDPs). We content ourselves with listing basic definitions and results without proofs. More extensive treatment of the subject is provided in classical books [18, 112].

We are concerned with sequential decision making problems under uncertainty where there is a need to make decisions at discrete time points $T = \{1, \ldots, N\}$, called *decision epochs*, throughout the lifetime of the system under consideration. Note that N need not be finite. At any given decision epoch t, the system occupies a *state* s in the *state space* $\mathcal{S}$. We assume that the actual state of the system, s, is completely known to the decision maker when a decision is to be made. We relax this assumption in Section 2.2. The decision maker has control over the system by modifying its trajectory through choosing an *action* a from the *action space* $\mathcal{A}_s$ at each decision epoch t. We assume that $\mathcal{S}$ and $\mathcal{A}_s$ are both discrete. Uncertainty in the system is governed by its probabilistic evolution. That is, when the system is in state s at decision epoch t and the decision maker chooses action a, the state of the system at the next decision epoch $t + 1$ is determined by the *transition probability*

distribution $P_t\{\cdot|s,a\}$. The fact that the next state of the system is only dependent on the current state and action is called the Markovian property of the process. Furthermore, as a result of choosing action a while the system is in state s at decision epoch t, the decision maker receives an *immediate reward*, $r_t(s,a)$. A more general reward function, $r_t(s,a,s')$, may be defined when the rewards are dependent not only on the current state and action pair (s,a) but also on the next state s'. However, for all practical purposes, this reward definition does not complicate our model as we can take its expected value by computing

$$r_t(s,a) = \sum_{s' \in \mathcal{S}} r_t(s,a,s') \cdot P_t\{s'|s,a\}.$$

Note that we can add more generality to our model by allowing the state space $\mathcal{S}$ and the action space $\mathcal{A}_s$ for all $s \in \mathcal{S}$ depend explicitly on t. This generalization would not affect the theory and the results, and, therefore, we remove the time index from these pieces of the model. The collection of objects $\{T, \mathcal{S}, \mathcal{A}_s, P_t\{\cdot|s,a\}, r_t(s,a)\}$ is referred to as a *Markov decision process*.

A *decision rule* is simply a prescription for choosing actions throughout the lifetime of the system. A decision rule d_t at decision epoch t is *Markovian* if the prescribed action choice is dependent only on the current system state, whereas a *history-dependent* decision rule prescribes actions that depend on the entire history of the system, which includes all the (state,action) pairs realized up until the current decision epoch. The number of possible histories can indeed be huge, therefore we are interested in conditions that guarantee the existence of Markovian decision rules that are optimal.

A decision rule is called *deterministic* if it prescribes to choose an action with certainty (i.e., with probability 1) for each state, whereas a *randomized* decision rule prescribes a non-degenerate probability distribution over the set of possible actions for at least one state. Note that a probability distribution is degenerate if it accumulates all of its mass onto a single element in its support. Due to its simplicity, we prefer an optimal deterministic decision rule over an optimal randomized decision rule. Indeed, it can be shown that if the state and action spaces are finite, and the states are completely observable, then there exists a deterministic Markovian decision rule that is optimal [112].

A *policy* δ specifies which decision rule to use at each decision epoch t, that is, $\delta = (d_1, d_2, \ldots, d_{N-1})$. A policy is called *stationary* if it specifies the same decision rule at each decision epoch (i.e., δ is stationary if $d_t = d$ for all $t \in T$). It can be shown that for an *infinite-horizon* MDP (i.e., $N = \infty$), if the state and action spaces are finite, and the states are completely observable, then there exists a stationary policy that is optimal [112].

Among the exponential number of available policies, we would like to choose one that maximizes the *expected total discounted reward*

$$\mathbb{E}\left[\sum_{t=0}^{N-1} \lambda^t r_t(s,a)\right] \tag{2.1}$$

where $0 \leq \lambda < 1$ is the discount rate. If $\lambda = 1$ in equation (2.1), it is called the *expected total reward criterion*. Other optimality criteria include *average reward criterion* [112].

In finite-horizon problems (i.e., $N < \infty$), any choice for $\lambda \in [0, 1]$ does not affect any theoretical results, but might affect the decision maker's preference for policies. Furthermore, the expected total reward and the average reward criteria give the same optimal policies in finite-horizon problems [112]. Therefore, we omit λ and let $u_t^*(s)$ be the optimal value function in state s when there are $N-t$ decision epochs left until the end of the horizon. The optimal value function that maximizes the expected total reward for finite-horizon problems can be computed by recursively solving the following Bellman equations:

$$u_t^*(s) = \max_{a \in \mathcal{A}_s} \left\{ r_t(s,a) + \sum_{s' \in \mathcal{S}} P_t\{s'|s,a\} \cdot u_{t+1}^*(s') \right\} \text{ for } t = 1, \ldots, N-1; s \in \mathcal{S} \tag{2.2}$$

with boundary conditions

$$u_N^*(s) = r_N(s) \quad \text{for } s \in \mathcal{S}, \tag{2.3}$$

where $r_N(s)$ represent the terminal rewards received when the process is in state s at the end of the planning horizon. The optimal action a_t^* at each stage $t = 1, \ldots, N-1$ is chosen as the argument $a \in \mathcal{A}_s$ that maximizes the right-hand side of equation (2.2). *Backward induction* algorithm provides an efficient method for solving the optimality equations (2.2)-(2.3).

For an infinite-horizon problem, an optimal stationary Markovian policy that maximizes the expected total discounted reward can be found by solving the following recursive equations:

$$v^*(s) = \max_{a \in \mathcal{A}_s} \left\{ r(s,a) + \lambda \cdot \sum_{s' \in \mathcal{S}} P\{s'|s,a\} \cdot v^*(s') \right\} \text{ for } s \in \mathcal{S}, \tag{2.4}$$

where $v^*(s)$ is the optimal value function in state s. The optimal action a^* is chosen as the argument $a \in \mathcal{A}_s$ that maximizes the right-hand side of equation (2.4). *Value iteration*, *policy iteration*, and *linear programming* are typical algorithms used for solving the optimality equations (2.4).

The literature on Markov decision processes and its applications is huge. Classical books by Bellman [15] and Howard [71] discuss early developments in this field. More recent developments in the subject with detailed solution algorithms and comprehensive lists of MDP applications can be found in [8, 18, 49, 52, 65, 112, 171]. Typical application areas include inventory control [32, 109, 125], queuing control [140, 144, 160], finance [127], communication networks [9], machine maintenance [31, 128], and medical decision making [1, 4, 72, 86, 92, 131]. We discuss MDP applications in health care in greater detail in Section 2.3.1

2.2 PARTIALLY OBSERVABLE MARKOV DECISION PROCESSES

In this section, we provide a brief overview of partially observable Markov decision processes (POMDPs). We content ourselves with listing basic definitions and results without proofs. More extensive treatment of the subject is provided in [17, 26, 99] and the references therein.

Many real-life problems does not allow the decision maker to completely observe the state of the system under consideration, in which case the completely observable MDP model of Section 2.1 no longer applies. The seemingly innocent relaxation of the assumption of complete observability to partial observability comes at a significant cost in complexity.

A POMDP model shares all the parameters of an MDP model, namely, T, $\mathcal{S}$, $\mathcal{A}_s$, $P_t\{\cdot|s,a\}$, and $r_t(s,a)$. A state $s \in \mathcal{S}$ in the POMDP model is usually referred to as a *core state* to distinguish it from a belief state (to be defined later). We relax the completely

observable system state assumption of an MDP model with partially observable system state. As a result of this relaxation, we add two more parameters to a POMDP model: (i) a *set of observations*, $\mathcal{Z}$, that is produced by the system and is accessible to the decision maker, and (ii) a set of *observation probabilities*, $O_t\{\cdot|s,a\}$, that gives probabilistic information about accessing observations. Specifically, $O_t\{z|s,a\}$ denotes the probability of receiving observation z at time t, while the system actually occupies core state s at time t upon taking action a at time $t-1$. Finally, the immediate rewards can be defined in a more general fashion as a function of the current state and action pair, (s,a), as well as the state and observation pair in the next decision epoch, (s',z'). This generalization, again, does not complicate our model as we can take its expected value by computing

$$r_t(s,a) = \sum_{s' \in \mathcal{S}} \sum_{z' \in \mathcal{Z}} r_t(s,a,s',z') \cdot O_t\{z'|s',a\} \cdot P_t\{s'|s,a\}.$$

The collection of objects $\{T, \mathcal{S}, \mathcal{A}_s, \mathcal{Z}, P_t\{\cdot|s,a\}, r_t(s,a), O_t\{\cdot|s,a\}\}$ is referred to as a *partially observable Markov decision process*. For notational convenience, we let $\mathcal{A} = \cup_{s \in \mathcal{S}} \mathcal{A}_s$ and drop the time index t from our discussions.

As in Section 2.1, we are interested in finding the optimal policy, but any policy mapping the core states into actions (i.e., Markovian with respect to core states) would not have any meaning in practice, as core states are not accessible to the decision maker. Extrapolating the Markovian idea to define a policy that maps observations to actions produces a legitimate policy, but can perform very poor as such policies may not be within the set of optimal policies for some problems [76]. Research considering policies defined over finite histories of the process revealed that such policies may be arbitrarily poor [170]. It turns out that the optimal policy of a POMDP requires remembering the entire history of the process.

Although it may not be possible to record the entire history of a process, all the information available in this history can be captured by a summary statistic, referred to as the *belief state* or *information state*. It can be shown that a belief state is a sufficient statistic for the entire history of the process [12, 145]. A belief state, $\boldsymbol{\pi}$, is simply defined to be a probability distribution over the set of core states, $\mathcal{S}$, where the s^{th} component of $\boldsymbol{\pi}$, denoted π_s, represents the probability of being in core state $s \in \mathcal{S}$. Let $\Pi(\mathcal{S})$ be the set of all probability distributions over the set $\mathcal{S}$.

Since we will use belief states to write optimality equations, we need to know how to make transitions among belief states. For this purpose, let $\boldsymbol{\pi}$ be a belief state on-hand at time t and z' be the observation received in the next decision epoch $t+1$ upon taking action a at time t. The j^{th} component of the updated belief state $\boldsymbol{\pi}'$, denoted π'_j, can be computed by

$$\pi'_j(\boldsymbol{\pi}, a, z') = \frac{\sum_{s\in\mathcal{S}} \pi_s \cdot P\{j|s,a\} \cdot O\{z'|j,a\}}{\sum_{j'\in\mathcal{S}} \sum_{s\in\mathcal{S}} \pi_s \cdot P\{j'|s,a\} O\{z'|j',a\}}. \tag{2.5}$$

which is derived using Bayes' formula and simple probability arguments.

For finite-horizon problems (i.e., $N < \infty$), the search for an optimal policy can be restricted to Markovian policies defined over the belief space, $\Pi(\mathcal{S})$, [12, 87], and the optimal value function can be found by solving the following set of recursive equations [26, 87]:

$$u_t^*(\boldsymbol{\pi}) = \max_{a\in\mathcal{A}} \left\{ r(\boldsymbol{\pi}, a) + \sum_{s\in\mathcal{S}} \sum_{s'\in\mathcal{S}} \sum_{z'\in\mathcal{Z}} \pi_s \cdot P\{s'|s,a\} \cdot O\{z'|s',a\} \cdot u_{t+1}^*(\boldsymbol{\pi}') \right\}$$
$$\text{for } t = 1, \ldots, N-1; \quad \boldsymbol{\pi} \in \Pi(\mathcal{S}) \tag{2.6}$$

with boundary conditions

$$u_N^*(\boldsymbol{\pi}) = \sum_{s\in\mathcal{S}} \pi_s \cdot r_N(s) \quad \text{for } \boldsymbol{\pi} \in \Pi(\mathcal{S}), \tag{2.7}$$

where $r(\boldsymbol{\pi}, a) = \sum_{s\in\mathcal{S}} \pi_s r(s,a)$. Note that $\boldsymbol{\pi}'$ in equation (2.6) is also dependent on the current belief state $\boldsymbol{\pi}$ as well as the observation z' and action a, however this dependency is suppressed for notational clarity.

For infinite-horizon problems (i.e., $N = \infty$), given a discount rate $\lambda \in [0, 1)$, the search for an optimal policy can be restricted to stationary Markovian policies defined over the belief space, $\Pi(\mathcal{S})$, [19, 87], and the optimal value function that maximizes the expected total discounted reward can be found by solving the following set of recursive equations [26, 99]:

$$v^*(\boldsymbol{\pi}) = \max_{a\in\mathcal{A}} \left\{ r(\boldsymbol{\pi}, a) + \lambda \sum_{s\in\mathcal{S}} \sum_{s'\in\mathcal{S}} \sum_{z'\in\mathcal{Z}} \pi_s P\{s'|s,a\} O\{z'|s',a\} v^*(\boldsymbol{\pi}') \right\}, \text{for } \boldsymbol{\pi} \in \Pi(\mathcal{S}). \tag{2.8}$$

Note again that $\boldsymbol{\pi}'$ in equation (2.8) is a function of current belief state $\boldsymbol{\pi}$ as well as the observation z' and action a as seen in equation (2.5), however, these dependencies are suppressed for notational clarity.

Partially observable systems have been studied at least since 1960s. Drake [44] provides the first explicit POMDP model. The first *exact* algorithm for solving POMDPs is Sondik's one-pass algorithm [141]. The main complication in solving the dynamic programming recursion for a POMDP problem is the fact that the belief space is a continuum. This means we essentially have infinitely many equations with potentially infinite number of variables in the optimality equations. A key result that resolves this complication is provided by Sondik [137, 141], who proves that the optimal value function for a POMDP is piecewise linear and convex for every finite decision epoch, and can be approximated arbitrarily closely by a piecewise linear and convex function over an infinite horizon. This key result allows the computation of the value function over the entire belief space in a finite time, however the complexity of the value function (i.e., the number of linear segments of the function) can grow exponentially with the number of dynamic programming steps performed [63]. Sondik's two-pass algorithm [142] determines the optimal policy of infinite horizon discounted POMDPs and solves a problem with 2 states, 2 actions, and 2 observations using this method. White [164] extends Sondik's one-pass algorithm to finite-horizon partially observable semi-Markov decision processes. Hansen [58] proposes a new policy iteration algorithm for POMDPs that represents a policy as a finite-state controller and is found to be more efficient than Sondik's algorithms. He reports that his policy iteration algorithm is, on the average, 40-50 times faster in finding the optimal solution compared to the value iteration algorithm. He does not report which problems he tested his algorithm. Several other exact algorithms have been proposed for solving POMDPs, most notably Cheng's relaxed region and linear support algorithms [33], as well as the witness algorithm [26, 27] and the variations of the incremental pruning algorithm [26, 28]. Interested readers are referred to [26, 90, 182] for detailed surveys of available algorithms.

Given that the current state of the art can solve POMDP problems with only a few states using the exact algorithms [58, 91], grid-based approximations became a natural and widely used technique to find approximate solutions. Finite-grid algorithms find approximate values to the optimal value functions by sampling a finite number of points, called the grid, from the continuous belief space and using interpolation to find the values of those belief points in the belief space that are not placed in the grid. Several variations exist in the literature. Lovejoy

[89] describes a *fixed-resolution regular grid* that samples points from the belief simplex in regular intervals and does not vary these points throughout the rest of the iterations. The advantage of this method is its efficient interpolation, whereas the obvious disadvantage is the exponential growth in the size of the grid as the number of core states or the resolution of the grid increase. Hauskrecht [61, 62] and Brafman [20] describe a *variable-resolution non-regular grid* method that samples points from the belief simplex in non-regular intervals and adds/removes to/from the grid over iterations as perceived necessary by heuristic rules. The advantage of this method is its more economic use of grid points in approximating the value function, whereas its disadvantage is inefficient interpolation times. Zhou and Hansen [183] describes a *variable-resolution regular grid* method that attempts to combine the strengths of the previous two methods. We describe a *fixed-resolution non-regular grid* method in Section 4.4.2, where we take advantage of a special structure of our problem. Interested readers are referred to Lovejoy [90] and Zhou and Hansen [183] for further details on grid-based methods.

Several researchers explore conditions that ensure structured optimal policies and value functions. Ross [120] considers a Markovian deteriorating process with 2 states, 3 actions, and costly imperfect information. He shows conditions under which the optimal policy has the monotonic at-most-four-region ($AM4R$) structure. Rosenfield [119] extends the results in [120] to more than 2 finite states. White [166] proves similar results under less restrictive conditions. Albright [7] considers a 2-state POMDP and derives conditions that guarantee optimal value functions and optimal policies to be monotone in the belief states. White [167] considers a machine replacement problem with two actions, where one of the actions puts the system in a state as good as new. He provides sufficient conditions for this problem that guarantee the existence of monotone optimal policies and value functions with respect to stochastically ordered belief states. White [169] considers single-stage partially observed problems and provides sufficient conditions for the optimal policy to be monotone on stochastically ordered belief states. White [168] provides conditions that guarantee the existence of monotone optimal policies for completely observed and completely unobserved POMDPs. Lovejoy [87] considers a general discrete-time, finite POMDP problem and uses the monotone likelihood ratio (MLR) partial order to show that the optimal value function

is monotone. Maillart [93] considers a maintenance optimization problem with condition monitoring and derives conditions under which the optimal value function is monotone with respect to *MLR*-ordered belief states when observations yield perfect information about the system conditions under which the optimal policy has *AM4R* structure. For further details on structural analysis, we refer the reader to references cited above and to the survey article by Monahan [99], as well as references therein.

Finally, typical application areas of POMDPs include machine maintenance and replacement [47, 93, 120, 165, 166, 167], inventory control [154], inspection of structural units such as paved roads, bridges, and buildings[48], network troubleshooting [151], search for a moving object [45, 108], medical diagnosis and treatment [67, 63, 72, 74, 94], health care systems analysis [138], human learning and instruction [135], optimal design of questionnaires when the answers are not truthful [163], optimal teaching strategies [136], and fisheries [83]. We refer the reader to [25, 99] for a more detailed review of POMDP applications. applications survey

2.3 APPLICATIONS IN MEDICAL DECISION MAKING

In this section, we summarize various applications in medical decision making. In Section 2.3.1, we consider MDP and POMDP models applied to health care problems. In Section 2.3.2, we describe mathematical models used in organ transplantation modeling.

2.3.1 MDP applications in health care

The material in this section is mostly from Schaefer et al. [126]. Lefevre [86] uses a infinite-horizon continuous-time MDP model for the problem of controlling an epidemic in a closed population. The state of the system is chosen to describe the number of infected people in the population. The decision maker control the system by deciding the number of people to quarantine and by the amount of medical treatment to apply to the infected population. The optimality criterion used is the minimization of the expected total discounted costs.

Magni et al. [92] uses a discrete-time MDP model to decide on therapy for mild hereditary spherocytosis. The state of patients is described through the severity of gallstones and the state of the spleen. They consider objective of maximizing the patient's quality-adjusted life years, where they used published mortality rates and previous studies to estimate the quality-of-life utilities as well as the transition probabilities among different states.

Shechter et al. [131] use an infinite-horizon discrete-time MDP model for the effective management of the HIV therapy planning. They specifically consider the optimal initiation time of highly-active antiretroviral therapy (HAART)for HIV care. They use prognostic variables such as patient's CD4 count in her blood as the state of their model. They use clinical data to estimate transition probabilities among various states and to estimate patient survival. They consider the objective of maximizing the total expected reward to the patient. They solve their model for a variety of patients and employ sensitivity analysis on various parameters of the model. Their main policy implication is that it is optimal to initiate therapy in each CD4 count.

Hu et al. [72] uses a POMDP model for the problem of choosing an appropriate drug infusion plan for administering anesthesia. The partially observable states of their model include patient parameters such as anesthesia concentration in the blood and clearance rate of a given drug. They use heuristics to implement numerical solutions for their model.

Hauskrecht and Fraser [63] develop a POMDP model for the management of patients with ischemic heart disease, a disease caused by an imbalance between the supply and demand of oxygen to the heart. The states of their POMDP model consists of a mix of perfectly and partially observable components. They generate data for their POMDP model informally through published results in consultation with a cardiologist. They use heuristic methods to solve the POMDP model, however the details of the heuristics are not clear nor are the details (such as the size) of the problems solved.

Ivy [74] uses a POMDP model to develop a cost-benefit analysis of mammogram frequency and treatment options for breast cancer with the objective of minimizing the total expected cost over a patient's lifetime. The model consisted of three states: no disease, non-invasive breast cancer, and invasive breast cancer. Two types of exams, clinical breast exams and mammograms, could be used to reveal further information about the partially

observable patient state. She estimates the parameters of her model from the published literature and characterizes the optimal decisions.

2.3.2 Models for organ transplantation

Several researchers consider the organ accept/reject problem from an individual patient's perspective ([1, 4, 6, 5, 39, 68, 70]) or from the society's perspective ([40, 41, 116, 121, 147, 177, 178, 181, 180]) or from a joint perspective ([146, 147, 148]). Most of this literature makes unrealistic assumptions such as patient health does not change over time, each organ may be offered to at most one patient, new patients do not arrive, listed patients do not die, all patients are homogeneous, organ quality does not deteriorate, offered organs cannot be declined, and all patients have the same pre-transplant life expectancy. We refer the reader to [2] for a detailed discussion of the organ transplantation literature.

A notably exception is the sequence of papers by Alagoz et al. [4, 5, 6], where they do not make any of the assumptions listed above. Alagoz et al. [4] consider the problem of optimally timing a living-donor liver transplant to maximize the patient's total expected discounted reward. They present a discrete-time infinite-horizon MDP model, in which the state of their model is described by patient's health status. Similar to our work, they use MELD scores in their numerical study to represent patient's health. They derive conditions that guarantee control-limit optimal policies. Their computational experiments parameterized by clinical data reveal that it is indeed optimal to delay the transplantation until patient's MELD score reaches a certain threshold. More details of this model is provided in Section 5.2.

Alagoz et al. [6] extend their previous living-donor model given in [4] for patients receiving cadaveric-donor livers. They present a Markov decision process (MDP) model in which the state is described by patient health and organ quality. For each possible state, provided an offer is made, the patient chooses to either accept or reject the offer so as to maximize her total expected reward. Their approach captures the effects of the waiting list *implicitly* through the organ arrival probabilities, which are assumed to be a function of patient health. Under the current liver allocation policy, however, the frequency and the quality of liver offers made to an individual patient are significantly affected by the physiology and the geographic

location of the other patients on the waiting list. Therefore, an *explicit* consideration of the waiting list information is needed, which is one of the goals of this proposal.

Alagoz et al. [5] present combined model, in which an ESLD patient waiting for a cadaveric-donor liver also has access to a living-donor liver. They present a combined MDP model, in which three possible actions are available to the the decision maker: accept the cadaveric liver offer, transplant the living-donor liver, or wait for one more decision epoch. They provide a structural analysis of this model and derive conditions under which the optimal policy has monotonic at-most-three region structure. Their computational results parameterized by clinical data also confirm this structure in the optimal policies.

2.4 OTHER RELATED LITERATURE

This section lists other literature that is found to be related to our study. Rather than going into details of this literature, we point out recent survey articles whenever available.

Several researchers use queuing models to analyze organ transplant waiting lists [146, 148, 177, 180]. Most of the queuing models are concerned with kidney transplantation. We refer to [179] for a recent review of queuing based models in kidney transplantation. However, the liver transplant waiting list is much more complex than a simple queue [69, 156] since the priorities assigned to patients are a function of geography and health, and this fact renders queuing approaches to liver transplantation inappropriate.

As noted in Section 1.2, our study is concerned with the value of accessing more precise waiting list information. Value of information analysis is recommended as particularly useful in framing complex decision-making problems characterized by large uncertainties and high stakes [104, 36]. The value of information literature is huge and we refer the reader to the classical book by Raiffa and Schlaifer [113] for an introduction to the topic and to the recent comprehensive survey by Yokota and Thompson [176] for a review of applications in health care. Typical applications of value of information analysis in the operations research literature include information flow in supply chains [56] and queuing systems with informed customers [173].

The public revelation of the waiting list information naturally gives rise to a gaming environment, where each patient, when making their own decisions, has to consider the possible decisions of other patients on the waiting list. To the best of our knowledge, the only organ transplantation paper to consider such game-theoretic aspects is Su and Zenios [146]. They model the kidney transplant waiting system as an $M/M/1$ queue with exponential reneging that represents patient death. They assume a homogeneous patient population despite their well-documented heterogeneity, which enables a detailed competitive equilibrium analysis, and discuss the effects of different queuing disciplines such as first-come-first-served and last-come-first-served on the system performance. They continue to acknowledge the associated shortage with the homogenous patient population assumption by saying that "...the absence of patient heterogeneity in our models may create suspicion about the validity of our findings." Unfortunately, such a queuing model is inappropriate for liver transplantation because patient priorities evolve dynamically with their health, and different livers may induce different priorities due to differences in geography and blood type. Furthermore, acknowledging the heterogeneity of the patients leads to an intractable asymmetric multi-player non-zero-sum stochastic game. We refer the interested reader to [102, 159] for recent reviews of the state-of-the-art in stochastic games.

A related stream of research analyzes equilibria in queueing models. We refer the interested reader to [10, 60] for recent surveys of this literature. This stream of research focuses on questions like when to join a queue, which queue to join when there are multiple queues, and what priority level to purchase when different priorities are allowed. All of these questions are related to customers' decisions at the time of their arrivals. However, in our case, patients are prioritized at the time of an organ arrival according to the liver allocation policy. Therefore, they do not have any choice at the time of their arrival but rather a prerogative to refuse an organ (service) offer. Furthermore, we also allow patient priorities to change over time and we explicitly use the rank information of the patient to make this decision. A methodologically similar paper that models the decision making (although the customers are not the decision makers) using rank information is [150], which formulates an MDP model to find optimal replacement policies for a single motion picture exhibitor ignoring the competition between theater chains. They contend with providing a numerical analysis of

the model without attempting any structural results. In their competitive equilibrium analysis within the kidney allocation system, [146] also use the rank information of the patients to characterize the rank-dependent threshold policies. However, these characterizations are heavily influenced by the assumption of homogeneous patients, which we can not justify in the context of this paper.

Recently, Hansen et al. [59] developed an exact algorithm for finite-horizon partially observable stochastic games that combines the dynamic programming updating for POMDPs and the iterated deletion of dominated strategies algorithm for normal form games. However, this algorithm is tested only on small problems, which leaves the questions surrounding its applicability for real-sized problems.

3.0 EXPLICIT WAITING LIST MODEL

This chapter considers the problem of optimally deciding which liver offers to accept and which ones to reject from an individual patient's perspective. Our goal in this chapter is to first develop and analyze a model for this problem that explicitly incorporates the waiting list information into patient's decision-making. Secondly, we want to use this model to estimate patient's price of privacy. Most of the material in this chapter is summarized in Sandıkçı et al. [124].

As discussed in Chapter 2, several researchers consider the organ accept/reject problem. Of this body of work, the most relevant is Alagoz et al. [6]. Their approach captures the effects of the waiting list *implicitly* through the organ arrival probabilities, which are assumed to be a function of patient health. As discussed in Chapter 1, under the current liver allocation policy, however, the frequency and the quality of liver offers made to an individual patient are significantly affected by the physiology and the geographic location of other patients on the waiting list. Therefore, a model that captures the effects of the waiting list in a more *explicit* fashion is needed.

Throughout this chapter, we assume that the waiting list is completely observable and, hence, patients have perfect information about their waiting list ranks. We also assume that patients are self-interested agents and do not explicitly consider the possible actions other candidates may take in making their own decisions. This latter assumption, while unlikely to hold, yields computational and analytical tractability and, hence, enables us to provide estimates for the true price of privacy.

The rest of this chapter is organized as follows. In Section 3.1, we present the Markov decision process model formulation. In Section 3.2, we introduce a new class of stochastic matrices, termed *CCD* matrices, which facilitate the derivation of structural results in Section

3.3. We present a detailed computational study in Section 3.4. Finally, we conclude in Section 3.5 by summarizing our results and pointing out some limitations.

3.1 MARKOV DECISION PROCESS MODEL

Consider an ESLD patient who must decide (with her physician and/or surgeon) whether to accept or reject a liver offered for transplantation so as to maximize her total expected discounted reward. We assume the patient makes this decision at discrete time periods. If a liver is not offered in a particular period, then the patient is forced to 'wait' until the next period. The non-trivial decision to be optimized is when there is a liver offered. If the patient chooses to wait, then she accrues an intermediate reward which is a function of her current health status, and faces the same problem at the next time period, provided she lives. If, on the other hand, the patient chooses to 'accept' the offer, then she receives a lump-sum terminal reward (e.g., the expected discounted post-transplant survival or quality-adjusted survival). This terminal reward is a function of the patient's current health status as well as the quality of the accepted liver. By choosing to transplant an offered liver, the patient terminates the process.

Among the organ acceptance models that consider the effect of the waiting list, Alagoz et al. [6] implicitly models the waiting list through the organ arrival probabilities. At the other extreme, a *fully explicit* model of the waiting list would track the health, location, blood type, and waiting time of all patients in the list. Our approach balances the additional complexity associated with incorporating information about the waiting list with practical considerations such as model calibration and solution time.

As discussed in Chapter 1, under the current liver allocation mechanism, the priorities assigned to patients are not only determined by the characteristics of the patients in the waiting list but also by the characteristics of the donated liver. Therefore, even in a hypothetical environment in which all the patients' characteristics are constant and no new patients arrive, a currently listed patient may be assigned different priorities for different livers. For instance, the priority of a patient for a liver donated in the same geographic ser-

Table 3.1: Patient characteristics for the hypothetical example

Patient	Geographic area	Blood type	Expected priority	Rank of expected priorities
a	1	A	1.5	1
b	1	AB	2.5	3
c	2	A	2	2
d	3	O	4	4

vice area as she is registered may be significantly different than her priority for an identical liver donated in a different service area. Furthermore, even for two different livers donated in the same geographic area, the patient may be assigned different priorities depending on her blood type compatibility with the donated livers. Recall from Section 1.1 that there are about 60 OPOs representing different geographic areas and 3 blood type compatibility levels (i.e., identical, compatible, and incompatible). Incorporating these two factors alone to model the rank of a patient over all possible livers would lead to a dramatic increase in the size of the state space.

As a compromise, we define a patient's 'rank' as a scalar, namely *the rank of the patient among all patients' expected priorities*, where the expectation is taken over all possible livers. To illustrate this definition, consider a hypothetical example with 4 patients and 2 livers. The characteristics of these patients are given in Table 3.1. Assume that both livers are blood type A; however, Liver 1 is procured in OPO 1, whereas Liver 2 is procured in OPO 2. All else held equal, the order of patient priorities based on the current UNOS policy, from first to last, would be a-b-c-d and c-a-b-d for Livers 1 and 2, respectively. The expected priorities and the rank of expected priorities of these patients are as shown in the respective columns of Table 3.1. For the sake of exposition, from now on, we use 'rank' to mean the rank of the patient among all patients' expected priorities.

We define a state, s, of the MDP model to be composed of the triplet (h, ℓ, k), where h is the patient's health status, ℓ is the quality of the liver being offered, and k is the rank of the patient. We assume the components of the state can take on the following

values: $h \in \Omega = \{1, 2, \ldots, H\}$, where the quality of health is decreasing as h increases; $\ell \in \Phi = \{1, 2, \ldots, L+1\}$, where the quality of the liver is decreasing as ℓ increases and $L+1$ represents no liver being offered; and $k \in \Psi = \{1, 2, \ldots\}$, where the patient moves further from the top of the waiting list as k increases.

For convenience, we add two absorbing states, Δ and ∇, to represent the dead and transplanted states, respectively. Therefore, the state space of the model becomes

$$\mathcal{S} = \mathcal{S}' \cup \{\Delta\} \cup \{\nabla\},$$

where

$$\mathcal{S}' = \{(h, \ell, k) | h \in \Omega, \ell \in \Phi, k \in \Psi\}.$$

The set of possible actions in state $s \in \mathcal{S}$ is

$$\mathcal{A}_s = \begin{cases} \{W\} & \text{if } s = \Delta \text{ or } s = \nabla \text{ or } s = \{(h, \ell, k) | \ell = L+1\}, \\ \{T, W\} & \text{otherwise,} \end{cases}$$

where 'W' stands for rejecting the offer and waiting for one more period, and 'T' stands for accepting the offer and transplanting. The immediate rewards for each possible state-action pair (s, a), such that $s \in \mathcal{S}$ and $a \in \mathcal{A}_s$, is given by

$$r(s, a) = \begin{cases} 0 & \text{if } s = \Delta \text{ or } s = \nabla, \\ r_W(h) & \text{if } s \in \mathcal{S}' \text{ and } a = W, \\ r_T(h, \ell) & \text{if } a = T. \end{cases}$$

The patient does not accumulate any additional rewards once she is dead or transplanted. Furthermore, a pre-transplant patient who chooses to wait for one more period receives an intermediate reward of $r_W(h)$, which is a function of the patient's health status only. Finally, if the patient chooses to transplant, she receives a lump-sum reward of $r_T(h, \ell)$, which is a function of both the patient's health status and the quality of the liver being offered.

The final component of the MDP model is the transition probabilities. When the patient chooses the transplant action, she transitions to the transplanted state, ∇, with probability

one (i.e., $P\{\nabla|s,T\}=1$ for all $s\in\mathcal{S}$ such that $T\in\mathcal{A}_s$). If, on the other hand, the patient chooses to wait, then the transition probabilities are defined as

$$P\{s'|s=(h,\ell,k),W\}=\begin{cases} p(h',\ell',k'|h,\ell,k) & \text{if } s'=(h',\ell',k'),\\ 1-\sum_{s'\in\mathcal{S}'}p(h',\ell',k'|h,\ell,k) & \text{if } s'=\Delta,\\ 0 & \text{if } s'=\nabla. \end{cases}$$

We assume the transition probabilities and rewards are stationary. We further assume that $p(h',\ell',k'|h,\ell,k)=\mathcal{H}\{h'|h\}\cdot\mathcal{K}\{k'|k\}\cdot\mathcal{L}\{\ell'|k'\}$ for all $h,h'\in\Omega,k,k'\in\Psi$, and $\ell,\ell'\in\Phi$, where $\mathcal{H}\{h'|h\}$ is the probability that the patient's health status will be h' at time $t+1$ given that her health at time t is h, $\mathcal{K}\{k'|k\}$ is the probability that the patient's rank will be k' at time $t+1$ given that her rank at time t is k, and $\mathcal{L}\{\ell'|k'\}$ is the probability that the patient will be offered an organ of quality ℓ' at time $t+1$ given that her rank at time $t+1$ is k'. We assume the transitions among health states and the transitions among rank states are independent. Admittedly, the patient's rank at time $t+1$ depends on her health at time $t+1$. However, for analytical and computational tractability, we choose to include the dependency on her rank at time t only. Since patient rank is the primary indicator of patient health, an additional dependency on health would not change the values of these probabilities significantly. Finally, we define the *rank transition probability matrix*, $\mathcal{K}$, as $\mathcal{K}=[\mathcal{K}\{k'|k\}]_{\forall k,k'\in\Psi}$, the *health probability matrix*, $\mathcal{H}$, as $\mathcal{H}=[\mathcal{H}\{h'|h\}]_{\forall h,h'\in\Omega}$, and the *liver offer probability matrix*, $\mathcal{L}$, as $\mathcal{L}=[\mathcal{L}\{\ell'|k'\}]_{\forall k'\in\Psi,\forall\ell'\in\Phi}$. We emphasize that $\sum_{k'\in\Psi}\mathcal{K}\{k'|k\}=\sum_{\ell'\in\Phi}\mathcal{L}\{\ell'|k'\}=1$ for all $k,k'\in\Psi$. We interpret $1-\sum_{h'\in\Omega}\mathcal{H}\{h'|h\}$ as the probability of dying when the patient's health is h.

Given discount rate $\lambda\in[0,1]$, the optimal solution to this problem can be obtained by solving the Bellman optimality equations [112]:

$$v(h,L+1,k)=r_W(h)+\lambda\sum_{(h',\ell',k')}\mathcal{H}\{h'|h\}\mathcal{K}\{k'|k\}\mathcal{L}\{\ell'|k'\}v(h',\ell',k')\quad\forall h\in\Omega,k\in\Psi,\quad(3.1)$$

and

$$v(h,\ell,k)=\max\Big\{r_T(h,\ell),r_W(h)+\lambda\sum_{(h',\ell',k')}\mathcal{H}\{h'|h\}\mathcal{K}\{k'|k\}\mathcal{L}\{\ell'|k'\}v(h',\ell',k')\Big\}$$
$$\forall h\in\Omega,k\in\Psi,\ell\in\Phi\setminus\{L+1\}.\quad(3.2)$$

The value associated with the absorbing states of death and post-transplant, Δ and ∇, is zero by construction and these states are, therefore, excluded from equations (3.1) and (3.2).

3.2 *CCD* MATRICES

In this section, we introduce a new class of stochastic matrices, termed Column-wise Concave with the maximum element of each column on the Diagonal (*CCD*) matrices, which facilitates the derivation of structural results associated with the rank component (k) of the MDP model given in Section 3.1. We start with its definition.

Definition 3.1. *An $n \times n$ stochastic matrix $\mathcal{P}$ is called CCD (Column-wise Concave with the maximum element of each column on the Diagonal) if, for $i = 1, \ldots, n-1$, it satisfies*

$$(i) \quad \mathcal{P}\{j|i\} \geq \mathcal{P}\{j|i+1\} \text{ for } j = 1, 2, \ldots, i, \tag{3.3}$$

$$(ii) \quad \mathcal{P}\{j|i\} \leq \mathcal{P}\{j|i+1\} \text{ for } j = i+1, \ldots, n. \tag{3.4}$$

This definition implies that within each column of a *CCD* matrix, the values are non-decreasing up to (and including) the diagonal element, and nonincreasing after the diagonal element. As a consequence, a necessary (but not sufficient) condition for a stochastic matrix to be *CCD* is to have the maximum value within each column at the diagonal entry.

In the context of our MDP model, a *CCD* rank transition probability matrix implies that the likelihood that a patient with rank i moves to a better rank $j_1 < i$ in the next period is at least as large as the likelihood of moving to the same rank j_1 from a rank that is further down the list. On the other hand, the likelihood that a patient with rank i moves to a worse rank $j_2 > i$ in the next period is no more than the likelihood of moving to the same rank j_2 if the patient is further down the list than i.

Next, we derive several results for *CCD* matrices, which will be used in proving the structural properties of our MDP model. Proposition 3.1 presents inequalities for the row differences of a *CCD* matrix, which are used in proving Theorem 3.3.

Proposition 3.1. *Let $f : \mathbb{R} \to \mathbb{R}_+$ and $g : \mathbb{R} \to \mathbb{R}$ be two functions. If $g(\cdot)$ is nonincreasing and $\mathcal{P}$ is a CCD transition probability matrix, then the following hold*

$$(i) \quad \sum_{j \le k} [\mathcal{P}\{j|k\} - \mathcal{P}\{j|k+1\}] f(j) g(j) \ge g(k) \cdot \sum_{j \le k} [\mathcal{P}\{j|k\} - \mathcal{P}\{j|k+1\}] f(j), \tag{3.5}$$

$$(ii) \quad \sum_{j > k} [\mathcal{P}\{j|k\} - \mathcal{P}\{j|k+1\}] f(j) g(j) \ge g(k) \cdot \sum_{j > k} [\mathcal{P}\{j|k\} - \mathcal{P}\{j|k+1\}] f(j). \tag{3.6}$$

Proof. (i) If $\mathcal{P}$ is a *CCD* matrix, then the term in the square brackets on the left-hand side of inequality (3.5) is nonnegative by definition (see equation (3.3)). Thus the coefficient of $g(j)$ is nonnegative for all j on the left-hand-side of (3.5), since we are given nonnegative $f(j)$. Furthermore, since $g(j) \ge g(k)$ for any $j \le k$, we can write:

$$\sum_{j \le k} [\mathcal{P}\{j|k\} - \mathcal{P}\{j|k+1\}] f(j) g(j) \ge g(k) \sum_{j \le k} [\mathcal{P}\{j|k\} - \mathcal{P}\{j|k+1\}] f(j),$$

which establishes the result.

(ii) Similar to the proof in case (i). □

Proposition 3.2 states that, for a given set of nonincreasing weights, the nonnegative linear combination of the values obtained by taking the difference of any two successive rows of a *CCD* matrix is always nonnegative.

Proposition 3.2. *Let $\{z_j\}_{j=1}^{\infty}$ be a given sequence of nonnegative and nonincreasing numbers. If $\mathcal{P}$ is a CCD transition probability matrix, then for any given k*

$$\sum_{j} [\mathcal{P}\{j|k\} - \mathcal{P}\{j|k+1\}] z_j \ge 0. \tag{3.7}$$

Proof. Starting with the left-hand side of inequality (3.7), we find

$$\begin{aligned}
\sum_{j} [\mathcal{P}\{j|k\} - \mathcal{P}\{j|k+1\}] z_j &= \sum_{j \le k} [\mathcal{P}\{j|k\} - \mathcal{P}\{j|k+1\}] z_j + \sum_{j > k} [\mathcal{P}\{j|k\} - \mathcal{P}\{j|k+1\}] z_j \\
&\ge \sum_{j \le k} [\mathcal{P}\{j|k\} - \mathcal{P}\{j|k+1\}] z_k + \sum_{j > k} [\mathcal{P}\{j|k\} - \mathcal{P}\{j|k+1\}] z_j \\
&\ge \sum_{j \le k} [\mathcal{P}\{j|k\} - \mathcal{P}\{j|k+1\}] z_k + \sum_{j > k} [\mathcal{P}\{j|k\} - \mathcal{P}\{j|k+1\}] z_k \\
&= z_k \Big\{ \sum_{j} [\mathcal{P}\{j|k\} - \mathcal{P}\{j|k+1\}] \Big\}
\end{aligned}$$

$$
\begin{aligned}
&= z_k[1-1] \\
&= 0,
\end{aligned}
$$

which establishes the result. □

An alternative proof of Proposition 3.2 follows from the fact that *CCD* matrices have the Increasing Failure Rate (*IFR*) property (see Proposition 3.3 below) and Lemma 4.7.2 of Puterman [112].

Finally, we show how the *CCD* class relates to other well-known matrix classes in the literature, namely, *IFR* matrices and TP_2 matrices. We start by recalling the definitions of *IFR* matrices and TP_2 matrices.

Definition 3.2 (Barlow and Proschan [13]). *An $n \times n$ stochastic matrix P is called IFR (Increasing Failure Rate) if, for $i = 1, \ldots, n-1$,*

$$
\sum_{j=k}^{n} P\{j|i\} \leq \sum_{j=k}^{n} P\{j|i+1\} \qquad \text{for } k = 1, \ldots, n.
$$

Definition 3.3 (Karlin [82]). *An $n \times n$ stochastic matrix P is called TP_2 (Totally Positive of order 2) if the determinants of all 2×2 submatrices of P are nonnegative.*

IFR and totally positive matrices have been extensively studied. We know that a TP_2 matrix is also *IFR* [82] and are interested in finding any relationship between the *CCD* class and these two classes. Proposition 3.3 states that every *CCD* matrix is also an *IFR* matrix.

Proposition 3.3. *If P is CCD, then it is also IFR.*

Proof. Let P be an $n \times n$ stochastic *CCD* matrix and pick an arbitrary index $i \in \{1, \ldots, n\}$.

Initially consider the first set of inequalities, (3.3), for the given i that must hold for a *CCD* matrix. Summing these inequalities for $j = 1$ to i_2, where $i_2 = 1, \ldots, i$, we obtain

$$
\begin{aligned}
&\sum_{j=1}^{i_2} P\{j|i\} \geq \sum_{j=1}^{i_2} P\{j|i+1\} \qquad \text{for } i_2 = 1, \ldots, i, \\
\Rightarrow \quad &1 - \sum_{j=i_2+1}^{n} P\{j|i\} \geq 1 - \sum_{j=i_2+1}^{n} P\{j|i+1\} \qquad \text{for } i_2 = 1, \ldots, i,
\end{aligned}
$$

$$\Rightarrow \quad \sum_{j=i_2+1}^{n} P\{j|i\} \leq \sum_{j=i_2+1}^{n} P\{j|i+1\} \qquad \text{for } i_2 = 1, \ldots, i,$$

$$\Rightarrow \quad \sum_{j=k}^{n} P\{j|i\} \leq \sum_{j=k}^{n} P\{j|i+1\} \qquad \text{for } k = 2, \ldots, i+1. \tag{3.8}$$

Now consider the second set of inequalities, (3.4), for the given i that must hold for a *CCD* matrix. Summing these inequalities for i_1 through $j = n$, where $i_1 = i+2, \ldots, n$, we obtain

$$\sum_{j=i_1}^{n} P\{j|i\} \leq \sum_{j=i_1}^{n} P\{j|i+1\} \qquad \text{for } i_1 = i+2, \ldots, n. \tag{3.9}$$

Finally, combining results (3.8) and (3.9), we obtain

$$\sum_{j=k}^{n} P\{j|i\} \leq \sum_{j=k}^{n} P\{j|i+1\} \qquad \text{for } k = 2, \ldots, n. \tag{3.10}$$

In addition, since both sides of the inequality in (3.10) are equal to 1 when $k = 1$, we conclude that the inequality holds for $k = 1, \ldots, n$, which is the definition of an *IFR* matrix. □

Remark 3.1 states that not every *IFR* matrix is *CCD*.

Remark 3.1. *If P is IFR, then it is not necessarily CCD as indicated by the following example. Consider the following stochastic matrix.*

$$P = \begin{bmatrix} 0.8 & 0.1 & 0.1 \\ 0.7 & 0.1 & 0.2 \\ 0.5 & 0.2 & 0.3 \end{bmatrix}.$$

P is IFR, but it is not CCD because the maximum element of the second column does not appear on the diagonal entry.

Having proved that *CCD* is a stronger condition than *IFR*, we now investigate how *CCD* is related to TP_2. Proposition 3.4 shows that *CCD* and TP_2 conditions are equivalent for 2×2 matrices. However, the answer remains ambiguous for matrices of larger size, as indicated in Remark 3.2.

Proposition 3.4. *For a 2×2 stochastic matrix P, P is CCD $\Leftrightarrow$ P is TP_2 $\Leftrightarrow$ P is IFR.*

Proof. Let P be given as

$$P = \begin{bmatrix} a & 1-a \\ b & 1-b \end{bmatrix},$$

where $0 \leq a, b \leq 1$. First, assume P is *CCD*. Then $a \geq b$ must hold. Therefore, $|P| = a(1-b) - b(1-a) = a - b \geq 0$, which implies P is TP_2. Next, assume P is TP_2. Then $|P| = a(1-b) - b(1-a) = a - b \geq 0$ must hold. Therefore, $a \geq b$, which implies P is *CCD*. This completes the proof for the equivalence of the *CCD* and TP_2 conditions.

Finally, assume P is *IFR*. Then $1-a \leq 1-b$ must hold. Therefore, $a \geq b$, which implies P is *CCD*. Combining this result with that of Proposition 3.3 completes the proof for the equivalence of the *CCD* and *IFR* conditions. □

Remark 3.2. *For $n > 2$, if P is CCD (TP_2), then it is not necessarily TP_2 (CCD) as indicated by the following examples.*

First, it is easily verified that the following matrix is CCD

$$P = \begin{bmatrix} 0.9 & 0.1 & 0.0 \\ 0.6 & 0.2 & 0.1 \\ 0.4 & 0.1 & 0.5 \end{bmatrix};$$

however, it is not TP_2, since the lower left 2×2 submatrix has a negative determinant.

Second, the following TP_2 matrix is not CCD, since the maximum element in the second column is not on the diagonal.

$$P = \begin{bmatrix} 0.4 & 0.4 & 0.2 \\ 0.3 & 0.3 & 0.4 \\ 0.2 & 0.3 & 0.5 \end{bmatrix}.$$

3.3 STRUCTURAL RESULTS

In this section, we establish several structural properties of the MDP model formulated in Section 3.1. Specifically, we identify conditions on the parameters that guarantee structured value functions and optimal policies. In addition to their analytical elegance, such results may provide deeper insight into the overall problem and help devise computationally faster solution approaches.

The following assumptions hold throughout this section:

(AS1) $r_W(h)$ is nonincreasing in h;

(AS2) $r_T(h, \ell)$ is nonincreasing in both h and ℓ.

AS1 implies that the intermediate reward of waiting does not increase as the patient deteriorates. Similarly, AS2 implies that the post-transplant reward does not increase as the patient deteriorates and/or the quality of the liver degrades.

Theorem 3.1 establishes the intuitive fact that it is always better to be offered a higher quality organ.

Theorem 3.1. *$v(h, \ell, k)$ is monotonically nonincreasing in ℓ for any $h \in \Omega$ and $k \in \Psi$.*

Proof. For any given h, k, and l, consider two possible cases. If $a(h, k, l+1) = T$, then $v(h, k, l+1) = r_T(h, l+1) \leq r_T(h, l) \leq v(h, k, l)$ by assumption AS2 and equation (3.2). If, on the other hand, $a(h, k, l+1) = W$, then

$$\begin{aligned} v(h, k, l+1) &= r_W(h) + \lambda \sum_{h'} \sum_{k'} \sum_{l'} \mathcal{H}\{h'|h\} \cdot \mathcal{K}\{k'|k\} \cdot \mathcal{L}\{l'|k'\} \cdot v(h', k', l') \\ &\leq v(h, k, l), \end{aligned}$$

where the last inequality follows from equation (3.2). This completes the proof. □

Similar to the definition given in [6], we define a *liver-based control-limit optimal policy* to be a policy among the optimal policies that, for a given health state, h, and rank, k, distinguishes a critical liver state ℓ^* and prescribes 'transplant' for all livers $\ell \leq \ell^*$ and 'wait' for all livers $\ell > \ell^*$. Theorem 3.2 shows that assumptions AS1 and AS2 are sufficient for the existence of a liver-based control-limit optimal policy.

Theorem 3.2. *There exists a liver-based control-limit optimal policy for all $h \in \Omega$ and $k \in \Psi$.*

Proof. For any given liver quality $\ell < L$, it suffices to prove that if $a^*(h, \ell+1, k) = T$, then $a^*(h, \ell, k) = T$ for all $h \in \Omega$ and $k \in \Psi$. For any given $h \in \Omega$ and $k \in \Psi$, if $a^*(h, \ell+1, k) = T$, then

$$\begin{aligned} v(h, \ell+1, k) &= r_T(h, \ell+1) \\ &\geq r_W(h) + \lambda \sum_{h'} \sum_{k'} \sum_{\ell'} \mathcal{H}\{h'|h\} \cdot \mathcal{K}\{k'|k\} \cdot \mathcal{L}\{\ell'|k'\} \cdot v(h', k', \ell'). \end{aligned}$$

Since $v(h, \ell, k) = \max\{r_T(h, \ell), r_W(h) + \lambda \sum_{h'} \sum_{k'} \sum_{\ell'} \mathcal{H}\{h'|h\}\mathcal{K}\{k'|k\}\mathcal{L}\{\ell'|k'\}v(h', k', \ell')\}$ and AS2 implies $r_T(h, \ell) \geq r_T(h, \ell+1)$ for $\ell = 1, \ldots, L-1$, we find $v(h, \ell, k) = r_T(h, \ell)$. Therefore, $a^*(h, \ell, k) = T$ for any $h \in \Omega$ and $k \in \Psi$. □

An immediate result of Proposition 3.1 is Corollary 3.1.

Corollary 3.1. *If $\mathcal{K}$ is a CCD matrix and $v(h, \ell, k)$ is nonincreasing in k for any $h \in \Omega$ and $\ell \in \Phi$, then the following hold*

$$\begin{aligned} (i) \sum_{\ell'} \sum_{k' \leq k} [\mathcal{K}\{k'|k\} - \mathcal{K}\{k'|k+1\}]\mathcal{L}\{\ell'|k'\}v(h', \ell', k') \\ \geq \sum_{\ell'} v(h', \ell', k) \sum_{k' \leq k} [\mathcal{K}\{k'|k\} - \mathcal{K}\{k'|k+1\}]\mathcal{L}\{\ell'|k'\}, \\ (ii) \sum_{\ell'} \sum_{k' > k} [\mathcal{K}\{k'|k\} - \mathcal{K}\{k'|k+1\}]\mathcal{L}\{\ell'|k'\}v(h', \ell', k') \\ \geq \sum_{\ell'} v(h', \ell', k) \sum_{k' > k} [\mathcal{K}\{k'|k\} - \mathcal{K}\{k'|k+1\}]\mathcal{L}\{\ell'|k'\}. \end{aligned}$$

Theorem 3.3 states that a patient's maximum expected total discounted reward does not increase as her rank deteriorates. The condition on $\mathcal{L}$ simply states that patient's chance of receiving a liver offer does not increase as her rank deteriorates.

Theorem 3.3. *If $\mathcal{K}$ is CCD and $\mathcal{L}\{\ell|k\}$ is monotonically nonincreasing in $k \in \Psi$ for all $\ell \neq L+1$, then $v(h, \ell, k)$ is monotonically nonincreasing in $k \in \Psi$ for any $h \in \Omega$ and $\ell \in \Phi$.*

Proof. (by induction on the steps of the value iteration algorithm)

We prove the theorem for $\ell \neq L+1$ and note that the proof for $\ell = L+1$ follows similarly.

If we can show that the value functions at each iteration of the value iteration algorithm are monotonically nonincreasing in $k \in \Psi$ for given $h \in \Omega$ and $\ell \in \Phi$, then the result holds by the convergence of value iteration. Let $v^i(h, \ell, k)$ be the value associated with state $(h, \ell, k) \in \mathcal{S}$ at the ith iteration of the value iteration algorithm and assume, without loss of generality, that the algorithm starts with a value of 0 for each state, i.e., $v^0(h, \ell, k) = 0$ for all $(h, \ell, k) \in \mathcal{S}$.

It is clear that $v^1(h, \ell, k)$ is constant and therefore nonincreasing in $k \in \Psi$ for all $h \in \Omega$ and $\ell \in \Phi$.

Next, assume, as the induction hypothesis, that for a given $h \in \Omega$ and $\ell \in \Phi$, $v^i(h, \ell, k) \geq v^i(h, \ell, k+1)$ for all $k \in \Psi$ for iterations $i = 2, \ldots, n$.

By equation (3.2)

$$v^{n+1}(h, \ell, k) = \max\Big\{r_T(h, \ell); \\ r_W(h) + \lambda \sum_{h'} \sum_{k'} \sum_{\ell'} \mathcal{H}\{h'|h\}\mathcal{K}\{k'|k\}\mathcal{L}\{\ell'|k'\}v^n(h', \ell', k')\Big\} \quad (3.11)$$

and

$$v^{n+1}(h, \ell, k+1) = \max\Big\{r_T(h, \ell); \\ r_W(h) + \lambda \sum_{h'} \sum_{k'} \sum_{\ell'} \mathcal{H}\{h'|h\}\mathcal{K}\{k'|k+1\}\mathcal{L}\{\ell'|k'\}v^n(h', \ell', k')\Big\}. \quad (3.12)$$

If $a^{n+1}(h, \ell, k+1) = T$, then $v^{n+1}(h, \ell, k+1) = r_T(h, \ell) \leq v^{n+1}(h, \ell, k)$.

If $a^{n+1}(h, \ell, k+1) = W$, then by equations (3.11) and (3.12)

$$\begin{aligned} &v^{n+1}(h, \ell, k) - v^{n+1}(h, \ell, k+1) \\ &\quad \geq \lambda \sum_{h'} \mathcal{H}\{h'|h\}\Big\{\sum_{k'} \sum_{\ell'} [\mathcal{K}\{k'|k\} - \mathcal{K}\{k'|k+1\}]\mathcal{L}\{\ell'|k'\}v^n(h', \ell', k')\Big\} \\ &\quad = \lambda \sum_{h'} \mathcal{H}\{h'|h\}\Big\{\sum_{\ell'} \sum_{k' \leq k} [\mathcal{K}\{k'|k\} - \mathcal{K}\{k'|k+1\}]\mathcal{L}\{\ell'|k'\}v^n(h', \ell', k') \\ &\qquad\qquad + \sum_{\ell'} \sum_{k' > k} [\mathcal{K}\{k'|k\} - \mathcal{K}\{k'|k+1\}]\mathcal{L}\{\ell'|k'\}v^n(h', \ell', k')\Big\}. \end{aligned} \quad (3.13)$$

Since $v^n(h', \ell', k')$ is nonincreasing in $k' \in \Psi$ for all $h' \in \Omega$ and $\ell' \in \Phi$, by the induction hypothesis, and the fact that $\mathcal{K}$ is *CCD*, Corollary 3.1 implies that inequality (3.13) is preserved if we replace $v^n(h', \ell', k')$ by $v^n(h', \ell', k)$ for all (h', ℓ', k'). Making this substitution and rearranging yields

$$\begin{aligned} & v^{n+1}(h, \ell, k) - v^{n+1}(h, \ell, k+1) \\ & \quad \geq \lambda \sum_{h'} \mathcal{H}\{h'|h\} \Big\{ \sum_{\ell' \neq L+1} v^n(h', \ell', k) \sum_{k'} [\mathcal{K}\{k'|k\} - \mathcal{K}\{k'|k+1\}] \mathcal{L}\{\ell'|k'\} \\ & \qquad\qquad + v^n(h', L+1, k) \sum_{k'} [\mathcal{K}\{k'|k\} - \mathcal{K}\{k'|k+1\}] \mathcal{L}\{L+1|k'\} \Big\}. \end{aligned}$$

Employing the identity $\mathcal{L}\{L+1|k'\} = 1 - \sum_{\ell' \neq L+1} \mathcal{L}\{\ell'|k'\}$ yields

$$\begin{aligned} & v^{n+1}(h, \ell, k) - v^{n+1}(h, \ell, k+1) \\ & \quad \geq \lambda \sum_{h'} \mathcal{H}\{h'|h\} \Big\{ \sum_{\ell' \neq L+1} v^n(h', \ell', k) \cdot \sum_{k'} [\mathcal{K}\{k'|k\} - \mathcal{K}\{k'|k+1\}] \mathcal{L}\{\ell'|k'\} \\ & \qquad\qquad - v^n(h', L+1, k) \sum_{k'} [\mathcal{K}\{k'|k\} - \mathcal{K}\{k'|k+1\}] \sum_{\ell' \neq L+1} \mathcal{L}\{\ell'|k'\} \\ & \qquad\qquad + v^n(h', L+1, k) \sum_{k'} [\mathcal{K}\{k'|k\} - \mathcal{K}\{k'|k+1\}] \Big\}. \end{aligned}$$

Eliminating the final term in the right-hand-side of the last inequality since it is always 0 and rearranging yields

$$\begin{aligned} & v^{n+1}(h, \ell, k) - v^{n+1}(h, \ell, k+1) \\ & \geq \lambda \sum_{h'} \mathcal{H}\{h'|h\} \Big\{ \sum_{\ell' \neq L+1} \Big[v^n(h', \ell', k) - v^n(h', L+1, k) \Big] \Big[\sum_{k'} [\mathcal{K}\{k'|k\} - \mathcal{K}\{k'|k+1\}] \mathcal{L}\{\ell'|k'\} \Big] \Big\}. \end{aligned}$$

Now $\lambda \geq 0$, $\mathcal{H}\{h'|h\} \geq 0$ for all $h, h' \in \Omega$, and $v^n(h', \ell', k) - v^n(h', L+1, k) \geq 0$ for all $h' \in \Omega, \ell' \in \Phi$, and $k \in \Psi$, by Proposition 3.1. Furthermore, since $\mathcal{K}$ is *CCD* and $\mathcal{L}$ is nonincreasing in $k' \in \Psi$ for fixed $\ell' \neq L+1$, Proposition 3.2 implies that $\sum_{k'} [\mathcal{K}\{k'|k\} - \mathcal{K}\{k'|k+1\}] \mathcal{L}\{\ell'|k'\} \geq 0$ for all $\ell' \neq L+1$. Therefore, $v^{n+1}(h, \ell, k) - v^{n+1}(h, \ell, k+1) \geq 0$, which completes the proof. □

We need the following technical result, before we present our main result concerning the structure of the optimal policy.

Lemma 3.1. *Let* $\varphi(k^{'}) = \sum_{h^{'}} \sum_{l^{'}} \mathcal{H}\{h^{'}|h\}\mathcal{L}\{l^{'}|k^{'}\}v^{n}(h^{'}, k^{'}, l^{'})$. *If* $\mathcal{K}$ *is CCD and* $\mathcal{L}\{l|k\}$ *is monotonically nonincreasing in* k, *then* $\varphi(k)$ *is also nonincreasing in* k.

Proof. For any $k' \in \Psi$,

$$
\begin{aligned}
&\varphi(k^{'}) - \varphi(k^{'}+1) \\
&= \sum_{h^{'}} \sum_{l^{'}} \mathcal{H}\{h^{'}|h\}\mathcal{L}\{l^{'}|k^{'}\}v(h^{'}, k^{'}, l^{'}) - \sum_{h^{'}} \sum_{l^{'}} \mathcal{H}\{h^{'}|h\}\mathcal{L}\{l^{'}|k^{'}+1\}v(h^{'}, k^{'}+1, l^{'}) \\
&= \sum_{h^{'}} \Bigg\{ \mathcal{H}\{h^{'}|h\} \sum_{l^{'}} \Big[\mathcal{L}\{l^{'}|k^{'}\}v(h^{'}, k^{'}, l^{'}) - \mathcal{L}\{l^{'}|k^{'}+1\}v(h^{'}, k^{'}+1, l^{'}) \Big] \Bigg\} \\
&= \sum_{h^{'}} \Bigg\{ \mathcal{H}\{h^{'}|h\} \Bigg[\sum_{l^{'} \neq L+1} \Big[\mathcal{L}\{l^{'}|k^{'}\}v(h^{'}, k^{'}, l^{'}) - \mathcal{L}\{l^{'}|k^{'}+1\}v(h^{'}, k^{'}+1, l^{'}) \Big] \\
&\qquad + \mathcal{L}\{L+1|k^{'}\}v(h^{'}, k^{'}, L+1) - \mathcal{L}\{L+1|k^{'}+1\}v(h^{'}, k^{'}+1, L+1) \Bigg] \Bigg\} \\
&\geq \sum_{h^{'}} \Bigg\{ \mathcal{H}\{h^{'}|h\} \Bigg[\sum_{l^{'} \neq L+1} \Big[\mathcal{L}\{l^{'}|k^{'}\}v(h^{'}, k^{'}, l^{'}) - \mathcal{L}\{l^{'}|k^{'}+1\}v(h^{'}, k^{'}, l^{'}) \Big] \\
&\qquad + \mathcal{L}\{L+1|k^{'}\}v(h^{'}, k^{'}, L+1) - \mathcal{L}\{L+1|k^{'}+1\}v(h^{'}, k^{'}, L+1) \Bigg] \Bigg\}
\end{aligned}
$$

where the last inequality follows since $v(h^{'}, k^{'}+1, l^{'}) \leq v(h^{'}, k^{'}, l^{'})$ by Theorem 3.3 and $\mathcal{L}\{l^{'}|k^{'}+1\}$ is nonnegative. Similarly, $\mathcal{L}\{L+1|k^{'}+1\}$ is nonnegative and $v(h^{'}, k^{'}+1, L+1) \leq v(h^{'}, k^{'}, L+1)$ by Theorem 3.3, therefore we can write

$$
\begin{aligned}
&\varphi(k^{'}) - \varphi(k^{'}+1) \\
&\geq \sum_{h^{'}} \Bigg\{ \mathcal{H}\{h^{'}|h\} \Bigg[\sum_{l^{'} \neq L+1} \Big[\mathcal{L}\{l^{'}|k^{'}\}v(h^{'}, k^{'}, l^{'}) - \mathcal{L}\{l^{'}|k^{'}+1\}v(h^{'}, k^{'}, l^{'}) \Big] \\
&\qquad + \mathcal{L}\{L+1|k^{'}\}v(h^{'}, k^{'}, L+1) - \mathcal{L}\{L+1|k^{'}+1\}v(h^{'}, k^{'}, L+1) \Bigg] \Bigg\} \\
&= \sum_{h^{'}} \Bigg\{ \mathcal{H}\{h^{'}|h\} \Bigg[\sum_{l^{'} \neq L+1} v(h^{'}, k^{'}, l^{'}) \Big[\mathcal{L}\{l^{'}|k^{'}\} - \mathcal{L}\{l^{'}|k^{'}+1\} \Big] \\
&\qquad + v(h^{'}, k^{'}, L+1) \Big[\mathcal{L}\{L+1|k^{'}\} - \mathcal{L}\{L+1|k^{'}+1\} \Big] \Bigg] \Bigg\}
\end{aligned}
$$

$$= \sum_{h'} \Bigg\{ \mathcal{H}\{h'|h\} \Bigg[\sum_{l' \neq L+1} v(h', k', l') \Big[\mathcal{L}\{l'|k'\} - \mathcal{L}\{l'|k'+1\} \Big] + v(h', k', L+1) \Big[1 - \sum_{l' \neq L+1} \mathcal{L}\{l'|k'\} - 1 + \sum_{l' \neq L+1} \mathcal{L}\{l'|k'+1\} \Big] \Bigg] \Bigg\}$$

$$= \sum_{h'} \Bigg\{ \mathcal{H}\{h'|h\} \Bigg[\sum_{l' \neq L+1} \Big[\mathcal{L}\{l'|k'\} - \mathcal{L}\{l'|k'+1\} \Big] \Big[v(h', k', l') - v(h', k', L+1) \Big] \Bigg] \Bigg\}$$

Finally, $\mathcal{H}\{h'|h\}$ is nonnegative for all h' by definition; $\mathcal{L}\{l'|k'\} \geq \mathcal{L}\{l'|k'+1\}$ for all $l' \neq L+1$ and k' by assumption ; and $v(h', k', l') \geq v(h', k', L+1)$ for all l' by Theorem 3.1. Therefore, $\varphi(k') - \varphi(k'+1) \geq 0$ and the proof is completed. □

The main result on the structure of the optimal policy is given in Theorem 3.4, which establishes conditions under which there exists a rank-based control-limit optimal policy. Analogous to a liver-based control-limit optimal policy, a *rank-based control-limit optimal policy* is an optimal policy that prescribes, for a given health state, h, and a liver quality, ℓ, 'wait' if the rank of the patient is below some threshold rank k^* and 'transplant' for all ranks greater than k^*.

Theorem 3.4. *If $\mathcal{K}$ is CCD and $\mathcal{L}\{\ell|k\}$ is monotonically nonincreasing in $k \in \Psi$ for all $\ell \neq L+1$, then there exists a rank-based control-limit optimal policy.*

Proof. (by contradiction) An alternative proof of this theorem can be found in Sandıkçı et al. [124].

We want to prove that given h and l, $a(h,k,l) = T$ for any k implies that $a(h,k',l) = T$ for all $k' \geq k$. Consider any rank $k \in \Psi$ and assume otherwise, i.e., assume $a(h,k,l) = T$ but $a(h,k+1,l) = W$ uniquely. This, respectively, implies that

$$\begin{aligned} v(h,k,l) &= r_T(h,l) \\ &\geq r_W(h) + \lambda \sum_{h'} \sum_{k'} \sum_{l'} \mathcal{H}\{h'|h\} \cdot \mathcal{K}\{k'|k\} \cdot \mathcal{L}\{l'|k'\} \cdot v(h', k', l') \end{aligned}$$

and

$$\begin{aligned} v(h,k+1,l) &= r_W(h) + \lambda \sum_{h'} \sum_{k'} \sum_{l'} \mathcal{H}\{h'|h\} \cdot \mathcal{K}\{k'|k+1\} \cdot \mathcal{L}\{l'|k'\} \cdot v(h', k', l') \\ &> r_T(h,l). \end{aligned}$$

Then these two together imply that

$$
\begin{aligned}
& r_T(h,l) - r_T(h,l) \\
& \quad > \lambda \Bigg\{ \sum_{h'} \sum_{k'} \sum_{l'} \mathcal{H}\{h'|h\} \mathcal{K}\{k'|k\} \mathcal{L}\{l'|k'\} v(h',k',l') \\
& \qquad\qquad - \sum_{h'} \sum_{k'} \sum_{l'} \mathcal{H}\{h'|h\} \mathcal{K}\{k'|k+1\} \mathcal{L}\{l'|k'\} v(h',k',l') \Bigg\} \\
& \quad = \lambda \Bigg\{ \sum_{k'} \mathcal{K}\{k'|k\} \left[\sum_{h'} \sum_{l'} \mathcal{H}\{h'|h\} \mathcal{L}\{l'|k'\} v(h',k',l') \right] \\
& \qquad\qquad - \sum_{k'} \mathcal{K}\{k'|k+1\} \left[\sum_{h'} \sum_{l'} \mathcal{H}\{h'|h\} \mathcal{L}\{l'|k'\} v(h',k',l') \right] \Bigg\} \\
& \quad = \lambda \Bigg\{ \sum_{k'} \mathcal{K}\{k'|k\} \cdot \varphi(k') - \sum_{k'} \mathcal{K}\{k'|k+1\} \cdot \varphi(k') \Bigg\} \\
& \quad = \lambda \Bigg\{ \sum_{k' \le k} \mathcal{K}\{k'|k\} \cdot \varphi(k') + \sum_{k' > k} \mathcal{K}\{k'|k\} \cdot \varphi(k') - \sum_{k' \le k} \mathcal{K}\{k'|k+1\} \cdot \varphi(k') \\
& \qquad\qquad - \sum_{k' > k} \mathcal{K}\{k'|k+1\} \cdot \varphi(k') \Bigg\} \\
& \quad = \lambda \Bigg\{ \sum_{k' \le k} \Big[\mathcal{K}\{k'|k\} - \mathcal{K}\{k'|k+1\} \Big] \varphi(k') + \sum_{k' > k} \Big[\mathcal{K}\{k'|k\} - \mathcal{K}\{k'|k+1\} \Big] \varphi(k') \Bigg\} \qquad (3.14)
\end{aligned}
$$

Now, since $\mathcal{K}$ is CCD by assumption and $\varphi(k)$ is nonincreasing in k by Lemma 3.1, Proposition 3.1 applies and implies that if $\varphi(k')$ is replaced by $\varphi(k)$ for all k' in the right-hand-side of inequality (3.14), then the result will be no bigger than the original. That is,

$$
\begin{aligned}
0 &= r_T(h,l) - r_T(h,l) \\
&> \lambda \Bigg\{ \varphi(k) \sum_{k' \le k} \Big[\mathcal{K}\{k'|k\} - \mathcal{K}\{k'|k+1\} \Big] + \varphi(k) \sum_{k' > k} \Big[\mathcal{K}\{k'|k\} - \mathcal{K}\{k'|k+1\} \Big] \Bigg\} \\
&= \lambda \varphi(k) \sum_{k'} \Big[\mathcal{K}\{k'|k\} - \mathcal{K}\{k'|k+1\} \Big] \\
&= 0,
\end{aligned}
$$

which is a contradiction. Therefore, $a(h,k+1,l)$ cannot be 'W'. □

Before we proceed with the results associated with the health component (h) of the MDP model, we provide a technical lemma that is used in the proof of Theorem 3.5. For notational convenience, we denote the dead state, Δ, as $H+1$ in the remainder of this section. Let the $(H+1)\times(H+1)$ *augmented health transition probability matrix* $\hat{\mathcal{H}}$, where the first H states of this matrix represent the health states and the last state represents death, be given by

$$\hat{\mathcal{H}} = \begin{bmatrix} \mathcal{H} & (I-\mathcal{H})e \\ 0 & 1 \end{bmatrix},$$

where I is the $H \times H$ identity matrix and e is an $H \times 1$ vector of ones.

Lemma 3.2. *Let $\hat{\mathcal{H}}$ be an IFR transition probability matrix and $v(h,\ell,k)$ be a nonincreasing function of $h \in \Omega$ for any $k \in \Psi$ and $\ell \in \Phi$. Then the following hold for all $h \in \Omega$*

$$(i) \quad \sum_{h' \le h} \Big\{ \Big[\hat{\mathcal{H}}\{h'|h\} - \hat{\mathcal{H}}\{h'|h+1\}\Big] \Big[\sum_{k'} \sum_{\ell'} \mathcal{K}\{k'|k\}\mathcal{L}\{\ell'|k'\} v(h',\ell',k')\Big]\Big\}$$
$$\ge \Big[\sum_{k'} \sum_{\ell'} \mathcal{K}\{k'|k\}\mathcal{L}\{\ell'|k'\} v(h,\ell',k')\Big] \sum_{h' \le h} \Big[\hat{\mathcal{H}}\{h'|h\} - \hat{\mathcal{H}}\{h'|h+1\}\Big], \tag{3.15}$$

$$(ii) \quad \sum_{h' > h} \Big\{ \Big[\hat{\mathcal{H}}\{h'|h\} - \hat{\mathcal{H}}\{h'|h+1\}\Big] \Big[\sum_{k'} \sum_{\ell'} \mathcal{K}\{k'|k\}\mathcal{L}\{\ell'|k'\} v(h',\ell',k')\Big]\Big\}$$
$$\ge \Big[\sum_{k'} \sum_{\ell'} \mathcal{K}\{k'|k\}\mathcal{L}\{\ell'|k'\} v(h+1,\ell',k')\Big] \sum_{h' > h} \Big[\hat{\mathcal{H}}\{h'|h\} - \hat{\mathcal{H}}\{h'|h+1\}\Big]. \tag{3.16}$$

Proof. (i) Starting with the left-hand-side of inequality (3.15), we can write:

$$\begin{aligned}
&\sum_{h' \le h} \Big\{ \Big[\hat{\mathcal{H}}\{h'|h\} - \hat{\mathcal{H}}\{h'|h+1\}\Big] \Big[\sum_{k'} \sum_{\ell'} \mathcal{K}\{k'|k\}\mathcal{L}\{\ell'|k'\} v(h',\ell',k')\Big]\Big\} \\
&\quad = \Big[\hat{\mathcal{H}}\{1|h\} - \hat{\mathcal{H}}\{1|h+1\}\Big] \Big[\sum_{k'} \sum_{\ell'} \mathcal{K}\{k'|k\}\mathcal{L}\{\ell'|k'\} v(1,\ell',k')\Big] \\
&\qquad + \sum_{h'=2}^{h} \Big\{ \Big[\hat{\mathcal{H}}\{h'|h\} - \hat{\mathcal{H}}\{h'|h+1\}\Big] \Big[\sum_{k'} \sum_{\ell'} \mathcal{K}\{k'|k\}\mathcal{L}\{\ell'|k'\} v(h',\ell',k')\Big]\Big\} \\
&\quad \ge \sum_{h'=1}^{2} \Big[\hat{\mathcal{H}}\{h'|h\} - \hat{\mathcal{H}}\{h'|h+1\}\Big] \Big[\sum_{k'} \sum_{\ell'} \mathcal{K}\{k'|k\}\mathcal{L}\{\ell'|k'\} v(2,\ell',k')\Big] \\
&\qquad + \sum_{h'=3}^{h} \Big\{ \Big[\hat{\mathcal{H}}\{h'|h\} - \hat{\mathcal{H}}\{h'|h+1\}\Big] \Big[\sum_{k'} \sum_{\ell'} \mathcal{K}\{k'|k\}\mathcal{L}\{\ell'|k'\} v(h',\ell',k')\Big]\Big\},
\end{aligned}$$

where the last inequality follows from the fact that $\hat{\mathcal{H}}$ *IFR* implies $\hat{\mathcal{H}}\{1|h\} - \hat{\mathcal{H}}\{1|h+1\} \ge 0$, $\mathcal{K}\{\cdot\}$ and $\mathcal{L}\{\cdot\}$ are nonnegative, and $v(1,\ell',k') \ge v(2,\ell',k')$. Repeating this argument (i.e., $\hat{\mathcal{H}}$

IFR implies $\sum_{h'=1}^{2}[\hat{\mathcal{H}}\{h'|h\}-\hat{\mathcal{H}}\{h'|h+1\}\geq 0$, $\mathcal{K}\{\cdot\}$ and $\mathcal{L}\{\cdot\}$ are nonnegative by definition, and $v(2,\ell',k')\geq v(3,\ell',k'))$ yields:

$$\begin{aligned}
&\sum_{h'\leq h}\Big\{\Big[\hat{\mathcal{H}}\{h'|h\}-\hat{\mathcal{H}}\{h'|h+1\}\Big]\Big[\sum_{k'}\sum_{\ell'}\mathcal{K}\{k'|k\}\mathcal{L}\{\ell'|k'\}v(h',\ell',k')\Big]\Big\}\\
&\qquad\geq\ \Big[\sum_{k'}\sum_{\ell'}\mathcal{K}\{k'|k\}\mathcal{L}\{\ell'|k'\}v(3,\ell',k')\Big]\sum_{h'=1}^{3}\Big[\hat{\mathcal{H}}\{h'|h\}-\hat{\mathcal{H}}\{h'|h+1\}\Big]\\
&\qquad\quad+\sum_{h'=4}^{h}\Big\{\Big[\hat{\mathcal{H}}\{h'|h\}-\hat{\mathcal{H}}\{h'|h+1\}\Big]\Big[\sum_{k'}\sum_{\ell'}\mathcal{K}\{k'|k\}\mathcal{L}\{\ell'|k'\}v(h',\ell',k')\Big]\Big\}.
\end{aligned}$$

Continuing in this manner until $v(h-1,\ell',k')$ is replaced by $v(h,\ell',k')$ establishes the result.

(ii) Similar to the proof in case (i). □

Theorem 3.5 presents the conditions under which the value function is monotone in patient health for fixed rank and liver quality.

Theorem 3.5. *If $\hat{\mathcal{H}}$ is IFR, then $v(h,\ell,k)$ is monotonically nonincreasing in h for any $k\in\Psi$ and $\ell\in\Phi$.*

Proof. (by induction on the steps of the value iteration algorithm)

We prove the theorem for $\ell\neq L+1$ and note that the proof for $\ell=L+1$ follows similarly.

Let $v^i(h,\ell,k)$ be the value associated with state $(h,\ell,k)\in\mathcal{S}$ at the ith iteration of the value iteration algorithm and assume, without loss of generality, that the algorithm starts with a value of 0 for each state, i.e., $v^0(h,\ell,k)=0$ for all $(h,\ell,k)\in\mathcal{S}$. It is clear, by assumptions AS1 and AS2, that the result holds for iteration 1. Given that $v^n(h,\ell,k)\geq v^n(h+1,\ell,k)$, we must show $v^{n+1}(h,\ell,k)\geq v^{n+1}(h+1,\ell,k)$. Since $v(H+1)=0$, by equation (3.2)

$$\begin{aligned}
v^{n+1}(h,\ell,k)=\max\Big\{&r_T(h,\ell);\\
&r_W(h)+\lambda\sum_{h'}\sum_{k'}\sum_{\ell'}\hat{\mathcal{H}}\{h'|h\}\mathcal{K}\{k'|k\}\mathcal{L}\{\ell'|k'\}v^n(h',\ell',k')\Big\}
\end{aligned}\tag{3.17}$$

and

$$v^{n+1}(h+1,\ell,k) = \max\Big\{r_T(h+1,\ell);\\ r_W(h+1) + \lambda \sum_{h'}\sum_{k'}\sum_{\ell'} \hat{\mathcal{H}}\{h'|h+1\}\mathcal{K}\{k'|k\}\mathcal{L}\{\ell'|k'\}v^n(h',\ell',k')\Big\}. \quad (3.18)$$

If $a^{n+1}(h+1,\ell,k) = T$, then $v^{n+1}(h+1,\ell,k) = r_T(h+1,\ell) \leq r_T(h,\ell) \leq v^{n+1}(h,\ell,k)$.

If $a^{n+1}(h+1,\ell,k) = W$, then by equations (3.17) and (3.18),

$$\begin{aligned} & v^{n+1}(h,\ell,k) - v^{n+1}(h+1,\ell,k) \\ \geq\ & r_W(h) - r_W(h+1) + \lambda \sum_{h'} \hat{\mathcal{H}}\{h'|h\}\Big\{\sum_{k'}\sum_{\ell'} \mathcal{K}\{k'|k\}\mathcal{L}\{\ell'|k'\}v^n(h',\ell',k')\Big\} \\ & -\lambda \sum_{h'} \hat{\mathcal{H}}\{h'|h+1\}\Big\{\sum_{k'}\sum_{\ell'} \mathcal{K}\{k'|k\}\mathcal{L}\{\ell'|k'\}v^n(h',\ell',k')\Big\} \\ \geq\ & \lambda \sum_{h'}\Big\{\Big[\hat{\mathcal{H}}\{h'|h\} - \hat{\mathcal{H}}\{h'|h+1\}\Big]\Big[\sum_{k'}\sum_{\ell'} \mathcal{K}\{k'|k\}\mathcal{L}\{\ell'|k'\}v^n(h',\ell',k')\Big]\Big\}. \quad (3.19) \end{aligned}$$

Since $\hat{\mathcal{H}}$ is *IFR* by assumption, $\mathcal{K}\{\cdot\}$ and $\mathcal{L}\{\cdot\}$ are nonnegative by definition, and $v^n(h',\ell',k')$ is nonincreasing in $h' \in \Omega$ for all $k' \in \Psi$ and $\ell' \in \Phi$, Lemma 3.2 implies that inequality (3.19) is preserved if we replace $v^n(h',\ell',k')$ by $v^n(h,\ell',k')$ for all (h',ℓ',k'). Therefore,

$$\begin{aligned} & v^{n+1}(h,\ell,k) - v^{n+1}(h+1,\ell,k) \\ \geq\ & \lambda\Big[\sum_{k'}\sum_{\ell'} \mathcal{K}\{k'|k\}\mathcal{L}\{\ell'|k'\}v^n(h,\ell',k')\Big]\sum_{h'}\Big[\hat{\mathcal{H}}\{h'|h\} - \hat{\mathcal{H}}\{h'|h+1\}\Big]. \end{aligned}$$

The result follows because $\sum_{h'}[\hat{\mathcal{H}}\{h'|h\} - \hat{\mathcal{H}}\{h'|h+1\}] = 0$. □

Analogous to a liver-based or a rank-based control-limit optimal policy, a *health-based control-limit optimal policy* is defined as an optimal policy that prescribes, for a given rank state, k, and a liver quality, ℓ, 'wait' in all health states up to (and including) a threshold health state h^* and 'transplant' in all health states greater than h^*. Given the result of Theorem 3.5 and similar conditions to the conditions of Theorem 3 in [4], it can easily be shown that there exists a health-based control-limit optimal policy. We present Theorem 3.6 without a proof.

Theorem 3.6. *If $\hat{\mathcal{H}}$ is IFR and $\mathcal{H}$ and $r_T(h,\ell)$ satisfy the following:*

$$\sum_{h'=j}^{H} \mathcal{H}\{h'|h\} \leq \sum_{h'=j}^{H} \mathcal{H}\{h'|h+1\} \text{ for } j = h+1,\ldots,H, \text{ and } h = 1,\ldots,H, \tag{3.20}$$

and for any given ℓ

$$\frac{r_T(h,\ell) - r_T(h+1,\ell)}{r_T(h,\ell)} \leq \lambda\Big[\mathcal{H}\{H+1|h+1\} - \mathcal{H}\{H+1|h\}\Big] \text{ for } h = 1,\ldots,H-1, \tag{3.21}$$

then there exists a health-based control-limit optimal policy for all $\ell \in \Phi$ and $\boldsymbol{\pi} \in \Pi(\Psi)$.

3.4 COMPUTATIONAL STUDY

In this section, we present numerical results driven by clinical data. Section 3.4.1 discusses the parameter estimation process for the MDP model formulated in Section 3.1 and presents a numerical example for a patient with Hepatitis B. Section 3.4.2 discusses the concept of price of privacy in greater detail and presents the results of a numerical study for 200 ESLD patients.

3.4.1 Parameter estimation and an example

In our computational experiments, we define each period to be one day and consider the objective of maximizing the patient's total expected remaining lifetime. Therefore, we set $r_W(h) = 1$ for all $h \in \Omega$ and estimate the patient specific total expected post-transplant life days, $r_T(h,\ell)$ for all $h \in \Omega$ and $\ell \in \Phi$, using the post-transplant survival model of Roberts et al. [117].

As discussed in Section 1.1, adult ESLD patients are classified by disease severity into Status 1 patients and MELD patients. We only consider MELD patients in this study since there are typically fewer than a dozen Status 1 patients nationwide at a given time. In the allocation mechanism, each MELD patient has an integer-valued MELD score between 6 (healthiest) and 40 (sickest). However, due to sparsity of the available data, we represent patient health (h) by MELD scores aggregated in groups of two. Since the natural history

of ESLD depends on the type of diagnosis, we estimate different $\mathcal{H}$ matrices for different disease groups using the natural history model of Alagoz et al. [3].

We also follow the liver quality classification scheme of Alagoz et al. [6], which considers 14 liver qualities as determined by the age, race, and gender of the donor [117]. Detailed descriptions of the liver quality assignment scheme and the estimation of $r_T(h, \ell)$ are provided in Alagoz et al. [6].

To estimate $\mathcal{K}$ and $\mathcal{L}$ we use the national liver allocation model of Shechter et al. [132], which simulates the evolution of the waiting list under various liver allocation policies for the United States based on clinical data. We resort to this simulation model to collect rank information, which is not available in any form from any of the clinical resources. Tracking the rank of a patient in the nationwide waiting list would result in enormously large $\mathcal{K}$ and $\mathcal{L}$ matrices given that the waiting list contains over 15,000 patients. Moreover, the vast majority of offers are made to patients in the same geographic area as the donated liver, since the allocation mechanism exhausts the local geographic area before considering other areas. Therefore we simulate the national waiting list, but track the rank of the patients within the geographic area where they are registered. Since OPOs represent different populations, for each OPO we estimate different $\mathcal{K}$ and $\mathcal{L}$ matrices of varying sizes depending on the size of the geographic area served by the OPO. For notational convenience, we drop the dependency on OPOs in the following discussion.

We use 30 independent replications of the simulation to estimate $\mathcal{K}$ and $\mathcal{L}$. Shechter et al. [132] also used 30 replications for their simulation model and found that the simulation output closely matches UNOS data for several important statistics such as number of new patients listed, number of cadaveric donors, number of transplants, median waiting time for a transplant, and 1-year survival rates for patients and organs after receiving the transplant. To estimate $\mathcal{L}$, for each OPO, we count the number of offers each rank receives during the simulation, provided a liver is donated, and transform these counts into probabilities by computing the relative frequencty of each count. Let $f(k)$ represent the probability that a rank k patient in an arbitrary OPO receives an offer given that a liver is offered in her OPO. When a liver is offered in an OPO, the patient having the highest priority in this OPO receives this offer and depending on her decision the liver may be offered to the patient

with the second highest priority and so on. Therefore, we assume that $f(1) = 1$ and $f(k)$ is monotonically nonincreasing in k. The second step in estimating each $\mathcal{L}$ matrix involves augmenting $f(k)$ to distinguish between liver qualities. To do so, we fit an exponential function of the form $\exp(-\theta k)$ to $f(k)$ using ordinary least squares. Then, to ensure that higher quality livers are accepted earlier, we perturb θ to obtain $\tilde{f}(k, \ell)$, the probability that a rank k patient receives a donated liver of quality ℓ given that this liver is offered in her OPO. In doing so, we assume that the probability that a rank k patient receives an average quality liver is at most twice that of the highest quality liver (i.e., $\ell = 1$) and is at most 2/3 of the probability of receiving the lowest quality liver (i.e., $\ell = 14$). Specifically, we obtain the perturbed θ values, θ_ℓ, by

$$\theta_\ell = \begin{cases} \theta(1 + 1/\ell) & \text{if } \ell = 1, \ldots, 7, \\ \theta(1 - 0.05(\ell - 7)) & \text{if } \ell = 8, \ldots, 14. \end{cases}$$

For a typical OPO, Figure 3.1 shows the original $f(k)$ as estimated from the simulation and the best exponential fit to $f(k)$ as discussed above ($r^2 = 0.9991$ for this fit). Figure 3.2 shows the resulting $\tilde{f}(k, \ell)$ functions for $\ell = 1, \ldots, 14$. Next, we assume that liver arrivals in each OPO follow a Poisson process such that the probability that there is a liver offer in this OPO during an arbitrary day is given by α, where α is estimated using data obtained from UNOS [157]. Lastly, let $\gamma(\cdot)$ be the organ quality distribution of an arriving liver, which we obtain from the simulation by counting the number of liver offers of each quality. Then,

$$\begin{aligned} \mathcal{L}\{\ell|k\} &= \alpha \cdot \gamma(\ell) \cdot \tilde{f}(k, \ell) \text{ for all } \ell \neq L + 1, k \in \Psi, \text{ and} \\ \mathcal{L}\{L + 1|k\} &= 1 - \sum_{\ell \in \Phi} \mathcal{L}\{\ell|k\} \text{ for all } k \in \Psi. \end{aligned}$$

Even within an OPO, the number of rank states can be as large as a few thousand for OPOs serving large populations, which may yield computationally intractable problems. For this reason, for each OPO, we aggregate the rank states into 30 new rank states, which is found to be computationally tractable for the numerical study presented in Section 3.4.2, using the original $f(k)$ estimates in the following manner. Given a number of original rank states, we start from rank 1 and consolidate the first j ranks, where j is the maximum number of ranks such that the difference between the average function value over these j

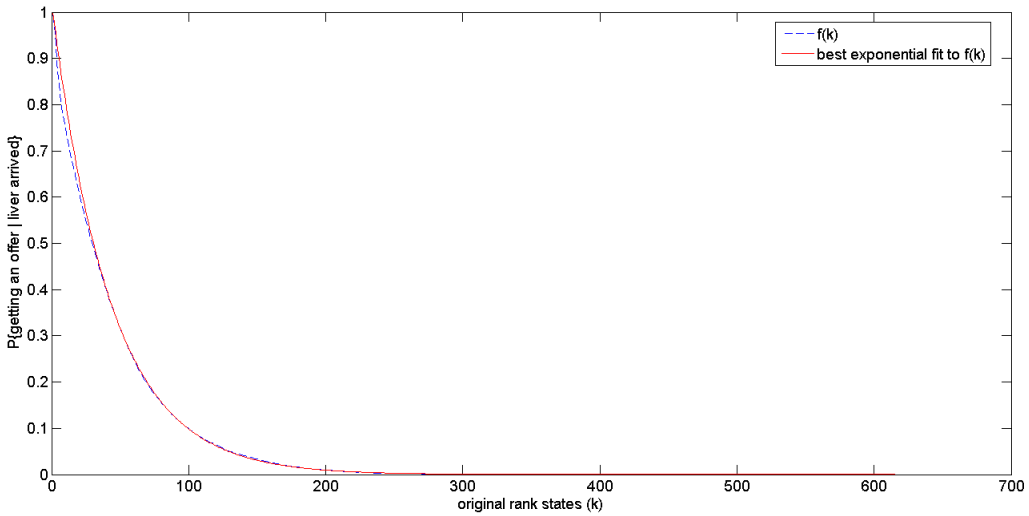


Figure 3.1: Liver offer probabilities as a function of rank given there is an offer for OPO 'A'

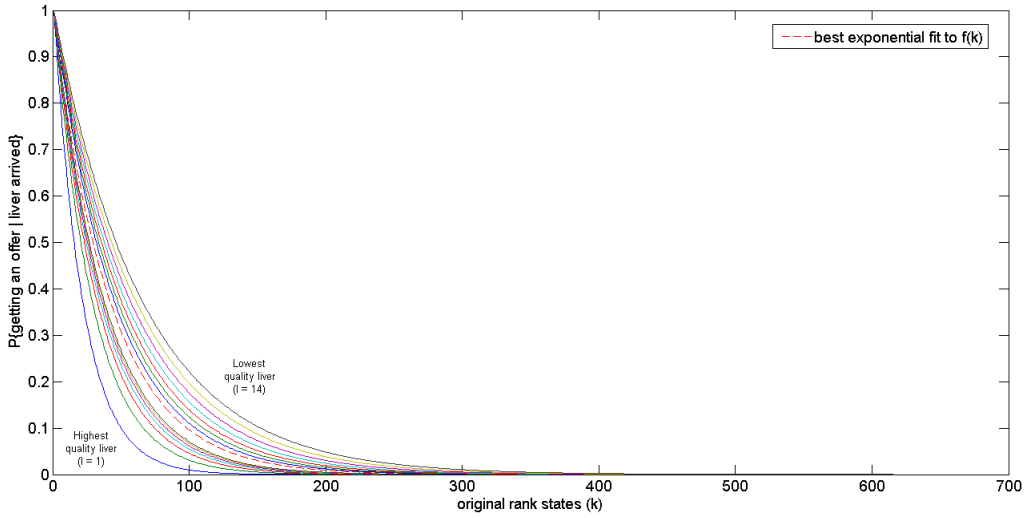
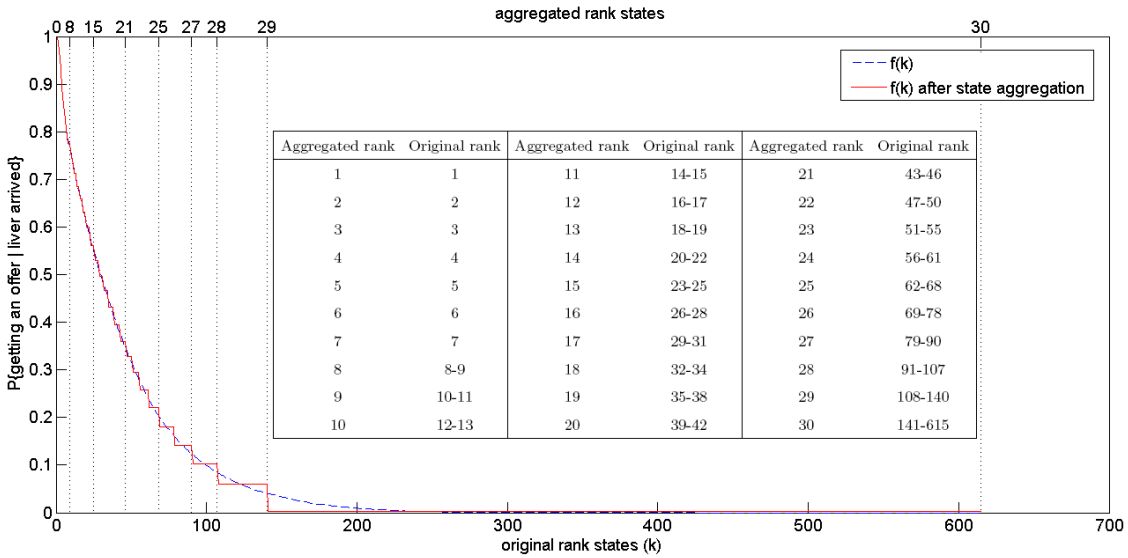


Figure 3.2: Liver offer probabilities as a function of rank given there is an offer of some quality for OPO 'A'



Aggregated rank	Original rank	Aggregated rank	Original rank	Aggregated rank	Original rank
1	1	11	14-15	21	43-46
2	2	12	16-17	22	47-50
3	3	13	18-19	23	51-55
4	4	14	20-22	24	56-61
5	5	15	23-25	25	62-68
6	6	16	26-28	26	69-78
7	7	17	29-31	27	79-90
8	8-9	18	32-34	28	91-107
9	10-11	19	35-38	29	108-140
10	12-13	20	39-42	30	141-615

Figure 3.3: Aggregated rank information for OPO 'A'

states and the function value at the $(j+1)^{st}$ state (e.g., $(1/j)\sum_{i=1}^{j} f(i) - f(j+1)$) is larger than some predetermined threshold. The threshold is found by line search so as to guarantee 30 final rank states. We then repeat this process starting from the $(j+1)$st rank and so on. Figure 3.3 depicts the approximation generated by this aggregation scheme for a typical OPO. Note that this OPO originally has 615 rank states. The first 7 original rank states remain unaggregated, the next 2 form rank 8 and so on.

To estimate the corresponding 30×30 $\mathcal{K}$ matrix for each OPO, we count the number of transitions during the simulation from each rank to every rank. We then transform these counts into transition probabilities by dividing each count by its row sum. Finally, we average the resulting transition probabilities across replications. Across all estimates of $\mathcal{K}$, the maximum standard error of the point estimates varies between 0.1245% and 0.3682% with an average of 0.1844% and a standard deviation of 0.0448%. Similarly, the maximum standard error of the point estimates in $\mathcal{L}$ varies between 0.0684% and 0.3616% with an average of 0.1594% and a standard deviation of 0.0709%.

Finally, we assume an annual discount rate of 0.97, which translates into a daily discount rate (λ) of 0.999917.

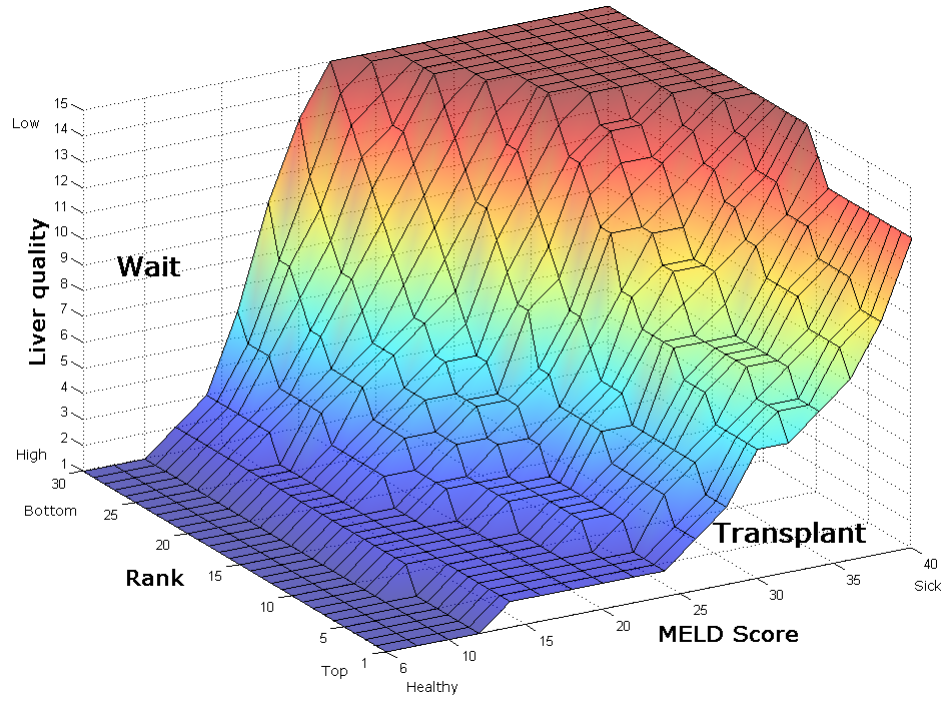


Figure 3.4: Control-limit optimal policy in all parameters for a 50-year old patient with Hepatitis B

Given these parameter estimates, Figure 3.4 depicts the optimal policy, which exhibits control-limit structure in all components, for a 50-year old female patient in OPO 'B' with Hepatitis B. In Figure 3.4, liver quality 15 represents the 'no liver offer' case. As shown in Figure 3.4, the optimal action varies across liver quality and health as measured by MELD score. If the patient is at the top of the list and has a MELD score of 20, then the optimal policy prescribes 'transplant' for a liver of quality 1 and 'wait' otherwise. This is an example of a liver-based control-limit optimal policy. Similarly, if the patient is at the top of the list and receives a liver offer of quality 2, then the optimal policy prescribes the 'wait' action if her MELD score is below 25 and the 'transplant' action otherwise. This is an example of a health-based control-limit optimal policy. Figure 3.4 further shows that the optimal action varies significantly by rank as well. The 'Transplant' region is smallest when the patient is at the top of the list and gradually grows as her rank deteriorates. In other words, the patient is more selective if she is at the top of the list and becomes less selective if her rank decreases. For instance, when the patient's MELD score is 20 and she is at the top of the list, she rejects all liver offers of quality $\ell \geq 2$ and accepts only the highest quality liver. However, at the same MELD score, if she is at the bottom of the list, she rejects only liver offers of quality $\ell \geq 12$ and accepts all liver offers of quality $\ell < 12$, which is an example of a rank-based control-limit optimal policy.

We randomly generate 200 test problems using the simulation model of Shechter et al. [132]. In all of these 200 instances, assumptions (AS1) and (AS2) are satisfied by the reward estimates $r_W(h)$ and $r_T(h,l)$, and the conditions on $\mathcal{L}$ in Theorems 3.3 and 3.4 are satisfied by the parameter estimates. However, the *CCD* condition on the $\mathcal{K}$ matrix is not always satisfied, although violations are not large. To quantify the magnitude of the violation of the *CCD* condition, we define the following metric:

$$\epsilon = \sum_{k'} \epsilon_{k'}$$

where

$$\epsilon_{k'} = \begin{cases} \sum_{k'} \max\{0, \mathcal{K}\{k'|k\} - \mathcal{K}\{k'|k+1\}\} & \text{for } k' = 1, \ldots, k-1, \\ \sum_{k'} \max\{0, \mathcal{K}\{k'|k+1\} - \mathcal{K}\{k'|k\}\} & \text{for } k' = k, k+1, \ldots. \end{cases}$$

Intuitively $\epsilon_{k'}$ measures the maximum non-negative difference in moving to a particular rank k' from any two successive ranks. Note that ϵ depends on the geographic area. Across all the estimates of $\mathcal{K}$ for each of the OPOs, ϵ varies between 0.001838 and 0.005425 with an average of 0.003332 and a standard deviation of 0.000806. Given the monotonicity of the value function in k and the rank-based control-limit optimal policy in all of the 200 test instances, we conclude that the results of Theorems 3.3 and 3.4 are fairly robust to small violations of the *CCD* requirement for the $\mathcal{K}$ matrix. We refer the reader to Alagoz et al. [4] for a discussion of violations on the $\hat{\mathcal{H}}$ matrix.

3.4.2 Estimating the price of privacy

The societal price of privacy is the aggregate benefit that society would accrue if the waiting list were made transparent. An estimate for the true price of privacy would be obtained by comparing a system (in which every patient has partial rank information as in the current allocation system and behaves optimally with this information) to a benchmark system (in which every patient has full rank information and behaves optimally). Our current model is unable to provide an exact value for the societal price of privacy since if the waiting list were to become transparent, the organ offer probabilities would change substantially as the allocation system moved to a new equilibrium, thus making precise parameter estimation using existing data impossible. Rather, due to difficulties in identifying an equilibrium in either of these systems, we focus on a special case where only one patient, who is provided the waiting list information, is considered. As a result, the quantities we provide can be viewed as estimates for the true values. We define the *patient's price of privacy* (*PPoP*) as the number of life days gained when she acts optimally based on full knowledge of the waiting list, as opposed to her optimal actions under the current allocation rules.

We provide an estimate of the *PPoP* by comparing the explicit waiting list model (EWLM) of Section 3.1 to the implicit waiting list model (IWLM) of Alagoz et al. [6]. More specifically, let π_a be an optimal IWLM policy and π_E be an optimal EWLM policy for the same patient. For any given $h \in \Omega$ and $\ell \in \Phi$, define

$$\pi_I(h, \ell, k) = \pi_a(h, \ell) \text{ for all } k \in \Psi.$$

This policy π_I may be viewed as the projection of the optimal IWLM policy onto the EWLM state space. Intuitively, if the patient does not have any rank information and solves IWLM, then her optimal actions as prescribed by this model should be same for each (h, ℓ) pair regardless of her rank, k. Let the benefit of using policy π_E over policy π_I in state $(h, \ell, k) \in \mathcal{S}'$ be given by:

$$v^{\pi_E}(h, \ell, k) - v^{\pi_I}(h, \ell, k),$$

where $v^{\pi_E}(h, \ell, k)$ and $v^{\pi_I}(h, \ell, k)$, respectively, is the total expected discounted reward for state (h, ℓ, k) associated with policy π_E and π_I.

Proposition 3.5 establishes that the benefit of using policy π_E over policy π_I is nonnegative in every state, which implies that an optimal policy recommended by IWLM may not provide the true optimal policy for the EWLM. The proof is obvious and omitted.

Proposition 3.5. *$b(h, \ell, k) \geq 0$ for all $h \in \Omega, \ell \in \Phi$ and $k \in \Psi$.*

Finally, let $\mathbf{b}$ be a probability distribution over the rank states Ψ. This probability distribution $\mathbf{b}$ is interpreted as the patient's probabilistic belief about her possible ranks (i.e., the ith component of the vector $\mathbf{b}$, b_i, indicates the probability that the patient believes her rank is i). We estimate patient-specific $\mathbf{b}$ vector as a function of patient's OPO and listing MELD score using the simulation model of Shechter et al. [132].

We provide an estimate of a patient's price of privacy ratio (i.e., the ratio of the patient's price of privacy to her optimal reward under the current allocation rules) using the following formula

$$\rho = \frac{\sum_{k \in \Psi} b_k \cdot \left[v^{\pi_E}(\tilde{h}, L+1, k) - v^{\pi_E}(\tilde{h}, L+1, k) \right]}{\sum_{k \in \Psi} b_k \cdot v^{\pi_I}(\tilde{h}, L+1, k)}, \tag{3.22}$$

where $\tilde{h}$ is the patient's health at the time of her registration to the waiting list. This metric measures the improvement associated with using the optimal EWLM policy over the optimal IWLM policy as a fraction of the optimal value of being in state $(\tilde{h}, L+1, k)$ and takes a weighted sum of the improvements over all rank states $k \in \Psi$ with weights being b_k. We choose state $(\tilde{h}, L+1, k)$ because a patient rarely receives a liver offer on the day she joins the list. The quantity given by equation (3.22) provides an estimate of the true *PPoP* ratio partly because of the fact that IWLM does not model the partial information availability in the current liver allocation system. Also note that Sandıkçı et al. [124] uses a special case

of the formula provided in equation (3.22), where they set $b_K = 1$ and $b_k = 0$ for $k \neq K$ (K being the last possible rank).

Consider, for example, the patient whose optimal policy is depicted in Figure 3.4. The estimate of her price of privacy ratio, as computed using equation (3.22), is 9.11%, which corresponds to 146.5 additional expected life days. Figure 3.5 adds the projected optimal IWLM policy to Figure 3.4 for the same patient. First note that, by construction, the optimal action of this projected policy does not vary across rank states. Furthermore, although the trends are same, the control limits h^* and ℓ^* of the projected policy do not always coincide with those of the optimal EWLM policy. In other words, ℓ^* is nondecreasing in h for fixed k in both policies, however they do not always coincide. Similar observations apply for h^*.

We compute the estimate ρ for the true $PPoP$ ratio for each of the 200 patients we have generated. Figure 3.6 presents a histogram of the ρ values for all 200 patients. The ρ values range between 0.31% and 15.57%, with a median value of 4.59%, an average value of 5.22%, and a standard deviation of 3.82%.

Table 3.2 presents the descriptive statistics associated with the estimated $PPoP$ ratios for the sampled patients in different disease groups; Table 3.3 presents these statistics by age range; and Table 3.4 presents these statistics by geographic area. We observe that the mean ρ values for patients in disease groups 1 and 4 and for those in disease groups 2 and 3 are close to each other. However, the mean ρ for patients in disease groups 1 and 4 is more than 25% higher compared to that for patients in disease groups 2 and 3. Patients in disease groups 1 and 4 have the best post-transplant survival (e.g., 10-year survival is approximately 75%), whereas patients in disease groups 2 and 3 have much poorer post-transplant survival (e.g., 10-year survival is approximately 60%). Therefore, it appears that the price of privacy ratio declines as the benefit of transplantation declines.

The mean ρ for patients younger than 20 and older than 70 is about 40% less than that for patients in the remaining age groups. Although we do not observe any major difference in the mean ρ values for patients in the remaining age groups, the results indicate a general decrease in the price of privacy ratio as age increases. Because elderly patients have shorter post-transplant survival, this observation further supports the hypothesis that a patient's price of privacy ratio decreases with decreased post-transplant survival.

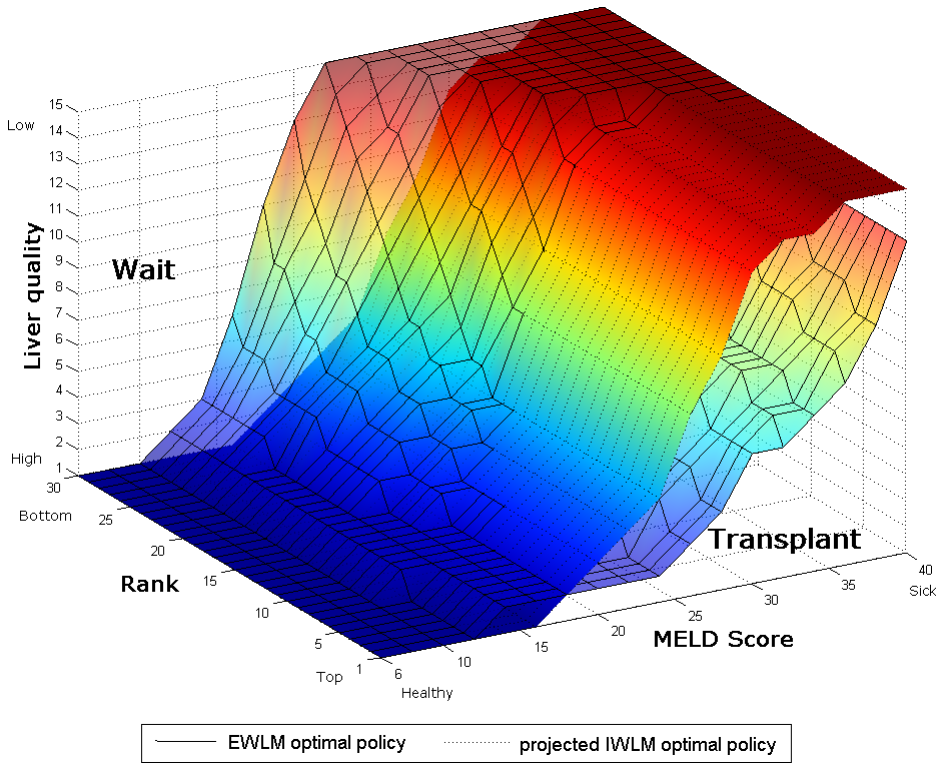


Figure 3.5: Comparison of EWLM and IWLM optimal policies

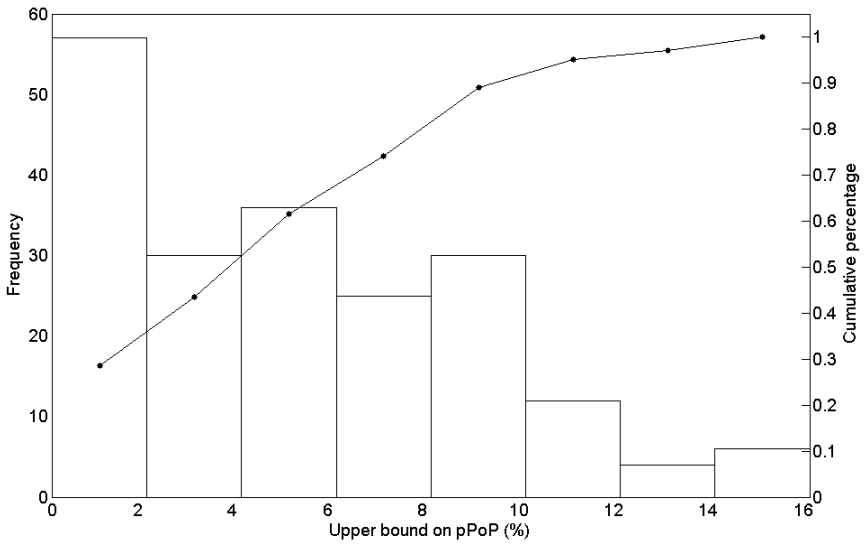


Figure 3.6: Histogram of the estimate for patient's price of privacy ratio for 200 patients generated from simulation

As seen in Table 3.4, Regions 6-11 serve significantly larger populations than those served by regions 1-5. The mean ρ for patients registered in OPOs serving larger populations is observed to be about 3.5 times the mean ρ for patients registered in OPOs serving smaller populations. This result is intuitive, because as the size of population served in a geographic area increases the liver offer probability differs significantly across ranks.

Finally, we also perform sensitivity analysis on ρ to understand the effect of the number of states used in our computations. Specifically, we picked 3 levels for the number of health states ($H = 4, 9, 18$), 2 levels for the number of liver qualities ($L = 4, 14$), and 3 levels for the number of rank states ($K = 10, 12, 30$). Table 3.5 summarizes the results of this analysis. As expected, we observe a slightly decreasing trend in average ρ values as we aggregate more rank states in our model. For instance, when $H = 4$ and $L = 4$ average ρ decreases from 5.21% at $K = 30$ to 4.32% at $K = 12$ and further down to 4.08% at $K = 10$. Similarly, simultaneously increasing H and L while keeping K constant also tends to increase the average ρ values.

Table 3.2: Descriptive statistics for estimated $PPoP$ ratio (ρ) by disease group

Disease group	# of patients	Min	Max	Median	Mean	Standard deviation
1	81	0.31%	15.01%	4.82%	5.61%	4.34%
2	71	0.47	11.53	4.35	4.79	2.96
3	22	0.75	10.36	2.88	3.85	2.82
4	26	0.60	15.57	6.34	6.36	4.55
ALL	200	0.31	15.57	4.59	5.22	3.82

Disease group 1 includes primary biliary cirrhosis, primary sclerosing cholangitis, alcoholic liver disease, and autoimmune disorders; disease group 2 includes Hepatitis B and C viruses; disease group 3 patients have acute liver failure; and disease group 4 patients have metabolic disorders (e.g., glycogen storage disease types I and II, and Gaucher's disease.)

Table 3.3: Descriptive statistics for estimated $PPoP$ ratio (ρ) by age group

Age group	# of patients	Min	Max	Median	Mean	Standard deviation
Age < 20	5	1.27%	5.38%	2.05%	2.66%	1.73%
$20 \leq$ Age < 30	8	1.09	11.78	4.46	5.27	4.06
$30 \leq$ Age < 40	34	0.31	15.57	4.95	5.78	4.09
$40 \leq$ Age < 50	54	0.73	14.56	6.02	5.96	3.32
$50 \leq$ Age < 60	56	0.35	14.50	3.92	4.76	4.09
$60 \leq$ Age < 70	39	0.40	15.01	4.18	4.84	4.01
Age ≥ 70	4	1.43	6.04	3.83	3.78	2.08
ALL	200	0.31	15.57	4.59	5.22	3.82

Table 3.4: Descriptive statistics for estimated $PPoP$ ratio (ρ) by geographic area

Region ID	# of patients	Min	Max	Median	Mean	Standard deviation
1	2	0.47%	0.60%	0.53%	0.53%	0.10%
2	6	0.64	2.91	0.88	1.18	0.86
3	8	0.40	2.34	1.33	1.28	0.70
4	7	0.44	3.70	1.47	1.84	1.18
5	6	0.31	5.40	2.25	2.55	2.15
6	20	0.72	7.68	3.09	3.11	1.96
7	25	1.41	9.18	4.18	4.66	2.26
8	28	0.55	15.01	5.77	5.18	3.76
9	34	0.54	15.57	5.01	6.33	5.30
10	15	1.76	10.66	5.51	6.71	2.89
11	49	0.59	12.14	8.35	7.32	3.12
ALL	200	0.31	15.57	4.59	5.22	3.82

Table 3.5: Effect of Number of states on ρ estimates

H	L	K	Median	Average	Standard deviation
4	4	10	2.64%	4.08%	3.99%
4	4	12	2.88	4.32	4.08
4	4	30	4.04	5.21	4.54
9	4	10	2.68	3.89	3.78
9	4	12	3.03	4.12	3.87
9	4	30	4.12	4.97	4.29
18	14	10	2.92	5.25	5.57
18	14	12	3.13	5.41	5.58
18	14	30	4.19	6.08	5.77

3.5 CONCLUSION

We develop an MDP model to optimize the accept/reject decision faced by ESLD patients. This model explicitly considers waiting list effects by augmenting the state space of the implicit waiting list model studied by [6]. We derive conditions under which the optimal value function is monotone in each dimension of the state space, namely, health, liver quality, and rank, and conditions under which control-limit optimal policies exist for each dimension. In establishing these results, we define a new class of stochastic matrices, termed *CCD* matrices, and explore its relationship to well known classes of matrices (i.e., *IFR* and TP_2 matrices). Computational experiments parameterized by clinical data reveal that complete knowledge of the composition of the waiting list significantly affects the optimal policy. In particular, a patient is much more selective if she knows that she is near the top of the waiting list and becomes gradually less selective as her position deteriorates.

We solve our explicit waiting list model for 200 randomly generated patients. Although the conditions of Theorems 3.3 (monotonicity in rank) and 3.5 (monotonicity in health) are not always satisfied, in all 200 cases the value function is monotone in each component of the model (i.e., h, ℓ, and k.) This result suggests that the monotonicity of the value function is robust to small violations of the *CCD* condition on the rank transition probability matrix, as well as the *IFR* condition on the augmented health transition probability matrix. In all of our computational experiments, the conditions of Theorem 3.2 hold, and therefore we find the optimal policy to be of liver-based control-limit type. Although the *CCD* condition of Theorem 3.4 is not always satisfied, the magnitudes of the violations are not significant, and, as a result, the optimal policy always has a rank-based control limit in our experiments. However, for some patients, the optimal policy does not have a health-based control limit. As noted by [6], as the patient deteriorates, the rate of increase in the probability of receiving higher quality liver offers may be sufficiently high so that she rejects low-quality livers that she would have accepted in better health in anticipation of higher-quality liver offers.

We use our explicit waiting list model to estimate a patient's price of privacy, which is incurred due to suboptimal decision making from a lack of complete waiting list information. By comparing the results of our model to those of the implicit waiting list model of [6],

we provide a quantitative estimate for a patient's price of privacy. Our computational experiments reveal that this quantity varies significantly by the particular etiology of ESLD. Indeed, although the majority of the patients would realize a less than 6% increase in total expected remaining lifetime by having complete information about the waiting list, there are patients in our study who realize improvements as high as 15%. In particular, patients diagnosed with diseases that yield shorter post-transplant survivals tend to have a smaller price of privacy ratio. Similarly, the price of privacy ratio tends to be smaller for older patients, who typically have shorter post-transplant survivals. Furthermore, our results indicate that patients registered in OPOs serving larger populations experience a higher price of privacy ratio. Finally, we have numerically confirmed, by solving different state aggregations, that the values we find may be underestimating the true price of privacy bound.

It is conceivable that patients registered at transplant centers that have a large market share in a geographic location may know (through their physicians) more than the revealed coarse descriptions about the waiting list. They may even be able to identify the precise ranks of the patients registered in the same location. The price of privacy for patients listed in such centers may be significantly smaller than what is suggested by our numerical study.

We emphasize that the estimate we provide for a patient's price of privacy can be further refined by incorporating the partially observable nature of the waiting list under the current liver allocation system. In Chapter 4, we model this partial information availability and compare it with EWLM to obtain better estimates of the price of privacy.

Throughout this chapter, we have suppressed the competition among patients when making their own decisions. However, a more realistic model should incorporate this competition and characterize the emerging equilibrium. This is an extremely challenging task as complete equilibrium analysis would require analyzing an asymmetric stochastic game with thousands of players each with non-zero-sum rewards. Unfortunately, describing all the equilibria of such a large-scale asymmetric stochastic game is well beyond the current state of the art. The three most important questions to be answered in stochastic games are existence, uniqueness, and computation of equilibrium. It can be shown that an equilibrium does exist, since every finite (i.e., finite state space and finitely many players with finite action space for

each player) discounted stochastic game has an equilibrium among the stationary strategies [143]. However, proving the uniqueness and computation of equilibria are more difficult. It is well known that there is a vast number equilibria for many normal-form games [95] and for stochastic games as well [64]. Even the extremely simplifying assumption of symmetry does not exclude the existence of other equilibria [55, 146].

Current literature that analyze competitive equilibrium makes such unrealistic assumptions (e.g., homogeneous patient population assumption in [146]). A methodologically similar paper to [146] is Altman and Shimkin [11], which analyzes the competitive equilibrium in a processor sharing system. They explicitly note the following: "*In the sequel we shall take special interest in equilibrium policies that are symmetric, namely the decision rules of all customers are identical. Such policies are natural here, since the specifications of all customers are the same, and they all face the same decision problem.*" However, acknowledging the well documented heterogeneity of the patients in liver transplantation yields intractable asymmetric game. Analysis of such a large-scale game-theoretic model is left for future research.

4.0 PARTIALLY OBSERVABLE WAITING LIST MODEL

In computing our estimate for a patient's price of privacy in Chapter 3, we compared the value associated with the optimal policy of the EWLM to the value associated with the optimal policy of the IWLM. However, the IWLM does not use all the information available to the patient under the current liver allocation system. Replacing the IWLM with a model that better represents the available information in the current system would provide a better estimate for a patient's price of privacy. A natural formulation using this information, which only provides partial knowledge about the waiting list, is a partially observable Markov decision process (POMDP) model.

Our goal in this chapter is to first develop and analyze a partially observable waiting list model (POWLM) for the purpose of optimizing the accept/reject decisions of an end-stage liver disease patient. Secondly, we want to use this model in conjunction with the explicit waiting list model of Chapter 3 to provide better estimates of the patient's price of privacy.

The remainder of this chapter is organized as follows. In Section 4.1, we review the information that is available under the current system. In Section 4.2, we present the details of the partially observable waiting list model. We derive structural results for this model in Section 4.3, which is followed by computational study in Section 4.4. Finally, we summarize our results and draw some conclusions in Section 4.5.

4.1 WHAT IS PARTIALLY OBSERVED IN THE CURRENT SYSTEM?

The current UNOS liver allocation system allows patients to retrieve partial information about the liver transplant waiting list through UNOS's website [106]. This information may

Table 4.1: Snapshot of the liver waiting list for the OPO serving Pittsburgh [106]

Disease severity	Number of registrations
MELD ≤ 10	152
$11 \leq$ MELD < 19	157
$19 \leq$ MELD < 25	45
MELD ≥ 25	33
ALL	387

help each patient to have better estimates of their respective ranks in the waiting list than the one assumed in the IWLM.

As an example, Table 4.1 presents a recent (condensed) snapshot of the liver transplant waiting list for the OPO labeled PATF-OP1, which serves the greater Pittsburgh area. Assuming a liver has been harvested from this OPO at the time of this snapshot (or shortly after during which the snapshot characteristics did not change), it is possible to determine a range of possible ranks for any patient. For instance, if the patient has a MELD score of 30, then she can be no lower than 33rd on the prioritized waiting list. If, on the other hand, the patient's MELD score is 18, then her rank could be as high as 78 or as low as 235. However, the patient cannot precisely pinpoint her rank within the given ranges.

Since blood type and waiting time are also part of the UNOS prioritization algorithm as described in Chapter 1, it is tempting to conclude that these pieces of information, which are also observable through UNOS's website, would enable the patient to better assess her rank. However, this conjecture is not true in general. Such blood type and waiting time information may only be useful if the patient's MELD score is equal to one of the end points of a MELD partition. Table 4.2 provides an expanded version of the waiting list snapshot given in Table 4.1. If, for instance, the patient's MELD score is 30 (or anything strictly between 25 and 40), then blood type and waiting time information provided in Table 4.2 would not add anything to her knowledge about her rank (i.e., all she knows is she is no lower than 33rd on the prioritized waiting list). If, however, her MELD score is 40 (the

Table 4.2: Expanded snapshot of the liver waiting list for the OPO serving Pittsburgh [106]

MELD	ABO	Days waited								
		< 30	30-89	90-179	180-364	365-729	730-1094	1095-1824	≥ 1825	ALL
≤ 10	O	1	4	3	11	17	11	8	14	69
	A	3	5	8	6	16	9	11	10	68
	B	-	-	-	1	3	3	1	3	11
	AB	-	-	2	-	1	1	-	-	4
	ALL	4	9	13	18	37	24	20	27	152
$11-18$	O	4	9	10	8	18	10	13	7	79
	A	4	8	9	13	15	9	4	4	66
	B	-	3	1	4	2	-	-	1	11
	AB	-	1	-	-	-	-	-	-	1
	ALL	8	21	20	25	35	19	17	12	157
$19-24$	O	3	4	4	3	2	1	4	1	22
	A	5	3	3	1	1	3	1	-	17
	B	-	2	1	1	-	-	-	-	4
	AB	2	-	-	-	-	-	-	-	2
	ALL	10	9	8	5	3	4	5	1	45
≥ 25	O	1	3	3	7	4	1	-	-	19
	A	-	1	1	3	-	1	2	-	8
	B	-	1	-	1	1	-	-	-	3
	AB	-	-	-	3	-	-	-	-	3
	ALL	1	5	4	14	5	2	2	-	33

upper end point of the last MELD partition) and she knows that her blood type (say, type A) is identical to the donated liver (which is revealed when the liver is offered), then this information along with the waiting time information may help the patient narrow her rank knowledge. If, say, this patient has been waiting for about 10 months on the waiting list, then she can narrow her rank down to top 6 $(3 + 1 + 2)$ on the prioritized waiting list. Our model described in Section 4.2 is general enough to handle such observations. However, we will restrict our observations to only patient's health status as measured by her MELD score in our computational study.

In addition to the information provided on UNOS's website, patients can also attain partial information about their waiting list rank by other means including the history of offers they have received and the information they may obtain through their physicians. For instance, not receiving any liver offers for a certain period may indicate that patient's rank is low. Furthermore, based on their expert opinion and knowledge of the characteristics of other patients in their transplant center, some physicians may be able to guess more precise estimates about their patient's possible ranks. Our model described in Section 4.2 is also general enough to capture such information. However, since these information are subjective and hard to obtain, we will restrict ourselves to the information published on UNOS's website in our computational study.

4.2 PARTIALLY OBSERVED MARKOV DECISION PROCESS MODEL

In this section, we formulate a discrete-time partially observed Markov decision process model that exploits the partially observable rank information in optimizing the accept/reject decision problem faced by an end-stage liver disease patient.

The set of *core* states, $\mathcal{S}$, of our POMDP is the same as the set of states in the EWLM given in Section 3.1. Furthermore, the POMDP model also uses the same action sets, $\mathcal{A}_s$ for $s \in \mathcal{S}$, rewards, $r(s, a)$ for $s \in \mathcal{S}, a \in \mathcal{A}_s$, and state transition probabilities, $P\{s'|s, a\}$ for $s, s' \in \mathcal{S}, a \in \mathcal{A}_s$, as in the EWLM. For notational convenience, let $\mathcal{A} = \bigcup_{s \in \mathcal{S}} \mathcal{A}_s$.

As for the observations, we assume that a patient can completely observe her health and the quality of the offered liver, but she may not directly know her rank. (Note that this assumption is not restrictive as our structural results would be valid if we allow health and liver states to be partially observed as well.) However, she can obtain probabilistic information about her rank through UNOS's website. Therefore, the best a patient can do is to observe a signal z that gives a probabilistic information about the actual state s. We define the set of possible observations as

$$\mathcal{Z} = \mathcal{Z}' \cup \{\Delta\} \cup \{\nabla\},$$

where Δ and ∇, respectively, represent the dead and transplanted states, and

$$\mathcal{Z}' = \{(h, \ell, y) | h \in \Omega, \ell \in \Phi, y \in \Upsilon\},$$

where Ω and Φ are, respectively, the sets of completely observable health and liver states as defined in Section 3.1, and $\Upsilon = \{Y_1, Y_2, \ldots, Y_p\}$ is the observation set associated with the partially observable rank state, where p is the number of observations one could realize at any point in time.

The special structure of the current system, as discussed in Section 4.1, dictates that an observation Y_i $(i = 1, \ldots p)$ is a *range of rank states* induced by Ω_i, which denotes the set of health states included in partition i. More formally, (i) $Y_i \subseteq \Psi$ for $i = 1, \ldots, p$; (ii) $Y_i \cap Y_j = \emptyset$ for any $i \neq j$, and (iii) $\cup_i Y_i = \Psi$, where Ψ is the set of rank states as defined in Section 3.1. For instance, considering the waiting list in Table 4.1, $\Omega_1 = \{6, \ldots, 10\} \Rightarrow Y_1 = \{236, \ldots, 387\}$, $\Omega_2 = \{11, \ldots, 18\} \Rightarrow Y_2 = \{79, \ldots, 235\}$, $\Omega_2 = \{19, \ldots, 24\} \Rightarrow Y_3 = \{34, \ldots, 78\}$, and $\Omega_4 = \{25, \ldots, 40\} \Rightarrow Y_4 = \{1, \ldots, 33\}$. Furthermore, a patient observes Y_i if and only if her health h is in partition Ω_i (i.e., $P\{y = Y_i\} = 1 \Leftrightarrow h \in \Omega_i$ for $i = 1, \ldots, p$.)

We define the observation probabilities (i.e, the probability of observing z' when the core state is s' at time $t + 1$ and action a was taken at time t), denoted $\mathcal{O}\{\cdot\}$, as follows. When the transplant action was taken, we observe ∇ with probability one (i.e., $\mathcal{O}\{\nabla | \nabla, T\} = 1$). When the wait action was taken and the patient dies in the next period, Δ is observed with

probability one (i.e., $\mathcal{O}\{\Delta|\Delta, W\} = 1$). When the wait action was taken and the patient survives the next time period, the observation probabilities for each (h, ℓ) are defined as:

$$\mathcal{O}\{Y_i|k, W\} = \begin{cases} 1 & \text{if } h \in \Omega_i, k \in Y_i, \\ 0 & \text{otherwise} \end{cases} \qquad \text{for } i = 1, \ldots, p.$$

It is clear that all the parameters associated with the observations (i.e., Y_i, Υ, and $\mathcal{O}\{\cdot\}$) are actually non-stationary, because the liver transplant waiting list is dynamic over time. That is, from this time point to the next, new patients may be registered to the list, existing patients may move out of the list due to death or being transplanted or other medical complications, and the disease severity of the existing patients may change. Therefore, the size of the waiting list and, more importantly, the sets Y_i $(i = 1, \ldots, p)$ will change over time, which implies that the observation probabilities will also change over time.

The non-stationary observation parameters would not add any complication to our conceptual model, however, for the sake of notational clarity, we do not attach a time index to Y_i, Υ, or $\mathcal{O}\{\cdot\}$. Furthermore, in our computational study, for the purposes of numerical tractability, we will assume a stationary process and use time-homogeneous observation probabilities, which we compute as a long-run time average of the realized observation probabilities. More details about this estimation are provided in Section 4.4.1.

Our modeling framework is also flexible enough to allow varying number of observations over time, however, again for notational convenience, we do not attach a time index to p. In our computational study, we interpret p as the number of partitions across MELD scores and we assume it to be fixed over time, which exactly represents the current UNOS liver allocation system. Furthermore, we will take $p = 4$ in our computational study, since the current system reveals the waiting list information by aggregating the MELD scores into 4 partitions.

In summary, the system is actually occupying the core state (h, ℓ, k), Δ, or ∇, but the patient only observes (h, ℓ, y), Δ, or ∇, where y represents the ranges of ranks she could possibly occupy.

Finally, note that in this representation, we are restricting our observations to the information retrieved through UNOS's website. Any other information, such as physician's

informed opinion about the patient's possible ranks, can be handled by virtually no change in our model.

4.2.1 Belief states

We have noted in Section 2.2 that any POMDP model can be transformed into an equivalent (completely observed) MDP with an enlarged state space (i.e., the space of probability distributions over the core states), which is known as the *belief* space. Any element of the belief space is referred to as a *belief* state.

Recall that the state space of our model is a mix of completely and partially observable states: the perfectly observable components of our state space are patient's health and liver quality whereas the partially observable component is patient's rank. It is actually not obvious what the appropriate belief vector definition would be under such a hybrid state space: should it be a probability distribution over the entire state space, $\mathcal{S}$, or only over the partially observable component, Ψ, of the state space? To answer this question, we will start with a belief vector definition over the entire state space and show that all the information available in this belief state can be summarized in a smaller belief state defined over only the partially observable components of the state space.

To start with, let $\boldsymbol{\eta}$ be a belief state defined over the entire state space $\mathcal{S}$. First note that $\boldsymbol{\eta}$ must satisfy the following: given $h \in \Omega$, $\ell \in \Phi$, if $\eta_{(h,\ell,k)} > 0$ for some $k \in \Psi$, then $\eta_{(h',\cdot,\cdot)} = 0$ for all $h' \neq h$ and $\eta_{(\cdot,\ell',\cdot)} = 0$ for all $\ell' \neq \ell$. This statement says that a legitimate belief state cannot assign positive probabilities to two different states which differ in perfectly observable components. This observation actually provides the first hint as to the sufficiency of a belief state defined only on the partially observable portion of the hybrid space $\mathcal{S}$.

We are now interested in updating this belief state using Bayes' rule. We have noted in Section 2.2 that the updated belief state is a function of the current belief state, the action taken during this decision epoch, and the observation received at the beginning of the next decision epoch. Let a be the action taken during this decision epoch and z' be the observation received at the beginning of the next decision epoch.

If $a = T$, then $z' = \nabla$ with probability 1 and the updated belief state $\boldsymbol{\eta'}$ is given by

$$\eta'_s = \begin{cases} 1 & \text{if } s = \nabla, \\ 0 & \text{if otherwise.} \end{cases}$$

If, on the other hand, $a = W$ and $z' = \Delta$, then the updated belief state $\boldsymbol{\eta'}$ is given by

$$\eta'_s = \begin{cases} 1 & \text{if } s = \Delta, \\ 0 & \text{if otherwise.} \end{cases}$$

Finally, if $a = W$ and $z' = (\tilde{h}, \tilde{\ell}, \tilde{y})$, then the updated belief state $\boldsymbol{\eta'}$ is given by

$$\eta'_s = \begin{cases} P\{\tilde{h}, \tilde{\ell}, \tilde{k} | \boldsymbol{\eta}, a, z'\} & \text{if } s = (\tilde{h}, \tilde{\ell}, \tilde{k}), \\ 0 & \text{if otherwise,} \end{cases} \qquad \text{for } \tilde{k} \in \Psi,$$

where $P\{\tilde{h}, \tilde{\ell}, \tilde{k} | \boldsymbol{\eta}, a, z'\}$ is the probability of being in state $(\tilde{h}, \tilde{\ell}, \tilde{k})$ given that the current belief is $\boldsymbol{\eta}$, action $a = W$ is taken, and observation $z' = (\tilde{h}, \tilde{\ell}, \tilde{y})$ is received. This expression shows that, if we have observed $(\tilde{h}, \tilde{\ell})$ at the beginning of the new decision epoch, the updated belief state $\boldsymbol{\eta'}$ will distribute all its mass across states (h, ℓ, k) such that $h = \tilde{h}$ and $\ell = \tilde{\ell}$. Therefore, it would be sufficient to define a belief state over only the partially observable component $k \in \Psi$.

Let $(\hat{h}, \hat{\ell}) = \{(h, \ell) | h \in \Omega, \ell \in \Phi, \eta_{(h,\ell,k)} > 0 \text{ for at least one } k \in \Psi\}$ (i.e., $\hat{h}$ and $\hat{\ell}$, respectively, records the perfectly observable health and liver state for which the current belief state assigns a positive probability.)

Also let $\boldsymbol{\pi}$ be a $K \times 1$ vector, where the kth component holds the probability of being in rank state $k \in \Psi$. That is, we have just introduced the belief state that is defined only over the partially observable rank states, which we will use frequently after this point. For any given $(\hat{h}, \hat{\ell})$, define $\pi_k = \eta_{(\hat{h},\hat{\ell},k)}$ for $k \in \Psi$.

Consider $P\{\tilde{h},\tilde{\ell},\tilde{k}|\boldsymbol{\eta},a,z'\}$. Since $a=W$ for this probability, we will drop it from our derivations for notational clarity. Substituting $z'=(\tilde{h},\tilde{\ell},\tilde{y})$ and using the definition of conditional probability twice we obtain

$$\begin{aligned} P\{\tilde{h},\tilde{\ell},\tilde{k}|\boldsymbol{\eta},a,z'\} &= \frac{P\{\tilde{k},\boldsymbol{\eta},\tilde{h},\tilde{\ell},\tilde{y}\}}{P\{\tilde{h},\tilde{\ell},\tilde{y},\boldsymbol{\eta}\}} \\ &= \frac{P\{\tilde{y}|\boldsymbol{\eta},\tilde{h},\tilde{\ell},\tilde{k}\}\cdot P\{\tilde{h},\tilde{\ell},\tilde{k},\boldsymbol{\eta}\}}{P\{\tilde{h},\tilde{\ell},\tilde{y},\boldsymbol{\eta}\}} \\ &= \frac{P\{\tilde{y}|\boldsymbol{\eta},\tilde{h},\tilde{\ell},\tilde{k}\}\cdot P\{\tilde{h},\tilde{\ell},\tilde{k}|\boldsymbol{\eta}\}}{P\{\tilde{h},\tilde{\ell},\tilde{y}|\boldsymbol{\eta}\}}, \end{aligned} \tag{4.1}$$

where we have used the identity $\frac{P\{\tilde{h},\tilde{\ell},\tilde{k},\boldsymbol{\eta}\}}{P\{\tilde{h},\tilde{\ell},\tilde{y},\boldsymbol{\eta}\}}=\frac{P\{\tilde{h},\tilde{\ell},\tilde{k}|\boldsymbol{\eta}\}}{P\{\tilde{h},\tilde{\ell},\tilde{y}|\boldsymbol{\eta}\}}$ in writing equation (4.1). Noting that $P\{\tilde{y}|\boldsymbol{\eta},\tilde{h},\tilde{\ell},\tilde{k}\}=\mathcal{O}\{\tilde{y}|\tilde{k}\}$ since the probability of observing $\tilde{y}$ is only dependent on the rank state and conditioning on (h,ℓ,k) to find $P\{\tilde{h},\tilde{\ell},\tilde{k}|\boldsymbol{\eta}\}$, we can re-write equation (4.1) as

$$P\{\tilde{h},\tilde{\ell},\tilde{k}|\boldsymbol{\eta},a,z'\}=\frac{\mathcal{O}\{\tilde{y}|\tilde{k}\}\cdot\sum_h\sum_\ell\sum_k P\{\tilde{h},\tilde{\ell},\tilde{k}|\boldsymbol{\eta},h,\ell,k\}P\{h,\ell,k,|\boldsymbol{\eta}\}}{P\{\tilde{h},\tilde{\ell},\tilde{y}|\boldsymbol{\eta}\}}. \tag{4.2}$$

We can replace $P\{\tilde{h},\tilde{\ell},\tilde{k}|\boldsymbol{\eta},h,\ell,k\}$ in the numerator of the right-hand side of equation (4.2) with $P\{\tilde{h},\tilde{\ell},\tilde{k}|h,\ell,k\}$, because we are given that the current state is (h,ℓ,k) and hence the probability distribution η over these states becomes redundant. Furthermore, since $P\{h,\ell,\cdot|\boldsymbol{\eta}\}=\eta_{(h,\ell,\cdot)}=0$ when $h\neq\hat{h}$ or $\ell\neq\hat{\ell}$, we can simplify equation (4.2) as

$$P\{\tilde{h},\tilde{\ell},\tilde{k}|\boldsymbol{\eta},a,z'\}=\frac{\mathcal{O}\{\tilde{y}|\tilde{k}\}\cdot\sum_k P\{\tilde{h},\tilde{\ell},\tilde{k}|\hat{h},\hat{\ell},k\}\cdot\eta_{(\hat{h},\hat{\ell},k)}}{P\{\tilde{h},\tilde{\ell},\tilde{y}|\boldsymbol{\eta}\}}. \tag{4.3}$$

Conditioning on (h,ℓ,k) for the current state and on (h',ℓ',k') for the state in the next decision epoch, we can write the term in the denominator of equation (4.3) as

$$\begin{aligned} P\{\tilde{h},\tilde{\ell},\tilde{y}|\boldsymbol{\eta}\} &= \sum_{h'}\sum_{\ell'}\sum_{k'}\sum_{h}\sum_{\ell}\sum_{k} P\{\tilde{h},\tilde{\ell},\tilde{y}|\boldsymbol{\eta},h,\ell,k,h',\ell',k'\}P\{h,\ell,k,h',\ell',k'|\boldsymbol{\eta}\} \\ &= \sum_{h'}\sum_{\ell'}\sum_{k'}\sum_{h}\sum_{\ell}\sum_{k} P\{\tilde{h},\tilde{\ell},\tilde{y}|h',\ell',k'\}P\{h,\ell,k,h',\ell',k'|\boldsymbol{\eta}\} && (4.4) \\ &= \sum_{k'}\sum_{h}\sum_{\ell}\sum_{k} P\{\tilde{h},\tilde{\ell},\tilde{y}|\tilde{h},\tilde{\ell},k'\}P\{h,\ell,k,\tilde{h},\tilde{\ell},k'|\boldsymbol{\eta}\} && (4.5) \\ &= \sum_{k'}\sum_{h}\sum_{\ell}\sum_{k} \mathcal{O}\{\tilde{y}|k'\}P\{h,\ell,k,\tilde{h},\tilde{\ell},k'|\boldsymbol{\eta}\} && (4.6) \\ &= \sum_{k'}\sum_{h}\sum_{\ell}\sum_{k} \mathcal{O}\{\tilde{y}|k'\}P\{\tilde{h},\tilde{\ell},k'|h,\ell,k,\boldsymbol{\eta}\}P\{h,\ell,k|\boldsymbol{\eta}\} && (4.7) \\ &= \sum_{k'}\mathcal{O}\{\tilde{y}|k'\}\cdot\sum_{k} P\{\tilde{h},\tilde{\ell},k'|h,\ell,k\}\cdot\eta_{(\hat{h},\hat{\ell},k)} && (4.8) \end{aligned}$$

Equation (4.4) follows since $P\{\tilde{h},\tilde{\ell},\tilde{y}|\boldsymbol{\eta},h,\ell,k,h',\ell',k'\} = P\{\tilde{h},\tilde{\ell},\tilde{y}|h',\ell',k'\}$; equation (4.5) follows from the fact that $P\{\tilde{h},\tilde{\ell},\tilde{y}|h',\ell',k'\} = 0$ when $h' \neq \tilde{h}$ or $\ell' \neq \tilde{\ell}$; equation (4.6) follows since the probability of observing $\tilde{y}$ is only dependent on the rank states; equation (4.7) follows since $P\{h,\ell,k,\tilde{h},\tilde{\ell},k'|\boldsymbol{\eta}\} = P\{\tilde{h},\tilde{\ell},k'|h,\ell,k,\boldsymbol{\eta}\}P\{h,\ell,k|\boldsymbol{\eta}\}$; and equation (4.8) follows from the fact that $P\{h,\ell,k|\boldsymbol{\eta}\} = \eta_{(h,\ell,k)} = 0$ when $h \neq \hat{h}$ or $\ell \neq \hat{\ell}$.

Substituting equation (4.8) back into equation (4.3), we obtain our main formula for updating a given belief state $\boldsymbol{\eta}$:

$$\eta'_{(\tilde{h},\tilde{\ell},\tilde{k})} = P\{\tilde{h},\tilde{\ell},\tilde{k}|\boldsymbol{\eta},a,z'\} = \frac{\mathcal{O}\{\tilde{y}|\tilde{k}\}\cdot\sum_k P\{\tilde{h},\tilde{\ell},\tilde{k}|\hat{h},\hat{\ell},k\}\cdot\eta_{(\hat{h},\hat{\ell},k)}}{\sum_{k'}\mathcal{O}\{\tilde{y}|k'\}\cdot\sum_k P\{\tilde{h},\tilde{\ell},k'|h,\ell,k\}\cdot\eta_{(\hat{h},\hat{\ell},k)}}. \tag{4.9}$$

where $\hat{h}$ and $\hat{\ell}$ are, respectively, the health and liver states for which the current belief state $\boldsymbol{\eta}$ assigns positive probabilities.

We can re-write the belief state updating formula (4.9) for belief vector $\boldsymbol{\pi}$, which is defined over the partially observed rank states only:

$$\pi'_{\tilde{k}} = \frac{\mathcal{O}\{\tilde{y}|\tilde{k}\}\cdot\sum_k P\{\tilde{h},\tilde{\ell},\tilde{k}|\hat{h},\hat{\ell},k\}\cdot\pi_k}{\sum_{k'}\mathcal{O}\{\tilde{y}|k'\}\cdot\sum_k P\{\tilde{h},\tilde{\ell},k'|\hat{h},\hat{\ell},k\}\cdot\pi_k}. \tag{4.10}$$

As a final step, we can further simplify the belief state updating formula given in equation (4.10) using the special structure of our core state transition probabilities. Using $P\{\tilde{h},\tilde{\ell},\tilde{k}|\hat{h},\hat{\ell},k\} = \mathcal{H}\{h'|h\}\cdot\mathcal{K}\{k'|k\}\cdot\mathcal{L}\{\tilde{\ell}|k'\}$ in (4.10), we obtain

$$\pi'_{\tilde{k}} = \frac{\mathcal{O}\{\tilde{y}|\tilde{k}\}\cdot\mathcal{L}\{\tilde{\ell}|\tilde{k}\}\cdot\sum_k \mathcal{K}\{\tilde{k}|k\}\cdot\pi_k}{\sum_{k'}\mathcal{O}\{\tilde{y}|k'\}\cdot\mathcal{L}\{\tilde{\ell}|k'\}\cdot\sum_k \mathcal{K}\{k'|k\}\cdot\pi_k}. \tag{4.11}$$

where we have assumed $\mathcal{H}\{\tilde{h}|\hat{h}\} \neq 0$ in obtaining equation (4.11). This assumption is needed for $\boldsymbol{\pi}'$ to be well defined. In other words, if $\mathcal{H}\{\tilde{h}|\hat{h}\} = 0$, then $\pi'_{\tilde{k}} = \frac{0}{0}$. We should further note that this assumption makes perfect intuitive sense and is not restrictive. To explain this, assume the system is in state (h,ℓ,k) and the decision maker chooses action a while in this state. As a result of this action, assume the decision maker observed $(\tilde{h},\tilde{\ell},\tilde{y})$ at the beginning of the next decision epoch. If this is the case, then there must be a nonzero transition probability from h to $\tilde{h}$ since health component of the state space is perfectly observable, otherwise the decision maker should not have have observed $\tilde{h}$.

We should also note that equation (4.11) reveals that the decision maker actually receives two pieces of information regarding her rank at every decision epoch: (i) the partition $\tilde{y}$, and (ii) the liver offered $\tilde{\ell}$. If, for instance, there were no organ offers today, that would suggest that the patient's rank is lower than what it would be if there was an organ offer. We perform a single Bayesian update based on both pieces of information.

4.2.2 Optimality equations

The optimal solution to this problem can be obtained by solving the Bellman optimality equations. We will start writing these equations for the larger belief vectors $\boldsymbol{\eta}$ and derive the equivalent optimality equations for the smaller belief vectors $\boldsymbol{\pi}$, which we will use in our computational study.

We first start with the definition of a probability simplex.

Definition 4.1. *Given a set $\mathbb{X}$ with n elements, the n-dimensional probability simplex defined over $\mathbb{X}$, denoted $\Pi(\mathbb{X})$, is the set of all probability mass functions on $\mathbb{X}$, i.e,*

$$\Pi(\mathbb{X}) = \left\{ \boldsymbol{\pi} \in \mathbb{R}^n : \sum_{i=1}^{n} \pi_i = 1, \quad \pi_i \geq 0 \quad \forall i \right\}.$$

Let $\lambda \in [0,1]$ be the discount rate. Substituting the parameters of our partially observable waiting list model into the generic optimality equation (2.8), we find for $\boldsymbol{\eta} \in \Pi(\mathcal{S})$

$$\begin{aligned}
&v(\boldsymbol{\eta}) \\
&= \max\Bigg\{ \sum_{(h,\ell,k)\in\mathcal{S}} \eta_{(h,\ell,k)} r_T(h,\ell); \qquad \sum_{(h,\ell,k)\in\mathcal{S}} \eta_{(h,\ell,k)} r_W(h) \quad + \\
&\quad \lambda \cdot \sum_{(h,\ell,k)\in\mathcal{S}} \sum_{(h',\ell',k')\in\mathcal{S}} \sum_{(\tilde{h},\tilde{\ell},\tilde{y})\in\mathcal{Z}} \eta_{(h,\ell,k)} P\{(h',\ell',k')|(h,\ell,k)\} P\{(\tilde{h},\tilde{\ell},\tilde{y})|(h',\ell',k')\} \cdot v(\boldsymbol{\eta}') \Bigg\} \\
&= \max\Bigg\{ r_T(\hat{h},\hat{\ell}); \qquad r_W(\hat{h}) \quad + \\
&\quad \lambda \cdot \sum_{k\in\Psi} \sum_{(h',\ell',k')\in\mathcal{S}} \sum_{(\tilde{h},\tilde{\ell},\tilde{y})\in\mathcal{Z}} \eta_{(\hat{h},\hat{\ell},k)} P\{(h',\ell',k')|(\hat{h},\hat{\ell},k)\} P\{(\tilde{h},\tilde{\ell},\tilde{y})|(h',\ell',k')\} \cdot v(\boldsymbol{\eta}') \Bigg\} \quad (4.12)
\end{aligned}$$

$$= \max\Bigg\{ r_T(\hat{h},\hat{\ell}); \qquad r_W(\hat{h}) \;+$$

$$\lambda \cdot \sum_{k\in\Psi} \sum_{(h',\ell',k')\in\mathcal{S}} \sum_{\tilde{y}\in\Upsilon} \eta_{(\hat{h},\hat{\ell},k)} P\{(h',\ell',k')|(\hat{h},\hat{\ell},k)\}\mathcal{O}\{\tilde{y}|k'\}\cdot v(\boldsymbol{\eta}')\Bigg\} \tag{4.13}$$

where equation (4.12) follows from the fact that $\eta_{(h,\ell,k)} = 0$ when $h \neq \hat{h}$ or $\ell \neq \hat{\ell}$ and equation (4.13) follows since $P\{(\tilde{h},\tilde{\ell},\tilde{y})|(h',\ell',k')\} = 0$ when $h' \neq \tilde{h}$ or $\ell' \neq \tilde{\ell}$.

The optimality equations for the smaller dimensional belief vectors $\boldsymbol{\pi} \in \Pi(\Psi)$ are obtained by replacing $\boldsymbol{\eta}$ by $(h, l, \boldsymbol{\pi})$:

$$v(h,\ell,\boldsymbol{\pi}) \;=\; \max\Big\{ r_T(h,\ell); \qquad r_W(h) +$$

$$\lambda \cdot \sum_{k} \sum_{h'} \sum_{\ell'} \sum_{k'} \sum_{y'} \pi_k \cdot P\{(h',\ell',k')|(h,\ell,k)\}\cdot \mathcal{O}\{y'|k'\}\cdot v\left(h',\ell',\boldsymbol{\pi}'\right)\Big\} \tag{4.14}$$

Note that, in optimality equation (4.14), the new belief vector $\boldsymbol{\pi}'$ is a function of the observation (h',ℓ',y') and the current belief vector $\boldsymbol{\pi}$, but this dependency is suppressed for notational clarity.

4.3 STRUCTURAL PROPERTIES

In this section, we derive some structural properties of the model developed in Section 4.2.

The following corollaries follow directly from the results in Section 3.3 and have identical interpretations as their analogs in Section 3.3.

Corollary 4.1. *$v(h,\ell,\boldsymbol{\pi})$ is monotonically nonincreasing in ℓ for any given $h \in \Omega$ and $\boldsymbol{\pi} \in \Pi(\Psi)$.*

Corollary 4.2. *There exists a liver-based control-limit optimal policy for all $h \in \Omega$ and $\boldsymbol{\pi} \in \Pi(\Psi)$.*

Corollary 4.3. *If $\hat{\mathcal{H}}$ is IFR, then $v(h,\ell,\boldsymbol{\pi})$ is monotonically nonincreasing in h for any $\ell \in \Phi$ and $\boldsymbol{\pi} \in \Pi(\Psi)$.*

Corollary 4.4. *If $\hat{\mathcal{H}}$ is IFR and $\mathcal{H}$ and $r_T(h,\ell)$ satisfy the following:*

$$\sum_{h'=j}^{H} \mathcal{H}\{h'|h\} \leq \sum_{h'=j}^{H} \mathcal{H}\{h'|h+1\} \text{ for } j = h+1,\ldots,H, \text{ and } h = 1,\ldots,H, \tag{4.15}$$

and for any given ℓ

$$\frac{r_T(h,\ell) - r_T(h+1,\ell)}{r_T(h,\ell)} \leq \lambda\Big[\mathcal{H}\{H+1|h+1\} - \mathcal{H}\{H+1|h\}\Big] \text{ for } h = 1,\ldots,H-1, \tag{4.16}$$

then there exists a health-based control-limit optimal policy for all $\ell \in \Phi$ and $\boldsymbol{\pi} \in \Pi(\Psi)$.

Now our goal is to prove the monotonicity of the value function in $\boldsymbol{\pi}$ for some fixed $h \in \Omega$ and $\ell \in \Phi$. Since $\boldsymbol{\pi}$'s do not necessarily have a natural order, we need to first find a way to order these belief states.

We start with the definition of monotone likelihood ratio order [51, 88, 172], which is used in comparing probability distributions, and the definition of a totally positive matrix [82], which is frequently used in reliability theory. One limitation of the monotone-likelihood ratio order is that it is not exhaustive (i.e., we may not be able to order all the points in a probability simplex using this order, rather it will be only a subset of probability distributions that can be ordered using this partial order).

Definition 4.2 (Ferguson [51])**.** *Let $\mathbb{X}$ be a completely ordered finite set with n elements. Given two vectors $\boldsymbol{\pi}^1$ and $\boldsymbol{\pi}^2$ in $\Pi(\mathbb{X})$, $\boldsymbol{\pi}^1$ is greater than $\boldsymbol{\pi}^2$ in monotone likelihood ratio (MLR) partial order, denoted $\boldsymbol{\pi}^1 \succeq_{LR} \boldsymbol{\pi}^2$, if*

$$\pi^1_i \pi^2_{i'} \geq \pi^1_{i'} \pi^2_i \quad \text{for } i \geq i' \text{ in } \mathbb{X}.$$

Definition 4.3 (Karlin [82])**.** *A transition probability matrix $P = [p_{ij}]$ is Totally Positive of order 2 (TP_2) if*

$$p_{ij} p_{i'j'} \geq p_{ij'} p_{i'j} \quad \text{for } i \leq i', j \leq j'.$$

Lovejoy [88] also studied the structural properties of a general POMDP model and derived conditions under which the optimal value function is monotone with respect to MLR-ordered belief states. He further presented conditions under which the optimal policy is monotone for the machine replacement problem presented in White [167] and conditions under which the optimal policy of a general POMDP problem is bounded by a monotone function. Our results differ from his results in several aspects. First, our model consists of a hybrid state space, whereas his model is only composed of partially observed state space. Second, the observations as well as the liver offer probabilities in our model provides some information to the decision maker as indicated by equation (4.11) and discussion surrounding it, whereas the decision maker in Lovejoy's model is only informed by the observations. As a result, in proving monotonicity of the value function, the conditions in [88] require the entire core state transition probability matrix to be TP_2, whereas we only require the $\mathcal{K}$ matrix to be TP_2. We find that our conditions will be sufficient for the conditions presented in [88]. Finally, the conditions presented in [88] for the existence of a monotone function that provides a bound for the optimal policy are more strict than the conditions that guarantee monotone value function, whereas we show that the same set of conditions guarantee monotonicity for the value function and the optimal policy.

Lemma 4.1 shows that, when the rank transition probability matrix, $\mathcal{K}$, is TP_2, the belief states preserve the MLR-partial order upon conditioning on the new information.

Lemma 4.1. *If $\mathcal{K}$ is TP_2, then $\boldsymbol{\pi}^1 \succeq_{LR} \boldsymbol{\pi}^2$ in $\Pi(\Psi)$ implies $\boldsymbol{\pi}'(\boldsymbol{\pi}^1) \succeq_{LR} \boldsymbol{\pi}'(\boldsymbol{\pi}^2)$.*

Proof. For convenience, we reiterate the belief state updating formula (4.11) for the given belief vectors $\boldsymbol{\pi}^1$ and $\boldsymbol{\pi}^2$ in $\Pi(\Psi)$.

$$\pi'_{\tilde{k}}(\boldsymbol{\pi}^i) = \frac{P\{\tilde{y}|\tilde{k}\} \cdot \mathcal{L}\{\tilde{\ell}|\tilde{k}\} \cdot \sum_k \mathcal{K}\{\tilde{k}|k\} \cdot \pi^i_k}{\sum_{k'} P\{\tilde{y}|k'\} \cdot \mathcal{L}\{\tilde{\ell}|k'\} \cdot \sum_k \mathcal{K}\{k'|k\} \cdot \pi^i_k} \qquad i = 1, 2. \tag{4.17}$$

If $\boldsymbol{\pi}^1 \succeq_{LR} \boldsymbol{\pi}^2$, then, by definition,

$$\pi^1_j \cdot \pi^2_i - \pi^1_i \cdot \pi^2_j \geq 0 \quad \text{for } j \geq i \text{ in } \Psi. \tag{4.18}$$

If $\mathcal{K}$ is TP_2, then, by definition,

$$\mathcal{K}\{k_1|j\} \cdot \mathcal{K}\{k_2|i\} - \mathcal{K}\{k_2|j\} \cdot \mathcal{K}\{k_1|i\} \geq 0 \quad \text{for } k_1 \geq k_2, \quad j \geq i \text{ in } \Psi. \tag{4.19}$$

Inequalities (4.18) and (4.19) together imply, for $k_1 \geq k_2$ in Ψ,

$$\begin{aligned}
0 \;\leq\; & \sum_{i=1}^{K}\sum_{j=i}^{K}\left[\mathcal{K}\{k_1|j\}\cdot\mathcal{K}\{k_2|i\}-\mathcal{K}\{k_2|j\}\cdot\mathcal{K}\{k_1|i\}\right]\left[\pi_j^1\cdot\pi_i^2-\pi_i^1\cdot\pi_j^2\right] \\
= \; & \sum_{i=1}^{K}\sum_{j=i}^{K}\mathcal{K}\{k_1|j\}\mathcal{K}\{k_2|i\}\pi_j^1\pi_i^2+\sum_{i=1}^{K}\sum_{j=i}^{K}\mathcal{K}\{k_2|j\}\mathcal{K}\{k_1|i\}\pi_i^1\pi_j^2 \\
& -\sum_{i=1}^{K}\sum_{j=i}^{K}\mathcal{K}\{k_1|j\}\mathcal{K}\{k_2|i\}\pi_i^1\pi_j^2-\sum_{i=1}^{K}\sum_{j=i}^{K}\mathcal{K}\{k_2|j\}\mathcal{K}\{k_1|i\}\pi_j^1\pi_i^2 \\
= \; & \sum_{i=1}^{K}\sum_{j=i}^{K}\mathcal{K}\{k_1|j\}\mathcal{K}\{k_2|i\}\pi_j^1\pi_i^2+\sum_{m=1}^{K}\sum_{n=1}^{m}\mathcal{K}\{k_2|m\}\mathcal{K}\{k_1|n\}\pi_n^1\pi_m^2 \\
& -\sum_{i=1}^{K}\sum_{j=i}^{K}\mathcal{K}\{k_1|j\}\mathcal{K}\{k_2|i\}\pi_i^1\pi_j^2-\sum_{m=1}^{K}\sum_{n=1}^{m}\mathcal{K}\{k_2|m\}\mathcal{K}\{k_1|n\}\pi_m^1\pi_n^2 \qquad (4.20) \\
= \; & \sum_{i=1}^{K}\sum_{j=1}^{K}\mathcal{K}\{k_1|j\}\mathcal{K}\{k_2|i\}\pi_j^1\pi_i^2-\sum_{i=1}^{K}\sum_{j=1}^{K}\mathcal{K}\{k_1|j\}\mathcal{K}\{k_2|i\}\pi_i^1\pi_j^2 \\
= \; & \left(\sum_{j=1}^{K}\mathcal{K}\{k_1|j\}\pi_j^1\right)\left(\sum_{i=1}^{K}\mathcal{K}\{k_2|i\}\pi_i^2\right)-\left(\sum_{j=1}^{K}\mathcal{K}\{k_1|j\}\pi_j^2\right)\left(\sum_{i=1}^{K}\mathcal{K}\{k_2|i\}\pi_i^1\right),
\end{aligned}$$

where we have changed the indices of the summations for the second and fourth terms in equation (4.20) and used algebraic manipulations in the rest.

To have $\boldsymbol{\pi}'(\boldsymbol{\pi}^1) \succeq_{LR} \boldsymbol{\pi}'(\boldsymbol{\pi}^2)$, the following must hold for $k_1 \geq k_2$ in Ψ:

$$\begin{aligned}
& \left[P\{\tilde{y}|k_1\}\cdot\mathcal{L}\{\tilde{\ell}|k_1\}\cdot\sum_k\mathcal{K}\{k_1|k\}\cdot\pi_k^1\right]\cdot\left[P\{\tilde{y}|k_2\}\cdot\mathcal{L}\{\tilde{\ell}|k_2\}\cdot\sum_k\mathcal{K}\{k_2|k\}\cdot\pi_k^2\right] \\
& \geq \left[P\{\tilde{y}|k_2\}\cdot\mathcal{L}\{\tilde{\ell}|k_2\}\cdot\sum_k\mathcal{K}\{k_2|k\}\cdot\pi_k^1\right]\cdot\left[P\{\tilde{y}|k_1\}\cdot\mathcal{L}\{\tilde{\ell}|k_1\}\cdot\sum_k\mathcal{K}\{k_1|k\}\cdot\pi_k^2\right],
\end{aligned}$$

which holds if

$$\left[\sum_k\mathcal{K}\{k_1|k\}\cdot\pi_k^1\right]\cdot\left[\sum_k\mathcal{K}\{k_2|k\}\cdot\pi_k^2\right]\geq\left[\sum_k\mathcal{K}\{k_2|k\}\cdot\pi_k^1\right]\cdot\left[\sum_k\mathcal{K}\{k_1|k\}\cdot\pi_k^2\right]. \qquad (4.21)$$

This completes the proof. □

Before we present our main results in this section, we also need the definition of stochastically ordered vectors and results associated with such vectors.

Definition 4.4. *Let $\mathbb{X}$ be a completely ordered finite set with n elements. Given two vectors $\boldsymbol{\pi}^1$ and $\boldsymbol{\pi}^2$ in $\Pi(\mathbb{X})$, $\boldsymbol{\pi}^1$ is stochastically greater than $\boldsymbol{\pi}^2$, denoted $\boldsymbol{\pi}^1 \succeq_S \boldsymbol{\pi}^2$, if*

$$\sum_{i=k}^{n} \pi_i^1 \geq \sum_{i=k}^{n} \pi_i^2 \quad \text{for } k = 1, \ldots, n.$$

An important result for stochastically ordered vectors is given in Lemma 4.2.

Lemma 4.2. *If $\mathbb{X}$ is a countable, completely ordered set, and $\boldsymbol{\pi}^1$ and $\boldsymbol{\pi}^2$ are elements of $\Pi(\mathbb{X})$, then $\boldsymbol{\pi}^1 \succeq_S \boldsymbol{\pi}^2$ implies $\sum_{i=1}^{n} \pi_i^1 f_i \leq \sum_{i=1}^{n} \pi_i^2 f_i$ for every nonincreasing sequence $\{f_i\}_{i \in \mathbb{X}}$.*

Proof. Define $f_{n+1} \equiv 0$. Then

$$\begin{aligned} \sum_{i=1}^{n} f_i \cdot \pi_i^1 &= \sum_{i=0}^{n-1} [f_{n-i} - f_{n-i+1}] \cdot \sum_{j=1}^{n-i} \pi_j^1 \\ &\leq \sum_{i=0}^{n-1} [f_{n-i} - f_{n-i+1}] \cdot \sum_{j=1}^{n-i} \pi_j^2 \qquad (4.22) \\ &= \sum_{i=1}^{n} f_i \cdot \pi_i^2 \end{aligned}$$

where inequality (4.22) follows from the definition of stochastically ordered vectors and we have used algebraic manipulations in the rest of the derivation. □

Recall the definition of an IFR transition probability matrix from Section 3.2 (Definition 3.2). If we denote the ith row of a transition probability matrix P by $\boldsymbol{\pi}^i$, then we can characterize an IFR matrix as follows: P is IFR if and only if $\boldsymbol{\pi}^i \succeq_S \boldsymbol{\pi}^j$ for all $i \geq j$. Similarly, we can characterize a TP_2 matrix as follows: P is TP_2 if and only if $\boldsymbol{\pi}^i \succeq_{LR} \boldsymbol{\pi}^j$ for all $i \geq j$.

The linkage between the stochastic order ($\succeq_S$) and MLR-order ($\succeq_{LR}$) is given in Lemma 4.3, which shows that MLR-order is a stronger condition than the stochastic order.

Lemma 4.3. *If $\boldsymbol{\pi}^1 \succeq_{LR} \boldsymbol{\pi}^2$, then $\boldsymbol{\pi}^1 \succeq_S \boldsymbol{\pi}^2$.*

Proof. See Rosenfield [119]. □

Lemma 4.4 provides sufficient conditions under which the Bayesian updates yield MLR-ordered beliefs with respect to the observation y'.

Lemma 4.4. *If $\mathcal{O}$ is TP_2, then $\boldsymbol{\pi}'(\boldsymbol{\pi}, h', l', y')$ is MLR-nondecreasing in y'.*

Proof. For any given $h' \in \Omega$ and $\ell' \in \Phi$, $\boldsymbol{\pi}'(\boldsymbol{\pi}, h', l', y')$ is MLR-nondecreasing in y' if and only if

$$\begin{aligned}
&\boldsymbol{\pi}'(\boldsymbol{\pi}, h', l', y') \succeq_{LR} \boldsymbol{\pi}'(\boldsymbol{\pi}, h', l', y'') && \text{for } y' \geq y'' \in \Upsilon \\
&\quad \Leftrightarrow \quad \pi'_i(\boldsymbol{\pi}, h', l', y') \cdot \pi'_j(\boldsymbol{\pi}, h', l', y'') \geq \pi'_j(\boldsymbol{\pi}, h', l', y') \cdot \pi'_i(\boldsymbol{\pi}, h', l', y'') && \text{for } i \geq j \in \Psi \\
&\quad \Leftrightarrow \quad \frac{\mathcal{O}\{y'|i\}\mathcal{L}\{\ell'|i\}\sum_k \mathcal{K}\{i|k\}\pi_k}{\sum_{k'} \mathcal{O}\{y'|k'\}\mathcal{L}\{\ell'|k'\}\sum_k \mathcal{K}\{k'|k\}\pi_k} \cdot \frac{\mathcal{O}\{y''|j\}\mathcal{L}\{\ell'|j\}\sum_k \mathcal{K}\{j|k\}\pi_k}{\sum_{k'} \mathcal{O}\{y''|k'\}\mathcal{L}\{\ell'|k'\}\sum_k \mathcal{K}\{k'|k\}\pi_k} \\
&\qquad \geq \frac{\mathcal{O}\{y'|j\}\mathcal{L}\{\ell'|j\}\sum_k \mathcal{K}\{j|k\}\pi_k}{\sum_{k'} \mathcal{O}\{y'|k'\}\mathcal{L}\{\ell'|k'\}\sum_k \mathcal{K}\{k'|k\}\pi_k} \cdot \frac{\mathcal{O}\{y''|i\}\mathcal{L}\{\ell'|i\}\sum_k \mathcal{K}\{i|k\}\pi_k}{\sum_{k'} \mathcal{O}\{y''|k'\}\mathcal{L}\{\ell'|k'\}\sum_k \mathcal{K}\{k'|k\}\pi_k} \\
&\quad \Leftrightarrow \quad \mathcal{O}\{y'|i\} \cdot \mathcal{O}\{y''|j\} \geq \mathcal{O}\{y'|j\} \cdot \mathcal{O}\{y''|i\} && \text{for } i \geq j \in \Psi,
\end{aligned}$$

which holds if $\mathcal{O}$ is TP_2. □

Lemma 4.5 provides sufficient conditions under which the Bayesian updates yield MLR-ordered beliefs with respect to the observation ℓ'.

Lemma 4.5. *If $\mathcal{L}\{\ell, k\}$ is monotonically nonincreasing in $k \in \Psi$ for all $\ell \neq L+1$, then, for any $h' \in \Omega$ and $y' \in \Upsilon$,*

$$\boldsymbol{\pi}'(\boldsymbol{\pi}, h', L+1, y') \succeq_{LR} \boldsymbol{\pi}'(\boldsymbol{\pi}, h', \ell', y') \qquad for \quad \ell' \neq L+1.$$

Proof. For any given $h' \in \Omega$, $y' \in \Upsilon$, and $\ell \neq L+1$,

$$\begin{aligned}
&\boldsymbol{\pi}'(\boldsymbol{\pi}, h', L+1, y') \succeq_{LR} \boldsymbol{\pi}'(\boldsymbol{\pi}, h', \ell', y') \\
&\quad \Leftrightarrow \quad \boldsymbol{\pi}'_i(\boldsymbol{\pi}, h', L+1, y') \cdot \boldsymbol{\pi}'_j(\boldsymbol{\pi}, h', \ell', y') \succeq_{LR} \boldsymbol{\pi}'_j(\boldsymbol{\pi}, h', L+1, y') \cdot \boldsymbol{\pi}'_i(\boldsymbol{\pi}, h', \ell', y') \text{ for } i \geq j \in \Psi \\
&\quad \Leftrightarrow \quad \frac{\mathcal{O}\{y'|i\}\mathcal{L}\{L+1|i\}\sum_k \mathcal{K}\{i|k\}\pi_k}{\sum_{k'} \mathcal{O}\{y'|k'\}\mathcal{L}\{L+1|k'\}\sum_k \mathcal{K}\{k'|k\}\pi_k} \cdot \frac{\mathcal{O}\{y'|j\}\mathcal{L}\{\ell'|j\}\sum_k \mathcal{K}\{j|k\}\pi_k}{\sum_{k'} \mathcal{O}\{y'|k'\}\mathcal{L}\{\ell'|k'\}\sum_k \mathcal{K}\{k'|k\}\pi_k} \\
&\qquad \geq \frac{\mathcal{O}\{y'|j\}\mathcal{L}\{L+1|j\}\sum_k \mathcal{K}\{j|k\}\pi_k}{\sum_{k'} \mathcal{O}\{y'|k'\}\mathcal{L}\{L+1|k'\}\sum_k \mathcal{K}\{k'|k\}\pi_k} \cdot \frac{\mathcal{O}\{y'|i\}\mathcal{L}\{\ell'|i\}\sum_k \mathcal{K}\{i|k\}\pi_k}{\sum_{k'} \mathcal{O}\{y'|k'\}\mathcal{L}\{\ell'|k'\}\sum_k \mathcal{K}\{k'|k\}\pi_k} \\
&\quad \Leftrightarrow \quad \mathcal{L}\{L+1|i\} \cdot \mathcal{L}\{\ell'|j\} \geq \mathcal{L}\{L+1|j\}\mathcal{L}\{\ell'|i\} && \text{for } i \geq j \in \Psi,
\end{aligned}$$

which is always true since

$$\frac{\mathcal{L}\{\ell'|j\}}{\mathcal{L}\{\ell'|i\}} \geq 1 \geq \frac{\mathcal{L}\{L+1|j\}}{\mathcal{L}\{L+1|i\}} \qquad \text{for } i \geq j \in \Psi.$$

This completes the proof. □

Lemma 4.6. *For any given $k' \in \Psi$, let*

$$A_{k'} = \sum_{h'} \sum_{\ell'} \sum_{y'} \mathcal{H}\{h'|h\}\mathcal{L}\{\ell'|k'\}\mathcal{O}\{y'|k'\}v^i\left(h', \ell', \boldsymbol{\pi}'(\boldsymbol{\pi}, h', \ell', y')\right).$$

If $\mathcal{O}$ is TP_2, $\mathcal{L}\{\ell, k\}$ is monotonically nonincreasing in $k \in \Psi$ for all $\ell \neq L+1$, and $v^i\left(h', \ell', \boldsymbol{\pi}'(\boldsymbol{\pi})\right)$ is MLR-nonincreasing in $\boldsymbol{\pi}$, then $A_{k'}$ is nonincreasing in k'.

Proof. For $k' \leq k'' \in \Psi$,

$$\begin{aligned}
&A_{k'} - A_{k''} \\
&= \sum_{h'} \sum_{\ell'} \sum_{y'} \mathcal{H}\{h'|h\}\mathcal{L}\{\ell'|k'\}\mathcal{O}\{y'|k'\}v^i\left(h', \ell', \boldsymbol{\pi}'(\boldsymbol{\pi}, h', \ell', y')\right) \\
&\quad - \sum_{h'} \sum_{\ell'} \sum_{y'} \mathcal{H}\{h'|h\}\mathcal{L}\{\ell'|k''\}\mathcal{O}\{y'|k''\}v^i\left(h', \ell', \boldsymbol{\pi}'(\boldsymbol{\pi}, h', \ell', y')\right) \\
&= \sum_{h'} \sum_{\ell' \neq L+1} \sum_{y'} \mathcal{H}\{h'|h\}\mathcal{L}\{\ell'|k'\}\mathcal{O}\{y'|k'\}v^i\left(h', \ell', \boldsymbol{\pi}'(\boldsymbol{\pi}, h', \ell', y')\right) \\
&\quad + \sum_{h'} \sum_{y'} \mathcal{H}\{h'|h\}\mathcal{O}\{y'|k'\}v^i\left(h', L+1, \boldsymbol{\pi}'(\boldsymbol{\pi}, h', L+1, y')\right) \\
&\quad - \sum_{h'} \sum_{\ell' \neq L+1} \sum_{y'} \mathcal{H}\{h'|h\}\mathcal{L}\{\ell'|k'\}\mathcal{O}\{y'|k'\}v^i\left(h', L+1, \boldsymbol{\pi}'(\boldsymbol{\pi}, h', L+1, y')\right) \\
&\quad - \sum_{h'} \sum_{\ell' \neq L+1} \sum_{y'} \mathcal{H}\{h'|h\}\mathcal{L}\{\ell'|k''\}\mathcal{O}\{y'|k''\}v^i\left(h', \ell', \boldsymbol{\pi}'(\boldsymbol{\pi}, h', \ell', y')\right) \\
&\quad - \sum_{h'} \sum_{y'} \mathcal{H}\{h'|h\}\mathcal{O}\{y'|k''\}v^i\left(h', L+1, \boldsymbol{\pi}'(\boldsymbol{\pi}, h', L+1, y')\right) \\
&\quad + \sum_{h'} \sum_{\ell' \neq L+1} \sum_{y'} \mathcal{H}\{h'|h\}\mathcal{L}\{\ell'|k''\}\mathcal{O}\{y'|k''\}v^i\left(h', L+1, \boldsymbol{\pi}'(\boldsymbol{\pi}, h', L+1, y')\right) \qquad (4.23) \\
&= \sum_{h'} \sum_{\ell' \neq L+1} \sum_{y'} \mathcal{H}\{h'|h\}v^i\left(h', \ell', \boldsymbol{\pi}'(\boldsymbol{\pi}, h', \ell', y')\right)\left[\mathcal{L}\{\ell'|k'\}\mathcal{O}\{y'|k'\} - \mathcal{L}\{\ell'|k''\}\mathcal{O}\{y'|k''\}\right] \\
&\quad + \sum_{h'} \sum_{y'} \mathcal{H}\{h'|h\}v^i\left(h', L+1, \boldsymbol{\pi}'(\boldsymbol{\pi}, h', L+1, y')\right)\left[\mathcal{O}\{y'|k'\} - \mathcal{O}\{y'|k''\}\right] \\
&\quad - \sum_{h'} \sum_{\ell' \neq L+1} \sum_{y'} \mathcal{H}\{h'|h\}v^i\left(h', L+1, \boldsymbol{\pi}'(\boldsymbol{\pi}, h', L+1, y')\right) \\
&\qquad\qquad \cdot \left[\mathcal{L}\{\ell'|k'\}\mathcal{O}\{y'|k'\} - \mathcal{L}\{\ell'|k''\}\mathcal{O}\{y'|k''\}\right]
\end{aligned}$$

$$= \sum_{h'} \sum_{\ell' \neq L+1} \sum_{y'} \mathcal{H}\{h'|h\} \cdot \left[v^i \left(h', \ell', \boldsymbol{\pi}'(\boldsymbol{\pi}, h', \ell', y')\right) - v^i \left(h', L+1, \boldsymbol{\pi}'(\boldsymbol{\pi}, h', L+1, y')\right) \right]$$
$$\cdot \left[\mathcal{L}\{\ell'|k'\} \mathcal{O}\{y'|k'\} - \mathcal{L}\{\ell'|k''\} \mathcal{O}\{y'|k''\} \right]$$
$$+ \sum_{h'} \sum_{y'} \mathcal{H}\{h'|h\} \cdot v^i \left(h', L+1, \boldsymbol{\pi}'(\boldsymbol{\pi}, h', L+1, y')\right) \cdot \left[\mathcal{O}\{y'|k'\} - \mathcal{O}\{y'|k''\} \right], \quad (4.24)$$

where we have used the identity $\mathcal{L}\{L+1|k'\} = \sum_{\ell' \neq L+1} \mathcal{L}\{\ell'|k'\}$ in writing equation (4.23).

Since $\mathcal{O}$ is TP_2 by assumption, Lemma 4.4 implies that $\boldsymbol{\pi}'(\boldsymbol{\pi}, h', L+1, y')$ is MLR-nondecreasing in y', which along with the monotonicity assumption of $v^i(\cdot)$ implies that $v^i \left(h', L+1, \boldsymbol{\pi}'(\boldsymbol{\pi}, h', L+1, y')\right)$ is MLR-nonincreasing in y'. This result taken together with $\mathcal{O}$ being TP_2 and $\mathcal{H}\{h'|h\} \geq 0$ for all h, h' yields, by Lemmas 4.2 and 4.3, for $k' \leq k'' \in \Psi$,

$$\sum_{h'} \sum_{y'} \mathcal{H}\{h'|h\} \cdot v^i \left(h', L+1, \boldsymbol{\pi}'(\boldsymbol{\pi}, h', L+1, y')\right) \cdot \left[\mathcal{O}\{y'|k'\} - \mathcal{O}\{y'|k''\} \right] \geq 0. \quad (4.25)$$

Using similar arguments along with Lemmas 4.4, 4.5, and Corollary 4.1, we find

$$\sum_{h'} \sum_{\ell' \neq L+1} \sum_{y'} \mathcal{H}\{h'|h\} \cdot \left[v^i \left(h', \ell', \boldsymbol{\pi}'(\boldsymbol{\pi}, h', \ell', y')\right) - v^i \left(h', L+1, \boldsymbol{\pi}'(\boldsymbol{\pi}, h', L+1, y')\right) \right]$$
$$\cdot \left[\mathcal{L}\{\ell'|k'\} \mathcal{O}\{y'|k'\} - \mathcal{L}\{\ell'|k''\} \mathcal{O}\{y'|k''\} \right] \geq 0. \quad (4.26)$$

Combining inequalities (4.25) and (4.26) into (4.24) yields

$$A_{k'} - A_{k''} \geq 0 \qquad \text{for} \quad k' \leq k'' \in \Psi,$$

which completes the proof. □

Lemma 4.7. *If $\mathcal{K}$ and $\mathcal{O}$ are TP_2, $\mathcal{L}\{\ell, k\}$ is monotonically nonincreasing in $k \in \Psi$ for all $\ell \neq L+1$, and $v^i \left(h', \ell', \boldsymbol{\pi}'(\boldsymbol{\pi})\right)$ is MLR-nonincreasing in $\boldsymbol{\pi}$, then $\sum_{k'} \mathcal{K}\{k'|k\} A_{k'}$ is nonincreasing in k.*

Proof. Since $\mathcal{K}$ is TP_2, its rows are in increasing MLR-order. Furthermore, $A_{k'}$ is nonincreasing in k', by Lemma 4.6, which along with Lemmas 4.2 and 4.3 implies

$$\sum_{k'} \mathcal{K}\{k'|i\} A_{k'} \leq \sum_{k'} \mathcal{K}\{k'|j\} A_{k'} \qquad \text{for } i \geq j \in \Psi. \quad (4.27)$$

This completes the proof. □

Our main result on the structure of the value function is given in Theorem 4.1, which states that the optimal value function is monotone with respect to the belief states that are in MLR order if the rank transition probability matrix is TP_2.

Theorem 4.1. *If $\mathcal{K}$ and $\mathcal{O}$ are TP_2, and $\mathcal{L}\{\ell, k\}$ is monotonically nonincreasing in $k \in \Psi$ for all $\ell \neq L+1$, then $v(h, \ell, \boldsymbol{\pi})$ is MLR-nonincreasing in $\boldsymbol{\pi}$ for any $h \in \Omega$ and $\ell \in \Phi$.*

Proof. For convenience, let, for any given $h \in \Omega$ and $\ell \in \Phi$,

$$\begin{aligned} f^{i+1}(\boldsymbol{\pi}) &= r_T(h, \ell), \quad \text{and} \\ g^{i+1}(\boldsymbol{\pi}) &= r_W(h) \\ &\quad + \lambda \cdot \sum_{k} \sum_{h'} \sum_{\ell'} \sum_{k'} \sum_{y'} \pi_k \cdot P\{h', \ell', k' | h, \ell, k\} \cdot \mathcal{O}\{y' | k'\} \cdot v^i\left(h', \ell', \boldsymbol{\pi}'(\boldsymbol{\pi})\right) \end{aligned} \tag{4.28}$$

for $i = 0, 1, 2, \ldots$, and $f(\boldsymbol{\pi})$ and $g(\boldsymbol{\pi})$, be, respectively, the same things without the superscripts i and $i+1$. We can now re-write the optimality equation (4.14) as

$$v(h, \ell, \boldsymbol{\pi}) = \max\left\{f(\boldsymbol{\pi}), g(\boldsymbol{\pi})\right\}.$$

Since the maximum of nonincreasing functions is itself nonincreasing, it suffices to show that $g(\boldsymbol{\pi})$ is MLR-nonincreasing in $\boldsymbol{\pi}$, as $f(\boldsymbol{\pi})$ is trivially nonincreasing in $\boldsymbol{\pi}$. We do so by induction on the steps of the value iteration algorithm. To start with, let $v^i(h, \ell, \boldsymbol{\pi})$ be the value of state $(h, \ell, \boldsymbol{\pi})$ at the ith iteration of the algorithm and assume, without loss of generality, that $v^0(h, \ell, \boldsymbol{\pi}) = 0$ for all h, ℓ, and $\boldsymbol{\pi}$. Therefore, $g^1(\boldsymbol{\pi}) = r_W(h)$ for all $\boldsymbol{\pi}$, which is trivially MLR-nonincreasing in $\boldsymbol{\pi}$.

As induction hypothesis, assume $v^i(h, \ell, \boldsymbol{\pi})$ is MLR-nonincreasing in $\boldsymbol{\pi}$ for iterations $i = 1, 2, \ldots, n$. Using the result of Lemma 4.1 along with the induction hypothesis, we find, for $\boldsymbol{\pi}^1 \succeq_{LR} \boldsymbol{\pi}^2$

$$v^i\left(h', \ell', \boldsymbol{\pi}'(\boldsymbol{\pi}^1)\right) \leq v^i\left(h', \ell', \boldsymbol{\pi}'(\boldsymbol{\pi}^2)\right) \qquad \text{for } i = 1, \ldots, n. \tag{4.29}$$

Since the coefficients λ, π_k, $P\{h', \ell', k'|h, \ell, k\}$, and $\mathcal{O}\{y'|k'\}$ are all non-negative in equation (4.28), replacing $v^i\left(h', \ell', \boldsymbol{\pi}'(\boldsymbol{\pi}^1)\right)$ with $v^i\left(h', \ell', \boldsymbol{\pi}'(\boldsymbol{\pi}^2)\right)$ for $g^{i+1}(\boldsymbol{\pi}^1)$ we obtain

$$
\begin{aligned}
& g^{i+1}\left(\boldsymbol{\pi}^1\right) \\
& \leq \; r_W(h) + \lambda \cdot \sum_k \sum_{h'} \sum_{\ell'} \sum_{k'} \sum_{y'} \pi_k^1 \cdot P\{h', \ell', k'|h, \ell, k\} \cdot \mathcal{O}\{y'|k'\} \cdot v^n\left(h', \ell', \boldsymbol{\pi}'(\boldsymbol{\pi}^2)\right) \\
& = \; r_W(h) + \lambda \cdot \sum_k \pi_k^1 \cdot \sum_{h'} \sum_{\ell'} \sum_{k'} \sum_{y'} P\{h', \ell', k'|h, \ell, k\} \cdot \mathcal{O}\{y'|k'\} \cdot v^n\left(h', \ell', \boldsymbol{\pi}'(\boldsymbol{\pi}^2)\right) \\
& = \; r_W(h) + \lambda \sum_k \pi_k^1 \sum_{k'} \mathcal{K}\{k'|k\} \sum_{h'} \sum_{\ell'} \sum_{y'} \mathcal{H}\{h'|h\} \mathcal{L}\{\ell'|k'\} \mathcal{O}\{y'|k'\} v^n\left(h', \ell', \boldsymbol{\pi}'(\boldsymbol{\pi}^2)\right) \\
& = \; r_W(h) + \lambda \cdot \sum_k \pi_k^1 \cdot \sum_{k'} \mathcal{K}\{k'|k\} \cdot A_{k'} \qquad (4.30)
\end{aligned}
$$

where

$$
A_{k'} = \sum_{h'} \sum_{\ell'} \sum_{y'} \mathcal{H}\{h'|h\} \mathcal{L}\{\ell'|k'\} \mathcal{O}\{y'|k'\} v^n\left(h', \ell', \boldsymbol{\pi}'(\boldsymbol{\pi}^2)\right). \qquad (4.31)
$$

But $\sum_{k'} \mathcal{K}\{k'|k\} A_{k'}$ is nonincreasing in k, by Lemma 4.7, which along with Lemmas 4.2 and 4.3 implies

$$
\sum_k \pi_k^1 \cdot \sum_{k'} \mathcal{K}\{k'|k\} \cdot A_{k'} \leq \sum_k \pi_k^2 \cdot \sum_{k'} \mathcal{K}\{k'|k\} \cdot A_{k'} \qquad \text{for } \boldsymbol{\pi}^1 \succeq_{LR} \boldsymbol{\pi}^2. \qquad (4.32)
$$

Using inequality (4.32) in (4.30), we find, for $\boldsymbol{\pi}^1 \succeq_{LR} \boldsymbol{\pi}^2$,

$$
\begin{aligned}
g^{i+1}\left(\boldsymbol{\pi}^1\right) \; & \leq \; r_W(h) + \lambda \cdot \sum_k \pi_k^2 \cdot \sum_{k'} \mathcal{K}\{k'|k\} \cdot A_{k'} \\
& = \; g^{i+1}\left(\boldsymbol{\pi}^2\right) \qquad \text{for } i = 1, \ldots, n,
\end{aligned}
$$

which implies

$$
v^{n+1}\left(h, \ell, \boldsymbol{\pi}^1\right) \leq v^{n+1}\left(h, \ell, \boldsymbol{\pi}^2\right).
$$

This completes the proof. □

Our main result on the structure of the optimal policy is given in Theorem 4.2, which proves the existence of a control-limit optimal policy among MLR-ordered belief states.

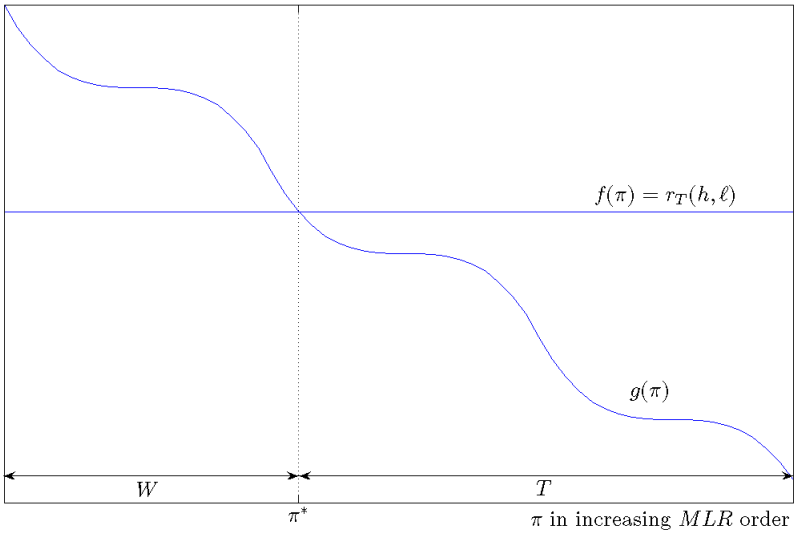


Figure 4.1: Control-limit optimal policy with respect to MLR-ordered belief vectors

Theorem 4.2. *If $\mathcal{K}$ and $\mathcal{O}$ are TP_2, and $\mathcal{L}\{\ell, k\}$ is monotonically nonincreasing in $k \in \Psi$ for all $\ell \neq L+1$, then there exists a threshold belief vector $\boldsymbol{\pi}^*$ such that the optimal action*

$$a^* = \begin{cases} W & \text{if } \boldsymbol{\pi} \preceq_{LR} \boldsymbol{\pi}^*, \\ T & \text{if } \boldsymbol{\pi} \succ_{LR} \boldsymbol{\pi}^*. \end{cases}$$

Proof. The result follows (See Figure 4.1) since $f(\boldsymbol{\pi})$ is constant for all $\boldsymbol{\pi}$ and $g(\boldsymbol{\pi})$ is MLR-nonincreasing in $\boldsymbol{\pi}$ by the result of Theorem 4.1. □

4.4 COMPUTATIONAL STUDY

In this section, we discuss and present the results of a grid-based computational study for the model given in Section 4.2. For this purpose, we only need to estimate the observation probability matrix to solve the POMDP model, since the rest of the parameters can be taken from the input data to the EWLM.

4.4.1 Estimating the observation probabilities

We have noted in Section 4.2 that the observation probabilities are indeed time-dependent due to the dynamic behavior of the liver transplant waiting list. However, to overcome the numerical problems associated with this non-stationarity, we will use time-homogeneous observation probabilities in our computational study.

We assume that the liver allocation system is in steady-state and let Θ be the random variable that denotes the stationary observation probability matrix. Then the dynamic behavior of the waiting list can be modeled by taking the expected value of Θ. We estimate this expectation using the national liver simulation model of Shechter et al. [132], in which we keep track of all the MELD partitions and take a long-run average of the observation probability matrices induced by these partitions as

$$\mathcal{O} = \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \mathcal{O}_t,$$

where $\mathcal{O}_t$ is the observation probability matrix at time t, $\mathcal{O}$ is the time-homogeneous observation probability matrix, and the average is taken as the usual component-wise matrix operation. Since OPOs represent different populations, we estimate OPO specific observation matrices, however, for notational convenience we drop the dependency on OPOs as we did in Section 3.4.1 in all of our discussions.

As an example of this estimation process, assume that there are 65 patients registered in a particular OPO at some arbitrary point in time. Let the MELD partitions be as given in the UNOS's website, i.e., $\Omega_1 = \{6, \ldots, 10\}$, $\Omega_2 = \{11, \ldots, 18\}$, $\Omega_3 = \{19, \ldots, 24\}$, and $\Omega_4 = \{25, \ldots, 40\}$. For a particular type of locally harvested organ, assume the Y_i's appear as:

$$Y_1 = \{1, \ldots, 10\}, \quad Y_2 = \{11, \ldots, 18\}, \quad Y_3 = \{19, \ldots, 35\}, \quad Y_4 = \{36, \ldots, 65\}.$$

This partition will induce an observation probability matrix, say $\mathcal{O}_1$, as in Figure 4.2(a). Let's further assume that from this time period to the next, the patient at the top list dies, the 2nd patient from the top of the list is transplanted and removed from the list, 1 new patient joined the list with a MELD score less than 10, the patient with rank 20 suddenly got severely sick and her MELD score jumped to 32, and 4 of the patients with MELD scores

less than 10 now have MELD score 15. Therefore, the Y_i's in the next time period would appear as:

$$Y_1 = \{1, \ldots, 9\}, \quad Y_2 = \{10, \ldots, 17\}, \quad Y_3 = \{18, \ldots, 37\}, \quad Y_4 = \{38, \ldots, 64\}.$$

This new partition will induce an observation probability matrix, say $\mathcal{O}_2$, as in Figure 4.2(b). Over these two time periods, we would estimate the time-homogeneous observation probability matrix by $\mathcal{O} = (\mathcal{O}_1 + \mathcal{O}_2)/2$, which would be as in Figure 4.2(c).

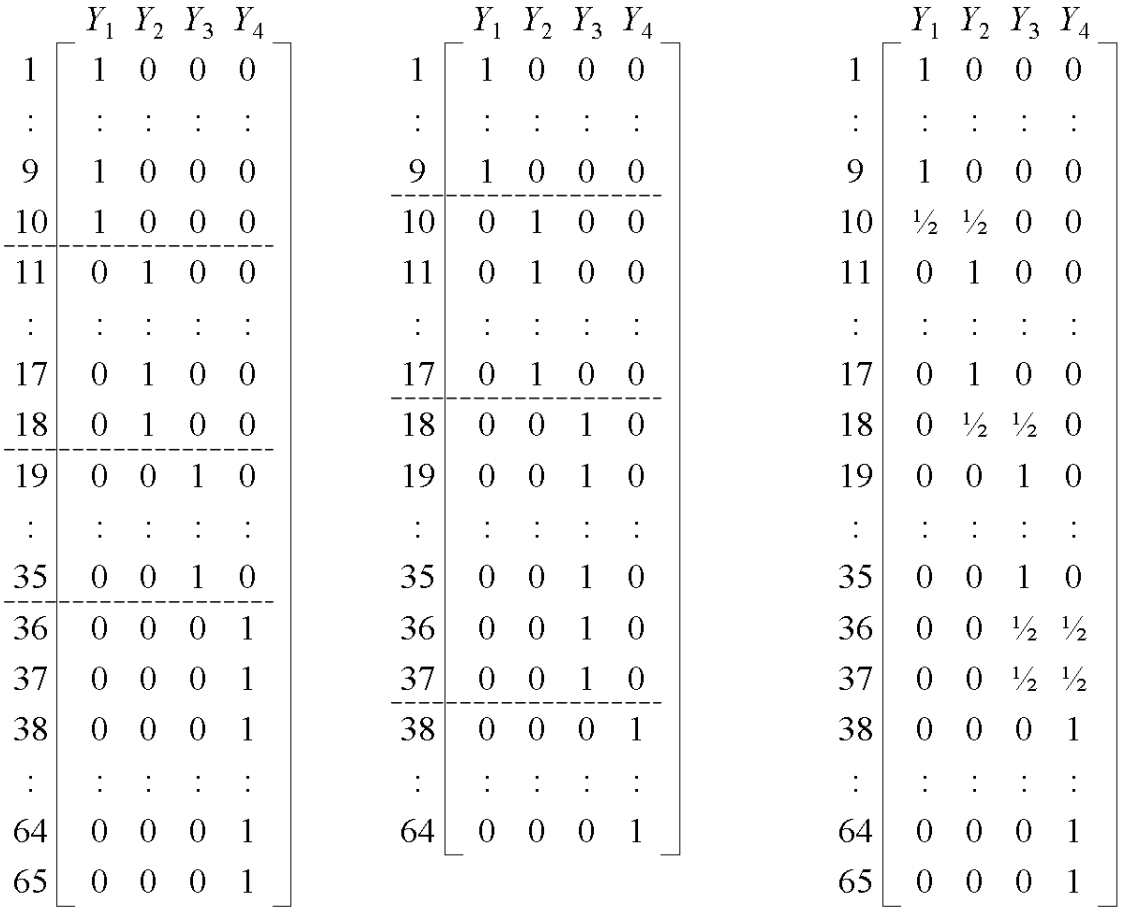
4.4.2 Grid-based solution methodology

The POMDP model has infinitely many states since the belief space is the entire probability simplex over the core states. Evaluating this belief space in its entirety suffers from computational intractability. Indeed, the current state of the art in POMDP literature can exactly solve problems with only a few core states [58, 91]. Therefore, we provide a grid-based solution methodology that samples from the entire belief space to approximate the optimal solution to our problem. This approach comes with two natural questions: (*i*) how to build a good grid, and (*ii*) how good is the resulting approximation.

4.4.2.1 Building the grid In this section, we discuss three alternative ways of building our grid.

One obvious way is to use what the literature refers to as *a fixed-resolution uniform grid* approach [89, 183]. This approach fixes the number of possible probability values that each dimension of the sampled belief vector can take. The obvious disadvantage with this approach is the need for an enormous number of sampled points as the dimension of the probability simplex grows. To see this growth, let

$$\begin{aligned} k &= \#\text{ of possible probability values in each dimension of the sampled belief vector} \\ m &= \#\text{ of core states, and} \\ F(m,k) &= \#\text{ of belief vectors to be sampled for given } m \text{ and } k. \end{aligned}$$



(a) $\mathcal{O}_1$ **(b)** $\mathcal{O}_2$ **(c)** Final observation matrix

Figure 4.2: An example of estimating time-homogeneous observation probabilities

Once m and k are given, the total number of belief vectors to be sampled can be determined by the following recursive formula

$$F(m,k) = F(m,k-1) + F(m-1,k)$$

with the boundary conditions

$$\begin{aligned} F(m,1) &= 1, \\ F(2,k) &= k. \end{aligned}$$

Possible values for m and k are positive integers starting from 2. Table 4.3 lists the number of belief vectors to sample under the uniform grid approach for $m \leq 30$ and Figure 4.3 displays the same data for $m \leq 100$. Although generating such a grid is very easy, the need for an enormous number of sample points forces us to search for alternative grid generation ideas.

A second approach for generating a grid is to generate a *uniform random grid* from the probability simplex. The obvious advantage of this approach is that we have the control of choosing how many points to sample, whereas the sample size was dictated by the resolution parameter (k) in the first approach.

When the number of core states (m) of the POMDP model is 2, this sampling can be done through generating Uniform[0,1] random numbers. However, when $m \geq 3$, uniform sampling from the probability simplex turns out to be nontrivial. The naive approach of sampling n Uniform[0,1] numbers (denoted u_i for $i = 1, \ldots, n$) and then dividing each by the sum of the sampled values ($\sum_i u_i$) does not generate a uniform sample over the probability space [26]. This approach produces points mostly from the center of the simplex and gradually less points from corners and borders of the simplex. Figure 4.4(a) shows 20,000 points generated using this naive approach when $m = 3$. The main reason for such a behavior is summarized in the following observation: "if one of u_i ($i = 1, \ldots, n$) is large (say, for instance, close to 1), then the probability that $\sum_i u_i$ exceeds 1 is very high. Once this happens, normalization causes each u_i for the given sample point to be decreased by a fraction of $\sum_i u_i$." Similar observation holds for the symmetric case ($\sum_i u_i < 1$), for which each u_i are inflated by a

Table 4.3: Number of belief vectors to sample using the uniform grid approach

		k									
		2	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
m	**2**	2	3	4	5	6	7	8	9	10	11
	3	3	6	10	15	21	28	36	45	55	66
	4	4	10	20	35	56	84	120	165	220	286
	5	5	15	35	70	126	210	330	495	715	1,001
	6	6	21	56	126	252	462	792	1,287	2,002	3,003
	7	7	28	84	210	462	924	1,716	3,003	5,005	8,008
	8	8	36	120	330	792	1,716	3,432	6,435	11,440	19,448
	9	9	45	165	495	1,287	3,003	6,435	12,870	24,310	43,758
	10	10	55	220	715	2,002	5,005	11,440	24,310	48,620	92,378
	11	11	66	286	1,001	3,003	8,008	19,448	43,758	92,378	184,756
	12	12	78	364	1,365	4,368	12,376	31,824	75,582	167,960	352,716
	13	13	91	455	1,820	6,188	18,564	50,388	125,970	293,930	646,646
	14	14	105	560	2,380	8,568	27,132	77,520	203,490	497,420	1,144,066
	15	15	120	680	3,060	11,628	38,760	116,280	319,770	817,190	1,961,256
	16	16	136	816	3,876	15,504	54,264	170,544	490,314	1,307,504	3,268,760
	17	17	153	969	4,845	20,349	74,613	245,157	735,471	2,042,975	5,311,735
	18	18	171	1,140	5,985	26,334	100,947	346,104	1,081,575	3,124,550	8,436,285
	19	19	190	1,330	7,315	33,649	134,596	480,700	1,562,275	4,686,825	13,123,110
	20	20	210	1,540	8,855	42,504	177,100	657,800	2,220,075	6,906,900	20,030,010
	21	21	231	1,771	10,626	53,130	230,230	888,030	3,108,105	10,015,005	30,045,015
	22	22	253	2,024	12,650	65,780	296,010	1,184,040	4,292,145	14,307,150	44,352,165
	23	23	276	2,300	14,950	80,730	376,740	1,560,780	5,852,925	20,160,075	64,512,240
	24	24	300	2,600	17,550	98,280	475,020	2,035,800	7,888,725	28,048,800	92,561,040
	25	25	325	2,925	20,475	118,755	593,775	2,629,575	10,518,300	38,567,100	131,128,140
	26	26	351	3,276	23,751	142,506	736,281	3,365,856	13,884,156	52,451,256	183,579,396
	27	27	378	3,654	27,405	169,911	906,192	4,272,048	18,156,204	70,607,460	254,186,856
	28	28	406	4,060	31,465	201,376	1,107,568	5,379,616	23,535,820	94,143,280	348,330,136
	29	29	435	4,495	35,960	237,336	1,344,904	6,724,520	30,260,340	124,403,620	472,733,756
	30	30	465	4,960	40,920	278,256	1,623,160	8,347,680	38,608,020	163,011,640	635,745,396

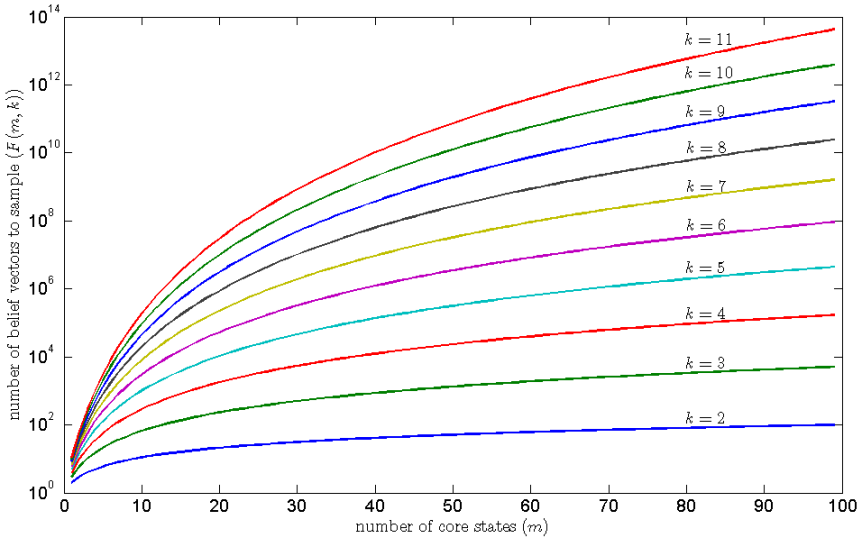
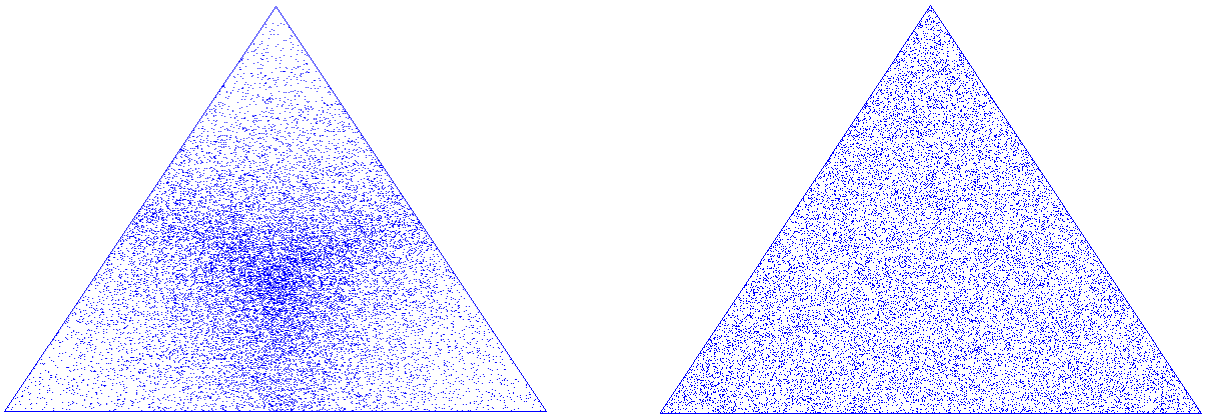


Figure 4.3: Number of belief vectors to sample using the uniform grid approach



(a) Naive algorithm

(b) Suresh's algorithm

Figure 4.4: Random sampling from the probability simplex

factor of $(\sum_i u_i)^{-1}$. Therefore, we see fewer points sampled from the vertices and edges of the probability simplex.

An algorithm for generating a uniform random sample from a probability simplex is suggested in [139]. An alternate view of the same algorithm is presented in [149]. This algorithm behaves exactly like the naive algorithm with a small exception: sample n random numbers from exponential distribution with parameter 1 and divide each by the sum of the sampled values. Figure 4.4(b) shows 20,000 points generated using this algorithm.

We are still not satisfied with the uniform random sampling approach despite its obvious advantage of controlling the number of points to be sampled. The main reason is that we do not want to sample points of little value. A sample point $\mathbf{u}$ has little value if the probability of reaching that particular point $\mathbf{u}$ is negligibly small. In other words, if the stochastic process is going to reach to this particular belief state $\mathbf{u}$ very rarely, then we do not want to spend our limited resources on this state since the contribution of this state will be negligible.

For our partially observed waiting list model given in Section 4.2, a belief point that assigns positive probabilities to rank states k_1 and k_2 ($k_1 < k_2$) while assigning zero probability for all ranks j, where $k_1 < j < k_2$, has little value, because it is practically impossible for a

patient to think that she could be in ranks k_1 and k_2 but not in between. This observation builds the backbone of our third approach for generating a grid.

The third grid generation approach, which we call the *fixed-resolution non-uniform grid*, is similar to the previous two approaches in keeping the number of grid points to be fixed over time. However, rather than sampling the points in regular intervals as in the fixed-resolution uniform grid approach or sampling the points randomly as in the random grid approach, this approach samples the points in such a way that the positive probabilities in the sampled belief vector are not interleaved with any zeros. For instance, if we have 5 core states, the vector

$$[0 \quad 1/3 \quad 1/3 \quad 1/3 \quad 0]^T$$

is a valid sample under this approach, however, the vector

$$[1/3 \quad 0 \quad 1/3 \quad 1/3 \quad 0]^T$$

is not a valid sample under this approach, whereas it is acceptable under fixed-resolution uniform grid approach.

More formally, let q be a positive integer representing the resolution of the grid and m be the number of core states. (Note that the resolution parameter q also controls the maximum number of consecutive positive entries in a sample point.) The fixed-resolution non-uniform grid approach selects vectors from the associated probability simplex that satisfy the following:

1. the strictly positive values in the vector are not interleaved by any zeros, and
2. all the values in the vector are non-negative integer multiples of $1/q$.

For instance, when $m = 5$ and $q = 3$, the grid generated by the fixed-resolution non-uniform grid approach would look like as in Figure 4.5. Note that there are a total of 16 belief points in this grid. In general, the following expression provides a formula to determine the number of sample points needed, $G(m, q)$, with this approach for arbitrary m and q:

$$G(m, q) = \sum_{i=0}^{\min\{q-1,m\}} \binom{q-1}{i} \cdot (m - i).$$

Table 4.4 lists the number of belief vectors to sample under the non-uniform grid approach for $m \leq 30$ and Figure 4.6 displays the same data for $m \leq 100$.

$$
\begin{pmatrix} 1/3 & 1/3 & 1/3 & 0 & 0 \\ 0 & 1/3 & 1/3 & 1/3 & 0 \\ 0 & 0 & 1/3 & 1/3 & 1/3 \end{pmatrix},
\begin{pmatrix} 2/3 & 1/3 & 0 & 0 & 0 \\ 0 & 2/3 & 1/3 & 0 & 0 \\ 0 & 0 & 2/3 & 1/3 & 0 \\ 0 & 0 & 0 & 2/3 & 1/3 \\ 1/3 & 2/3 & 0 & 0 & 0 \\ 0 & 1/3 & 2/3 & 0 & 0 \\ 0 & 0 & 1/3 & 2/3 & 0 \\ 0 & 0 & 0 & 1/3 & 2/3 \end{pmatrix},
\begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}.
$$

Figure 4.5: Example grid points sampled via non-uniform grid approach ($m = 5$, $k = 3$)

4.4.2.2 Bounds on the optimal value function In this section we derive some bounding inequalities using the optimal POWLM value function, the grid-based approximate POWLM value function, and the optimal EWLM value function. We choose to condider generic models in this section to show that the results are not restricted to the specific model presented in Section 4.2.

Without loss of generality, we assume that the state space is solely composed of partially observable component (note that this assumption is only for notational clarity.) With some abuse of notation, we use the following throughout the rest of this section:

$$
\begin{aligned}
\mathcal{S} &= \text{core state space,} \\
m &= \text{number of elements in } \mathcal{S}, \\
\Pi(\mathcal{S}) &= \text{the probability (belief) simplex defined over } \mathcal{S}, \\
\mathcal{Z} &= \text{observation space,} \\
\boldsymbol{\pi} &= \text{belief state (i.e., probability distribution over core states),} \\
\mathbf{b}^t &= \text{sampled belief point } t \text{ in the grid,}
\end{aligned}
$$

Table 4.4: Number of belief vectors to sample using the non-uniform grid approach

		q									
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
m	**2**	2	3	4	5	6	7	8	9	10	11
	3	3	5	8	12	17	23	30	38	47	57
	4	4	7	12	20	32	49	72	102	140	187
	5	5	9	16	28	48	80	129	201	303	443
	6	6	11	20	36	64	112	192	321	522	825
	7	7	13	24	44	80	144	256	448	769	1,291
	8	8	15	28	52	96	176	320	576	1,024	1,793
	9	9	17	32	60	112	208	384	704	1,280	2,304
	10	10	19	36	68	128	240	448	834	1,536	2,816
	11	11	21	40	76	144	272	512	960	1,792	3,328
	12	12	23	44	84	160	304	576	1,088	2,048	3,840
	13	13	25	48	92	176	336	640	1,216	2,304	4,352
	14	14	27	52	100	192	368	704	1,344	2,560	4,864
	15	15	29	56	108	208	400	768	1,472	2,816	5,376
	16	16	31	60	116	224	432	832	1,600	3,062	5,888
	17	17	33	64	124	240	464	896	1,728	3,328	6,400
	18	18	35	68	132	256	496	960	1,856	3,584	6,912
	19	19	37	72	140	272	528	1,024	1,984	3,840	7,424
	20	20	39	76	148	288	560	1,088	2,112	4,096	7,936
	21	21	41	80	156	304	592	1,152	2,240	4,352	8,448
	22	22	43	84	164	320	624	1,216	2,368	4,608	8,960
	23	23	45	88	172	336	656	1,280	2,496	4,864	9,472
	24	24	47	92	180	352	688	1,344	2,624	5,120	9,984
	25	25	49	96	188	368	720	1,408	2,752	5,376	10,496
	26	26	51	100	196	384	752	1,472	2,880	5,632	11,008
	27	27	53	104	204	400	784	1,536	3,008	5,888	11,520
	28	28	55	108	212	416	816	1,600	3,136	6,144	12,032
	29	29	57	112	220	432	848	1,664	3,264	6,400	12,544
	30	30	59	116	228	448	880	1,728	3,392	6,656	13,056

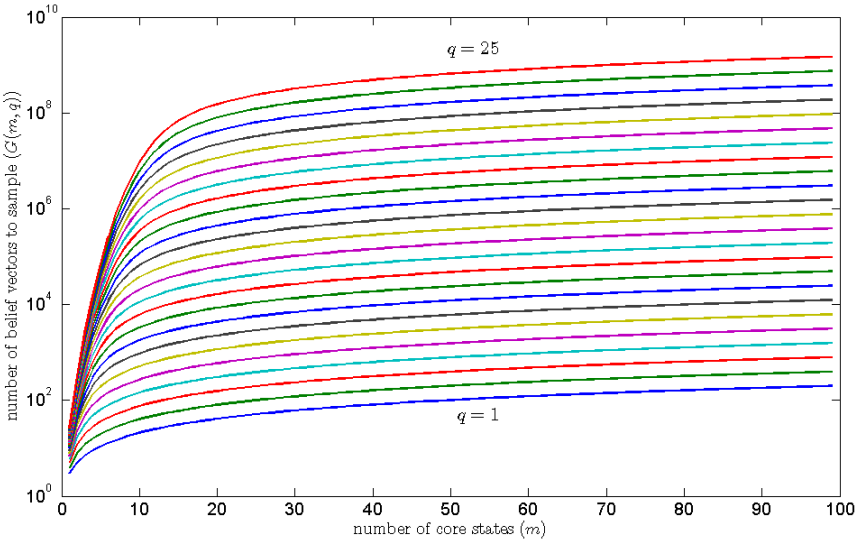


Figure 4.6: Number of belief vectors to sample using the non-uniform grid approach

$\mathcal{G}$ = the grid formed by the sampled belief points $\mathbf{b}^t$,

N = number of elements in $\mathcal{G}$,

a = action taken by the decision maker,

$\mathcal{A}$ = action space,

p_{ij}^a = probability of transiting from state i at time t to state j at time $t+1$ upon taking action a, $i, j \in \mathcal{S}, a \in \mathcal{A}$,

q_{jk}^a = probability of receiving observation k while the state is j at time t and action a was taken, $j \in \mathcal{S}, k \in \mathcal{Z}, a \in \mathcal{A}$,

$r(i,a)$ = the immediate reward associated with taking action a in state i, $i \in \mathcal{S}, a \in \mathcal{A}$,

$u^*(\cdot)$ = the optimal value function of the MDP model,

$v^*(\cdot)$ = the optimal value function of the POMDP model,

$\hat{v}(\cdot)$ = the approximate value function of the POMDP model based on the grid,

λ = discount rate.

For the sake of completeness, we repeat the optimality equations for an MDP model

$$u^*(i) = \max_{a\in\mathcal{A}} \left\{ r(i,a) + \lambda \sum_{j\in\mathcal{S}} p_{ij}^a u^*(j) \right\} \qquad \text{for } i \in \mathcal{S}, \tag{4.33}$$

and for a POMDP model

$$v^*(\boldsymbol{\pi}) = \max_{a\in\mathcal{A}} \left\{ r(\boldsymbol{\pi},a) + \lambda \sum_{i\in\mathcal{S}} \sum_{j\in\mathcal{S}} \sum_{k\in\mathcal{Z}} \pi_i p_{ij}^a q_{jk}^a v^*(\boldsymbol{\pi}') \right\} \qquad \text{for } \boldsymbol{\pi} \in \Pi(\mathcal{S}), \tag{4.34}$$

where $r(\boldsymbol{\pi},a) = \sum_{i\in\mathcal{S}} \pi_i r(i,a)$. Note that $\boldsymbol{\pi}'$ in equation (4.34) is also dependent on the current belief state $\boldsymbol{\pi}$ as well as the observation k and action a, however this dependency is suppressed for notational clarity.

We also recall the Bayesian updating formula for belief states.

$$\pi'_j = \frac{\sum_{i\in\mathcal{S}} \pi_i p_{ij}^a q_{jk}^a}{\sum_{i\in\mathcal{S}} \sum_{j\in\mathcal{S}} \pi_i p_{ij}^a q_{jk}^a}. \tag{4.35}$$

We assume that the grid $\mathcal{G}$ includes the extreme points of the probability simplex $\Pi(\mathcal{S})$. That is, we assume that the unit vectors $\mathbf{e}^i$ are included in the grid (note that $\mathbf{e}^i$ is a probability distribution that accumulates all its mass on i). Therefore, any belief point $\boldsymbol{\pi}$ can be written as a convex combination of the points in $\mathcal{G}$, that is

$$\boldsymbol{\pi} = \sum_{t=1}^{N} \alpha_t \mathbf{b}^t, \tag{4.36}$$

where $\alpha_t \geq 0$ for $t = 1, \ldots, N$ and $\sum_{t=1}^N \alpha_t = 1$.

We compute the approximate value function $\hat{v}(\cdot)$ associated with a belief state $\mathbf{b}^i \in \mathcal{G}$ using a modified version of the POMDP optimality equations (4.34). When we write the optimality equation (4.34) for a given belief state $\mathbf{b}^i \in \mathcal{G}$, the problem will arise on the right-hand-side as we will need the value associated with the updated belief vector, denote $\mathbf{b}'$, which is not known if $\mathbf{b}' \notin \mathcal{G}$. However, we can estimate the value associated with $\mathbf{b}'$ using convex combination of the known values of the grid points $\mathbf{b}^i \in \mathcal{G}$. That is, replacing $v^*(\mathbf{b}')$ on the right-hand-side of equation (4.34) with a convex interpolation estimate $\sum_{t=1}^N \alpha_t \hat{v}(\mathbf{b}^t)$ yields

$$\hat{v}(\mathbf{b}^i) = \max_{a\in\mathcal{A}} \left\{ r(\mathbf{b}^i,a) + \lambda \sum_{i\in\mathcal{S}} \sum_{j\in\mathcal{S}} \sum_{k\in\mathcal{Z}} b_i p_{ij}^a q_{jk}^a \sum_{t=1}^{N} \alpha_t \hat{v}(\mathbf{b}^t) \right\} \qquad \text{for } \mathbf{b}^i \in \mathcal{G}. \tag{4.37}$$

We now derive inequalities that relate $v^*(\cdot)$, $\hat{v}(\cdot)$, and $u^*(\cdot)$. We first need two important results: the first is the fundamental result for POMDPs by Sondik [141] and the second is the well known Jensen's inequality [122].

Theorem 4.3. *$v^*(\cdot)$ is piece-wise linear and convex.*

Proof. See Sondik [141]. □

Theorem 4.4. *Let f be a convex real-valued function and X be a discrete random variable. Then*

$$f\left(\mathbb{E}\left[X\right]\right) \leq \mathbb{E}\left[f\left(X\right)\right].$$

Proof. See Rudin [122]. □

Our first result is given in Theorem 4.5, which shows that the grid-based approximation, $\hat{v}(\cdot)$, provides an upper bound for the true optimal value function, $v^*(\cdot)$, for the POMDP problem.

Theorem 4.5. *$\hat{v}(\mathbf{b}^i) \geq v^*(\mathbf{b}^i)$ for $\mathbf{b}^i \in \mathcal{G}$.*

Proof. We prove the result by induction on the steps of the value iteration algorithm. Let $v_n^*(\mathbf{b})$ and $\hat{v}_n(\mathbf{b})$ denote the true value and the grid-based approximate value, respectively, for belief point $\mathbf{b}$ in the nth iteration of the algorithm. Initialize the algorithm with $v_0^*(\mathbf{b}^i) = \hat{v}_0(\mathbf{b}^i) = 0$ for $\mathbf{b}^i \in \mathcal{G}$, for which the intended result holds trivially. Assume, as the induction hypothesis, that the same inequality holds for iterations $1, 2, \ldots, n$. That is,

$$\hat{v}_n(\mathbf{b}^i) \geq v_n^*(\mathbf{b}^i) \text{ for } \mathbf{b}^i \in \mathcal{G}.$$

For iteration $n+1$,

$$\begin{aligned}
v_{n+1}^*(\mathbf{b}^i) &= \max_{a \in \mathcal{A}} \left\{ r(\mathbf{b}^i, a) + \lambda \sum_{i \in \mathcal{S}} \sum_{j \in \mathcal{S}} \sum_{k \in \mathcal{Z}} b_i p_{ij}^a q_{jk}^a \cdot v_n^*\left(\mathbf{b}'\right) \right\} \\
&= \max_{a \in \mathcal{A}} \left\{ r(\mathbf{b}^i, a) + \lambda \sum_{i \in \mathcal{S}} \sum_{j \in \mathcal{S}} \sum_{k \in \mathcal{Z}} b_i p_{ij}^a q_{jk}^a \cdot v_n^*\left(\sum_{t=1}^{N} \alpha_t \mathbf{b}^t\right) \right\} \qquad (4.38) \\
&\leq \max_{a \in \mathcal{A}} \left\{ r(\mathbf{b}^i, a) + \lambda \sum_{i \in \mathcal{S}} \sum_{j \in \mathcal{S}} \sum_{k \in \mathcal{Z}} b_i p_{ij}^a q_{jk}^a \cdot \sum_{t=1}^{N} \alpha_t \cdot v_n^*\left(\mathbf{b}^t\right) \right\} \qquad (4.39)
\end{aligned}$$

$$\begin{aligned} &\leq \max_{a\in\mathcal{A}}\left\{r(\mathbf{b}^i,a)+\lambda\sum_{i\in\mathcal{S}}\sum_{j\in\mathcal{S}}\sum_{k\in\mathcal{Z}}b_i p_{ij}^a q_{jk}^a\cdot\sum_{t=1}^{N}\alpha_t\cdot\hat{v}_n\left(\mathbf{b}^t\right)\right\} \qquad (4.40)\\ &= \hat{v}_{n+1}(\mathbf{b}^i), \end{aligned}$$

where we have used (4.36) in writing equation (4.38), Theorems 4.3 and 4.4 in writing inequality (4.39), and the induction hypothesis in writing inequality (4.40). □

Our second result concerns a comparison of the optimal value function for the POMDP problem, $v^*(\cdot)$, with that of the MDP problem, $u^*(\cdot)$. A probability distribution that assigns all its mass onto a single core state, say i, is represented by the unit vector $\mathbf{e}^i$ and such a belief state for the POMDP problem is practically equivalent to being in state i for the MDP problem. We are, therefore, interested in the relationship between $v^*(\mathbf{e}^i)$ and $u^*(i)$. We start with some preliminary results.

Lemma 4.8 provides a closed-form formula for the α_t coefficients for finding the updated belief state for a given unit vector $\mathbf{e}^i$.

Lemma 4.8. *The updated belief vector $\boldsymbol{\pi}'$ for the current belief vector $\mathbf{e}^i$ $(i\in\mathcal{S})$ is given by*

$$\boldsymbol{\pi}'=\sum_{t\in\mathcal{S}}\alpha_t\mathbf{e}^t,$$

where

$$\alpha_t=\frac{p_{it}^a r_{tk}^a}{\sum_{j\in\mathcal{S}}p_{ij}^a r_{jk}^a}\quad \textit{for } t\in\mathcal{S}.$$

Proof. Applying formula (4.35) to $\mathbf{e}^i$ yields that the tth component of the updated belief vector $\boldsymbol{\pi}'$ is

$$\pi_t'=\frac{p_{it}^a r_{tk}^a}{\sum_{j\in\mathcal{S}}p_{ij}^a r_{jk}^a}\quad \text{for } t\in\mathcal{S},$$

from which the result follows trivially. □

Lemma 4.9 provides an inequality for the optimal POMDP value function $v^*(\cdot)$, which will be used proving Theorem 4.6.

Lemma 4.9.

$$v^*(\mathbf{e}^i)\leq\max\left\{r(i,a)+\lambda\sum_{j\in\mathcal{S}}p_{ij}^a v^*(\mathbf{e}^j)\right\}\quad \textit{for } i\in\mathcal{S}.$$

Proof. Since belief state $\mathbf{e}^i$ assigns all its mass onto state i, equation (4.34) reduces to

$$v^*(\mathbf{e}^i) = \max_{a \in \mathcal{A}} \left\{ r(i,a) + \lambda \sum_{j \in \mathcal{S}} \sum_{k \in \mathcal{Z}} p_{ij}^a r_{jk}^a \cdot v^*(\boldsymbol{\pi}') \right\} \quad \text{for } i \in \mathcal{S}.$$

Using $\boldsymbol{\pi}' = \sum_{t \in \mathcal{S}} \alpha_t \cdot \mathbf{e}^t$ and applying Jensen's inequality along with Theorem 4.3 on this equation yields

$$\begin{aligned} v^*(\mathbf{e}^i) &= \max_{a \in \mathcal{A}} \left\{ r(i,a) + \lambda \sum_{j \in \mathcal{S}} \sum_{k \in \mathcal{Z}} p_{ij}^a r_{jk}^a \cdot v^* \left(\sum_{t \in \mathcal{S}} \alpha_t \cdot \mathbf{e}^t \right) \right\} \\ &\leq \max_{a \in \mathcal{A}} \left\{ r(i,a) + \lambda \sum_{j \in \mathcal{S}} \sum_{k \in \mathcal{Z}} p_{ij}^a r_{jk}^a \cdot \sum_{t \in \mathcal{S}} \alpha_t \cdot v^* \left(\mathbf{e}^t \right) \right\} \quad \text{for } i \in \mathcal{S}. \end{aligned}$$

Finally, using the result of Lemma 4.8, canceling the term $\sum_{j \in \mathcal{S}} p_{ij}^a r_{jk}^a$ from either side of the fraction, and noting that $\sum_{k \in \mathcal{Z}} r_{tk}^a = 1$ gives

$$\begin{aligned} v^*(\mathbf{e}^i) &\leq \max_{a \in \mathcal{A}} \left\{ r(i,a) + \lambda \sum_{j \in \mathcal{S}} \sum_{k \in \mathcal{Z}} p_{ij}^a r_{jk}^a \cdot \sum_{t \in \mathcal{S}} \frac{p_{it}^a r_{tk}^a}{\sum_{i' \in \mathcal{S}} p_{ii'}^a r_{i'k}^a} \cdot v^* \left(\mathbf{e}^t \right) \right\} \\ &= \max_{a \in \mathcal{A}} \left\{ r(i,a) + \lambda \sum_{k \in \mathcal{Z}} \sum_{t \in \mathcal{S}} p_{it}^a r_{tk}^a \cdot v^* \left(\mathbf{e}^t \right) \right\} \\ &= \max_{a \in \mathcal{A}} \left\{ r(i,a) + \lambda \sum_{t \in \mathcal{S}} p_{it}^a \cdot v^* \left(\mathbf{e}^t \right) \right\} \quad \text{for } i \in \mathcal{S}, \end{aligned}$$

which completes the proof. □

Our second result is given in Theorem 4.6, which shows that the optimal solution for the completely observable MDP problem provides an upper bound for its partially observable counterpart.

Theorem 4.6. *$u^*(i) \geq v^*(\mathbf{e}^i)$ for $i \in \mathcal{S}$.*

Proof. We complete the proof by induction on the steps of the value iteration algorithm. Let $v_n^*(\mathbf{e}^i)$ denote the value of state $\mathbf{e}^i$ for the POMDP problem at the nth iteration of the algorithm. Similarly, let $u_n^*(i)$ denote the value of state i for the perfectly observable MDP problem at the nth iteration of the algorithm. Initialize the algorithm with $v_0^*(\mathbf{e}^i) = u_0^*(i) = 0$ for $i \in \mathcal{S}$, for which the intended result holds trivially. Assume, as the induction hypothesis, that the same inequality holds for iterations $1, 2, \ldots, n$. That is,

$$u_n^*(i) \geq v_n^*(\mathbf{e}^i) \text{ for } i \in \mathcal{S}.$$

For iteration $n+1$, using the result of Lemma 4.9,

$$\begin{aligned} v_{n+1}^*(\mathbf{e}^i) &\leq \max_{a \in \mathcal{A}} \left\{ r(i,a) + \lambda \sum_{j \in \mathcal{S}} p_{ij}^a \cdot v_n^* \left(\mathbf{e}^j\right) \right\} \\ &\leq \max_{a \in \mathcal{A}} \left\{ r(i,a) + \lambda \sum_{j \in \mathcal{S}} p_{ij}^a \cdot u_n^*(j) \right\} \\ &= u_{n+1}^*(i), \end{aligned}$$

which completes the proof. □

Our third result, given in Theorem 4.7, compares the optimal MDP value for a given state i to the grid-based approximate value for the corresponding state $\mathbf{e}^i$ and shows that they are equal.

Theorem 4.7. *$u^*(i) = \hat{v}(\mathbf{e}^i)$ for $i \in \mathcal{S}$.*

Proof. The proof is similar to the proof of Theorem 4.6 and is omitted. □

We need one more inequality before our final result. This inequality is given in Lemma 4.10, which simply states that the convex combination of maximums is at least as much as the maximum of the convex combinations, and will be used in proving Theorem 4.8.

Lemma 4.10. *Given a set of scalars a_i^j for $i = 1, \ldots, I$, $j = 1, \ldots, J$, and a set of non-negative coefficients b_i such that $\sum_{i=1}^I b_i = 1$, the following holds*

$$\max_j \left\{ \sum_i b_i a_i^j \right\} \leq \sum_i b_i \max_j \left\{ a_i^j \right\}.$$

Proof. For $i = 1, \ldots, I$ and $k = 1, \ldots, J$,

$$a_i^k \leq \max_j \left\{ a_i^j \right\}.$$

Multiplying both sides of these inequalities by the nonnegative scalars b_i, we obtain

$$b_i \cdot a_i^k \leq b_i \cdot \max_j \left\{ a_i^j \right\} \quad \text{for } i = 1, \ldots, I; k = 1, \ldots, J.$$

Summing these inequalities for $i = 1, \ldots, I$ yields

$$\sum_i b_i \cdot a_i^k \leq \sum_i b_i \cdot \max_j \left\{ a_i^j \right\} \quad \text{for } k = 1, \ldots, J,$$

which directly implies the intended result. □

Our final result, given in Theorem 4.8, compares the grid-based approximate value for an arbitrary grid point $\mathbf{b}$ to the optimal MDP values at states $i \in \mathcal{S}$ and shows that the convex combination of the latter is at least as much as the former.

Theorem 4.8. *$\hat{v}(\mathbf{b}) \leq \sum_{i \in \mathcal{S}} b_i \cdot u^*(i)$ for $\mathbf{b} \in \mathcal{G}$.*

Proof. Let $\mathbf{b}'$ be the updated belief vector for the given belief $\mathbf{b}$. Using the unit vectors $\mathbf{e}^1, \mathbf{e}^2, \ldots, \mathbf{e}^m$ from the grid to represent the updated belief, we re-write equation (4.37) as

$$\hat{v}(\mathbf{b}) = \max_{a \in \mathcal{A}} \left\{ r(\mathbf{b}, a) + \lambda \sum_{i \in \mathcal{S}} \sum_{j \in \mathcal{S}} \sum_{k \in \mathcal{Z}} b_i p_{ij}^a q_{jk}^a \sum_{t \in \mathcal{S}} \alpha_t \hat{v}(\mathbf{e}^t) \right\}. \tag{4.41}$$

Equations (4.35) and (4.36) together imply that

$$\alpha_t = \frac{\sum_{i \in \mathcal{S}} b_i p_{it}^a q_{tk}^a}{\sum_{i \in \mathcal{S}} \sum_{j \in \mathcal{S}} b_i p_{ij}^a q_{jk}^a} \text{ for } t \in \mathcal{S}. \tag{4.42}$$

Substituting equation (4.42) into (4.41), canceling the term $\sum_{i \in \mathcal{S}} \sum_{j \in \mathcal{S}} b_i p_{ij}^a q_{jk}^a$ in either side of the resulting fraction, and reorganizing the terms yield

$$\begin{aligned} \hat{v}(\mathbf{b}) &= \max_{a \in \mathcal{A}} \left\{ r(\mathbf{b}, a) + \lambda \sum_{i \in \mathcal{S}} b_i \cdot \sum_{t \in \mathcal{S}} p_{it}^a \cdot \hat{v}(\mathbf{e}^t) \cdot \sum_{k \in \mathcal{Z}} q_{tk}^a \right\} \\ &= \max_{a \in \mathcal{A}} \left\{ \sum_{i \in \mathcal{S}} b_i \cdot r(i, a) + \lambda \sum_{i \in \mathcal{S}} b_i \cdot \sum_{t \in \mathcal{S}} p_{it}^a \cdot \hat{v}(\mathbf{e}^t) \right\} \\ &= \max_{a \in \mathcal{A}} \left\{ \sum_{i \in \mathcal{S}} b_i \cdot \left[r(i, a) + \lambda \sum_{t \in \mathcal{S}} p_{it}^a \cdot \hat{v}(\mathbf{e}^t) \right] \right\}, \end{aligned} \tag{4.43}$$

where, in deriving equation (4.43), we have used $r(\mathbf{b}, a) = \sum_{i \in \mathcal{S}} b_i r(i, a)$ and $\sum_{k \in \mathcal{Z}} q_{tk} = 1$ for all $k \in \mathcal{Z}$. Invoking the result of Theorem 4.7, we can re-write the last equation as

$$\begin{aligned} \hat{v}(\mathbf{b}) &= \max_{a \in \mathcal{A}} \left\{ \sum_{i \in \mathcal{S}} b_i \cdot \left[r(i, a) + \lambda \sum_{t \in \mathcal{S}} p_{it}^a \cdot u^*(t) \right] \right\}, \\ &\leq \sum_{i \in \mathcal{S}} b_i \cdot \max_{a \in \mathcal{A}} \left\{ \left[r(i, a) + \lambda \sum_{t \in \mathcal{S}} p_{it}^a \cdot u^*(t) \right] \right\}, \qquad (4.44) \\ &= \sum_{i \in \mathcal{S}} b_i \cdot u^*(i), \end{aligned}$$

where we have applied the result of Lemma 4.10 in writing the inequality (4.44). This completes the proof. □

Corollary 4.5. *$v^*(\mathbf{b}) \leq \hat{v}(\mathbf{b}) \leq \sum_{i \in \mathcal{S}} b_i \cdot u^*(i)$ for $\mathbf{b} \in \mathcal{G}$.*

Proof. The result follows trivially from the results of Theorems 4.5 and 4.8. □

4.4.2.3 Bayesian updating for the grid We have seen in Section 4.4.2.2 that for a given a set of grid points sampled from the appropriate probability simplex, the Bayesian updating of a belief point does not guarantee the updated point to be in the grid. This necessitates forcing the updating to yield points within the grid in order to solve the problem as a regular MDP.

Formally, given a grid $\mathcal{G} = \{\mathbf{b}^1, \mathbf{b}^2, \ldots, \mathbf{b}^N\}$ and an updated belief point $\mathbf{b}'$, we have represented the updated belief as a convex combination of the points in the grid (see equation (4.36)). However, in this representation, we did not specify which grid points to use to represent $\mathbf{b}'$.

To represent any vector in an m-dimensional space, we need at most m linearly independent vectors. Since our grid, by construction, includes all the extreme points of the belief simplex (note that there are m such points, because we have m core states) we can always use these points to represent any updated belief $\mathbf{b}'$ that is not in the grid, as these unit vectors will span the entire probability space. We have found in Section 4.4.2.2 that, if we restrict our basis to the unit vectors for every updated belief, the grid-based approximate value for a given belief $\mathbf{b}$ ($\hat{v}(\mathbf{b})$) is in between the true optimal value of this belief ($v^*(\mathbf{b})$) and

the convex combination of the optimal values of the associated completely observed MDP ($\sum_i b_i u^*(i)$).

However, there may be other vectors in our grid that span the updated belief $\mathbf{b}'$ while, at the same time, yielding better approximate values (better in the sense that is closer to $v^*(\mathbf{b})$).

To find such set of vectors, let x_i be the decision variable representing the α-coefficient of the ith grid point ($\mathbf{b}^i$). Obviously, the following must be satisfied to have a set of $\mathbf{b}^i$ vectors span the updated belief:

$$\begin{aligned} \sum_{i \in \mathcal{G}} \mathbf{b}^i x_i &= \mathbf{b}', \\ \sum_{i \in \mathcal{G}} x_i &= 1, \\ x_i &\geq 0 \quad \text{for } i \in \mathcal{G}. \end{aligned}$$

In light of Corollary 4.5 and the above constraints, we conclude that the tightest bound on $v^*(\mathbf{b})$ through $\hat{v}(\mathbf{b})$ can be achieved by solving the following linear program:

$$\begin{aligned} \text{minimize} \quad & \sum_{i \in \mathcal{G}} \hat{v}(\mathbf{b}^i) \cdot x_i \\ \text{subject to} \quad & \sum_{i \in \mathcal{G}} \mathbf{b}^i x_i = \mathbf{b}', \\ & \sum_{i \in \mathcal{G}} x_i = 1, \\ & x_i \geq 0 \quad \text{for } i \in \mathcal{G}. \end{aligned}$$

A similar linear programming formulation to find the tightest bounds is also presented in [183].

$$\begin{bmatrix}
0 & 0 & 0.000411 & 0.999589 \\
0 & 0 & 0.006340 & 0.993660 \\
0 & 0 & 0.040423 & 0.959577 \\
0 & 0 & 0.192499 & 0.807551 \\
0 & 0 & 0.610468 & 0.389532 \\
0 & 0 & 0.944424 & 0.055576 \\
0 & 0.000074 & 0.997930 & 0.001996 \\
0 & 0.006137 & 0.993863 & 0 \\
0 & 0.049697 & 0.950303 & 0 \\
0 & 0.239447 & 0.760553 & 0 \\
0 & 0.778042 & 0.221958 & 0 \\
0.412087 & 0.587834 & 0.000079 & 0
\end{bmatrix}.$$

Figure 4.7: An example observation matrix for OPO serving Pittsburgh ($k = 12$, $p = 4$)

4.4.3 Numerical results

In this section, we present numerical results parameterized by clinical data. We use the same methodologies discussed in Section 3.4.1 to estimate the rewards ($r_W(h)$ and $r_T(h, \ell)$) and core state transition probabilities ($\mathcal{H}$, $\mathcal{K}$, and $\mathcal{L}$) for POWLM. We also use the same annual discount rate ($\lambda = 0.97$). Finally, we estimate the observation probabilities ($\mathcal{O}$) for POWLM as discussed in Section 4.4.1. An example $\mathcal{O}$ matrix is depicted in Figure 4.7. Note that we obtain patient-specific post-transplant rewards ($r_T(\cdot)$) and OPO-specific $\mathcal{K}$, $\mathcal{L}$, and $\mathcal{O}$ matrices.

We have tested all of our instances on a 64-bit machine with 16GB of RAM, 8 parallel Intel Xeon processors running at 2.33 GHz. The operating system is Ubuntu - a Linux-based operating system. We have implemented the grid-based POMDP solution algorithm (namely, the POMDP-adapted policy iteration algorithm) in C. The two major modifications to the regular policy iteration algorithm are: (i) incorporation of the belief state updating

for POMDPs, and (ii) approximating the value of updated beliefs that are not in the grid as discussed in Section 4.4.2.3. In this implementation, we have used two libraries: (*i*) CPLEX 11.0 callable library to solve the linear programs discussed in Section 4.4.2.3, and (*ii*) Intel Math Kernel Library 10.0 to solve the system of linear equations in the policy evaluation step of the policy iteration algorithm.

Despite the computational power used, our results were affected by the curse of dimensionality. To be able to solve our instances, we have aggregated the MELD scores into 4 health states (namely, $\leq$ 10, 11-18, 19-24, $\geq$ 25), the liver types into 4 qualities as determined by donor's age (the most important factor in determining the quality of a donated liver [117]) (namely, $<$ 30, 30-44, 45-59, $\geq$ 60), and the number ranks to 12 using the same method discussed Section 3.4.1. This setup allowed us to increase the grid resolution parameter q up to 7, which samples 576 points from the 12-dimensional probability simplex (see Table 4.4). It took about 1 hour to solve each problem instance when the grid resolution parameter is set to 7, adding up to a total of about 200 hours (about 8 days and 8 hours) for the 200 instances solved.

4.4.3.1 Estimating the price of privacy We compute our revised price of privacy estimates by comparing the results of the EWLM and POWLM similar to the formula given in Section 3.4.2. For this purpose, let

$$\begin{aligned} u^*(h,\ell,k) &= \text{the optimal value of EWLM for state } (h,\ell,k), \\ \hat{v}(h,\ell,\mathbf{b}) &= \text{the grid-based approximate value of POWLM for state } (h,\ell,\mathbf{b}). \end{aligned}$$

Given an initial belief $\mathbf{b}$ over the possible rank states, we compute the revised estimate of a patient's price of privacy ratio, denoted ρ', by

$$\rho' = \frac{\sum_{k\in\Psi} b_k \cdot u^*(\tilde{h}, L+1, k) - \hat{v}(\tilde{h}, L+1, \mathbf{b})}{\hat{v}(\tilde{h}, L+1, \mathbf{b})}. \tag{4.45}$$

where $\tilde{h}$ is the patient's health at the time of her registration to the waiting list and $L+1$ indicates that no liver is being offered. This estimate ρ' provides a lower bound on the true price of privacy ratio, because of the grid-based approximation (recall the result of Corollary

Table 4.5: Summary statistics for ρ' for 200 patients with various grid resolutions (q)

q	Min	Max	Median	Mean	Standard deviation	% change in mean
1	0.000%	0.000%	0.000%	0.000%	0.000%	-
2	0.000	0.091	0.011	0.017	0.018	-
3	0.000	0.113	0.014	0.022	0.022	25.20%
4	0.000	0.223	0.041	0.055	0.046	150.54%
5	0.000	0.263	0.061	0.073	0.057	33.63%
6	−0.003	0.282	0.065	0.079	0.061	8.52%
7	0.000	0.252	0.061	0.077	0.057	−3.47%
ρ	0.013	15.123	2.883	4.319	4.078	-

4.5). The true price of privacy would be obtained if we replace $\hat{v}(\cdot)$ in equation (4.45) by $v^*(\cdot)$. Therefore, Corollary 4.5 implies that the true price of privacy is always non-negative.

Note that, when computing ρ' in equation (4.45), we use the national liver simulation model of Shechter et al. [132] to estimate initial belief $\mathbf{b}$ for each patient, where $\mathbf{b}$'s are assumed to be a function of health state and OPO.

We solve POWLM using the grid-based approach discussed in Section 4.4.2 for each of the 200 patients . A summary of the descriptive statistics for various values of the grid resolution parameter, $q = 1, \ldots, 7$, are given in Table 4.5. When we increased the grid resolution parameter to 8 the computational time increased to about 9 hours for the couple of instances solved, which would have added up to about 1800 hours (75 days) if we had solved all 200 instances. Considering the tradeoff between the computational time and the percent change in the mean price of privacy estimates (see the last column of Table 4.5), we have not attempted solving all 200 instances for $q = 8$. We have also appended the summary statistics for ρ, the price of privacy estimate obtained by comparing the EWLM and the IWLM solutions, from Table 3.5 to the bottom row of Table 4.5.

The first thing to note in these results is that when $q = 1$, $\sum_{k\in\Psi} b_k \cdot u^*(h, \ell, k) = \hat{v}(h, \ell, \mathbf{b})$ for all 200 instances, implying $\rho' = 0\%$ for all instances. This result is as expected, because

when $q = 1$, the generated grid consists only of the unit vectors $\mathbf{e}^i$ $(i = 1, \ldots, 12)$, and, therefore, $\hat{v}(h, \ell, \mathbf{b}) = \sum_k b_k \hat{v}(h, \ell, \mathbf{e}^k)$ for any $\mathbf{b} \in \Pi(\Psi)$. But Theorem 4.7 dictates that $\hat{v}(h, \ell, \mathbf{e}^k) = u^*(h, \ell, k)$, hence $\hat{v}(h, \ell, \mathbf{b}) = \sum_{k \in \Psi} b_k \cdot u^*(h, \ell, k)$ for any $\mathbf{b} \in \Pi(\Psi)$. This suggests that solving a POMDP problem using a grid-based approach, where the grid is restricted to vertices of the belief simplex, would be as if treating the partially observable state as completely observable and solving the resulting MDP problem.

The second thing to note is that the price of privacy ratios across all patients for all the grid resolutions is negligibly small: the maximum ratio obtained is 0.282% (computed when $q = 6$); the mean ρ' across all grid resolutions is less than 0.1%. Comparing these results to the mean ρ (4.319%) obtained by comparing the EWLM and the IWLM solutions, we conclude that one can do significantly better by implementing the POWLM optimal policy as opposed to the optimal policy of IWLM. In other words, one can nearly eliminate the price of privacy by implementing the partially observed waiting list model.

A third observation from our numerical results is that we have encountered some instances that produced negative price of privacy ratios. The -0.003% in the 'Min' column for $q = 6$ in Table 4.5 is one such example. This happened only in 7 cases out of a total of 1,400 instances (200 patients $\times$ 7 grid resolutions) solved, 5 of which are due to one patient These 7 negative ratios are -0.000387%, -0.003165%, -0.000039%, -0.000374%, -0.000113%, -0.000023%, and -0.000018%. There is no theoretical explanation for such results. We attribute them to numerical precision used in floating-point arithmetic.

Finally, we do not expect, in general, any monotone behavior with respect to the grid resolution parameter q, since a grid based on, say, $q = 7$ samples points very differently than a grid based on, say, $q = 5$. However, we do expect to have monotone behavior in the results associated with two grids, where one of the grids contains all of the points in the other grid (for instance, a grid based on $q = 4$ contains all the points of a grid based on $q = 2$; similarly a grid based on $q = 6$ contains all the points of a grid based on $q = 3$; and all grids contain all the points of a grid based on $q = 1$.) We have encountered very few violations to this expectation due to numerical precision, which are summarized in Table 4.6

We also test whether state aggregations have a major affect on the negligibly small price of privacy estimates summarized in Table 4.5. For this purpose, we have included two levels

Table 4.6: Number of instances that violated monotone behavior in q

		q						
		1	2	3	4	5	6	7
q	1	-	0	1	2	1	2	1
	2	-	-	NA	6	NA	3	NA
	3	-	-	-	NA	NA	2	NA
	4	-	-	-	-	NA	NA	NA
	5	-	-	-	-	-	NA	NA
	6	-	-	-	-	-	-	NA
	7	-	-	-	-	-	-	-

NA: Not Appropriate for comparison

Table 4.7: Affect of state aggregations on ρ'

				Mean ρ' when q			
H	L	K	Mean ρ	1	2	3	4
4	4	10	4.079%	0.000%	0.022%	0.030%	0.063%
4	4	12	4.319	0.000	0.017	0.022	0.055
4	4	30	5.214	0.000	0.006	0.007	0.026
9	4	10	3.893	0.000	0.021	0.029	0.066
9	4	12	4.117	0.000	0.018	0.023	0.059
9	4	30	4.968	0.000	0.006	0.007	0.026

for the number of health states (i.e., $H = 4, 9$), and 3 levels for the number of rank states (i.e., $K = 10, 12, 30$). The results of these experiments are summarized in Table 4.7. The first three columns in this table list down the number of health, liver, and rank states used in the experiments. The fourth column displays the mean price of privacy estimates obtained using formula (3.22) for the 200 hundred patients we have studied earlier. The last four columns display the mean price of privacy estimates obtained using formula (4.45) for the same 200 patients for various grid resolution parameter (q) values. We observe that aggregating core states does not significantly affect the price of privacy estimates. The mean ρ across all state aggregations is on the order of 4.5%, whereas the mean ρ' is still negligibly small.

4.5 CONCLUSIONS

This chapter considers the accept/reject decision problem faced by an end-stage liver disease patient when partial information about the status of the patient is available. The model presented in this chapter extends the one given in Chapter 3 by allowing the rank of the patient to be partially observable as in the current liver transplantation system.

Our contributions with this chapter are manifold. First, we provide a partially observable Markov decision process model with a hybrid state space. To the best of our knowledge, this study is the first to incorporate partial information into patient's decision making in organ transplantation. Second, we derive structural results of this partially observable model, including conditions under which control-limit optimal policies exist. Furthermore, we provide a grid-based solution methodology for numerical computation and provide several bounding inequalities relating the optimal value of the partially observable model, the optimal value of the completely observable counterpart, and the grid-based approximate value of the partially observable model. Third, we numerically solve the partially observable model for 200 patients using clinical data. Finally, we use this partially observable model to refine our estimate of a patient's price of privacy.

We conclude from the analysis conducted in this chapter that the optimal accept/reject decisions of end-stage liver disease patients can change significantly depending on where they

believe they are on the waiting list. As a result of this, a patient can do significantly better in terms of overall life expectancy by taking advantage of the partial information available to her through UNOS's website when compared to ignoring this information. Since we do not know how tight our grid-based approximation is, it is not exactly true to argue that patients can almost completely eliminate the price of privacy by exploiting the available partial waiting list information. However, as we have found that the maximum price of privacy to be less than 0.3% over all grid configurations, it is highly possible that this partial waiting list information can significantly reduce the patients' price of privacy.

There are some limitations of the study we presented. One is that we aggregated states in our computations for numerical tractability, which obviously affected our numerical results. It is hard to assess the direction of this effect, however it is conceivable that our price of privacy estimates might have been larger with coarser state representations. This limitation will be overcome with more computational power. We also assumed time-homogeneity in estimating the observation probabilities for our computational study. A more realistic study should allow these probabilities to be time-dependent. Such an implementation have immense data requirements and computational power and we leave it for future work.

Furthermore, in our conceptual model and computational study, we did not consider any objective other than maximizing the patient's total expected discounted lifetime. Consideration of other objectives such as quality-adjusted life expectancy or incorporation of risk sensitivity in objectives is left for future work. Indeed, we can incorporate quality adjusted life expectancy into our current model by including a quality-adjustment parameter and the current structural results would not be affected by this change. However, we are not aware of any study that reports utility weights associated with different stages of the liver disease. It would be interesting to see how the current results would be affected if such utility weights can be assessed. However, incorporating risk sensitivity of patients and other objectives can change our model drastically. This area of improvement is left for future work.

Another limitation of our current work is that we neglected game issues involved in this decision making process. An appropriate model to incorporate game into this situation would be a partially observable stochastic game. This challenging task is left for future work.

5.0 EMPIRICAL ANALYSIS OF LIVER ACCEPT/REJECT DECISIONS

The price of privacy estimates we have computed in Chapters 3 and 4 are based on the assumption that patients are making their accept/reject decisions optimally so as to maximize their life expectancy. This chapter examines that assumption by analyzing historical liver accept/reject decisions. For this purpose, we develop a non-stationary infinite-horizon Markov decision process (MDP) model and test its performance in an extensive empirical study, in which we compare the decisions suggested by the model to the actual decisions made by real patients.

In this empirical analysis, we restrict our study to patients transplanted with living-donors only. The main reason for this limitation is the unavailability of data for patients transplanted with cadaveric donors that would reveal the decisions made at various rank states. In practice, most patients who end up being transplanted with living-donor livers also join the UNOS waiting list, which is used only for prioritizing patients for cadaveric-donor livers. In our analysis, we use a patient's listing date to the UNOS waiting list and the date she received the living-donor liver transplant to compute her observed pre-transplant life. By doing so, we assume that a patient's accept/reject decision process is initiated when she joins the waiting list. Indeed, we are not aware of any other systematic way to compute the observed pre-transplant life of a patient if she did not join the waiting list.

This chapter is organized as follows. In Section 5.1, we discuss our data sources and patterns we observe in these data. In Section 5.2, we review the stationary MDP model of Alagoz et al. [4] and compare the decisions suggested by this model to the actual decisions of patients. In Section 5.3, we introduce the non-stationary MDP formulation and discuss computational results associated with this improved model. Finally, we provide some concluding remarks in Section 5.4.

5.1 HISTORICAL DATA PATTERNS

We use the publicly available data provided by UNOS for our computational study. This data set includes all the end-stage liver disease patients who have listed in the UNOS waiting list between December 1985 and April 2007 and received a liver transplant between September 1987 and April 2007. There are a total of 84,054 liver recipients in this data set. We eliminate all recipients transplanted with cadaveric-donor livers (80,715) and all non-adult recipients (aged 17 or younger) transplanted with living-donor livers (1,097). These reduce our data set to 2,242 patients. For our analysis, we need a patient's MELD score both at the time she listed in the waiting list and at the time she received the transplant. From the remaining data set, we eliminate all patients who are missing at least one of these MELD scores (1,223), leaving 1,019 patients for further analysis.

The UNOS data is very sparse between listing and transplant. Moreover, most of the patients in the UNOS data are pre-MELD patients, rendering the use of this data set for estimating the natural history of patients. To estimate the health transition probabilities (i.e., the $\mathcal{H}$ matrix) for our model, we use a more detailed data set obtained from the University of Pittsburgh Medical Center (UPMC), which includes only patients who joined the UNOS waiting list through The Thomas E. Starzl Transplantation Institute at UPMC. This data set is not publicly available and is the same data set used by Alagoz [2]. There are a total of 3,009 patients in this data set who joined the system between 1991 and 2000. Eliminating those patients with missing observations reduces this data set to 1,997 patients.

Roberts et al. [117] classify different liver diseases into 10 groups. Alagoz [2] further aggregates these groups into 5 categories: (1) cirrhotic diseases that include primary biliary cirrhosis, primary sclerosing cholangitis, alcoholic liver disease, and autoimmune disorders, (2) hepatitis infections that include hepatitis B and C viruses, (3) acute liver diseases, (4) cancers, and (5) other liver diseases. We use the natural history model of Alagoz et al. [3], which incorporates cubic splines to fill in missing observations in the data, to estimate disease-specific $\mathcal{H}$ matrices. We omit patients in disease groups 4 and 5 due to limited sample size in the UPMC data set. This final filter leaves 721 patients in the UNOS data set for our analysis (387 in disease group 1, 307 in disease group 2, and 27 in disease group 3).

Figures 5.1 and 5.2 display the distribution of MELD scores at the time of listing and at the time of transplant, respectively, for this final cohort of 721 patients. For the entire cohort, the average MELD score at the time of listing and at the time of transplant is 13.59 and 14.21, respectively. These averages are about the same for each of the three disease groups.

Figure 5.3 displays scatter plots of the MELD scores at listing against the MELD scores at transplant for each of the three disease groups and the entire cohort. It shows that 67.18% of the patients in disease group 1 are transplanted with a living-donor liver at the same MELD score they joined the system or better (i.e., smaller). This fraction drops to 59.28% for patients in disease group 2, and to 55.56% for patients in disease group 3. Overall, 63.38% of all patients received their living-donor liver transplant at a MELD score that is no worse than their listing MELD score. In other words, more than half of all patients' health, as measured by the MELD scoring system, improves (or at least stays same) from the time they are listed in the system until they receive a living-donor liver transplant. This observation is more frequent in disease group 1 compared to disease groups 2 and 3. One of the main reasons for such a behavior is the fact that many patients are listed during an exacerbation of their disease, when they are acutely ill. Therapy in the hospital often makes such patients transiently better.

Finally, Figure 5.4 displays scatter plots of the MELD scores at listing against the number of days until transplant. It shows that, generally, the higher a patients listing MELD score the less she waits before she receives the living-donor liver transplant. For instance, all but 3 patients with a MELD score of at least 20 received their transplants within one year after being listed in the system (see the bottom-right plot in Figure 5.4). Furthermore, 83.22% of all patients received their transplant within one year after listing and 95.43% received within two years after listing.

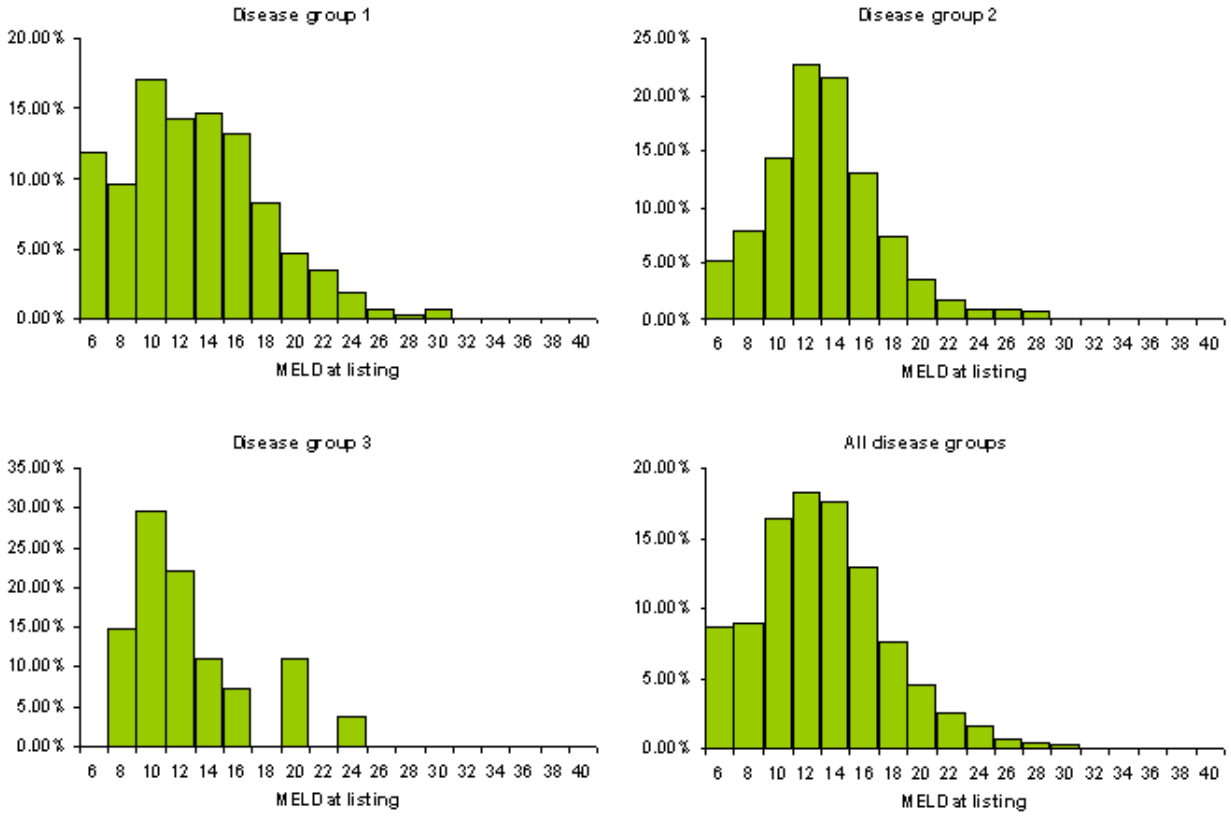


Figure 5.1: Distribution of MELD scores at the time of listing

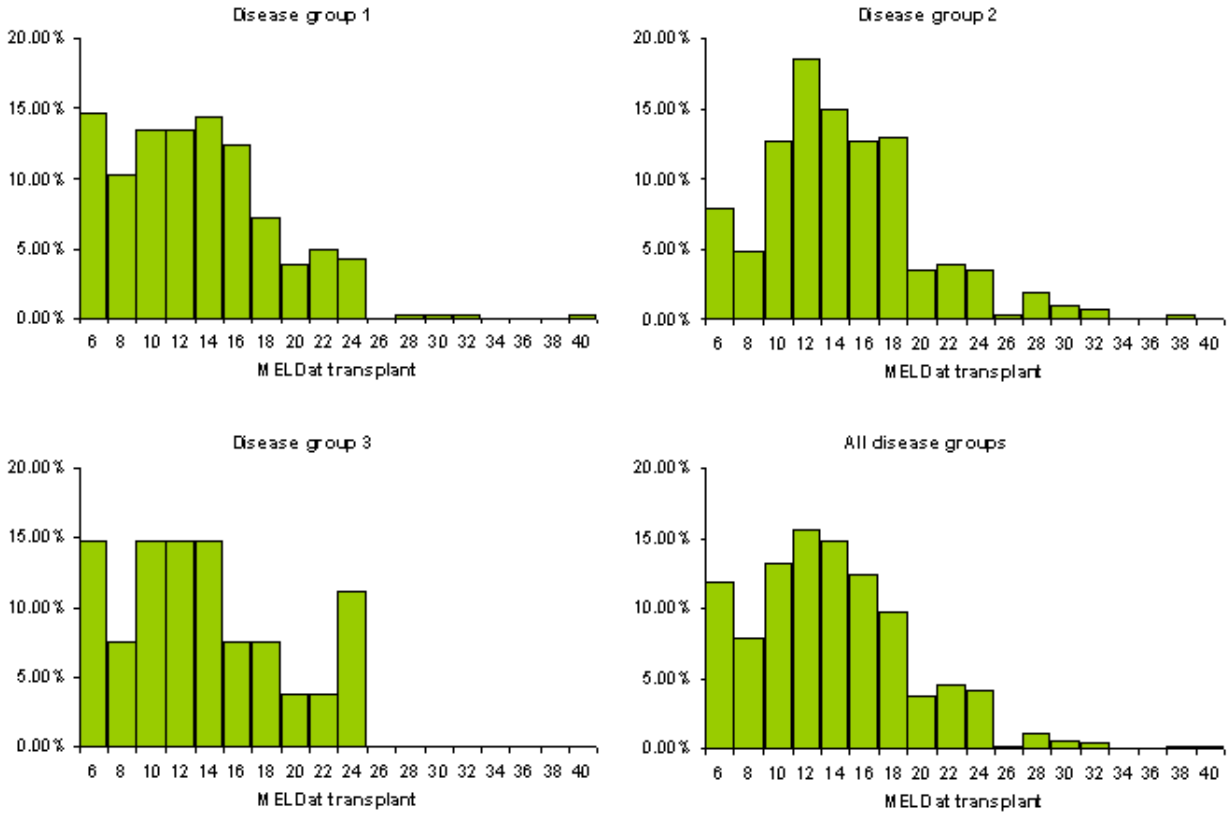


Figure 5.2: Distribution of MELD scores at the time of transplant

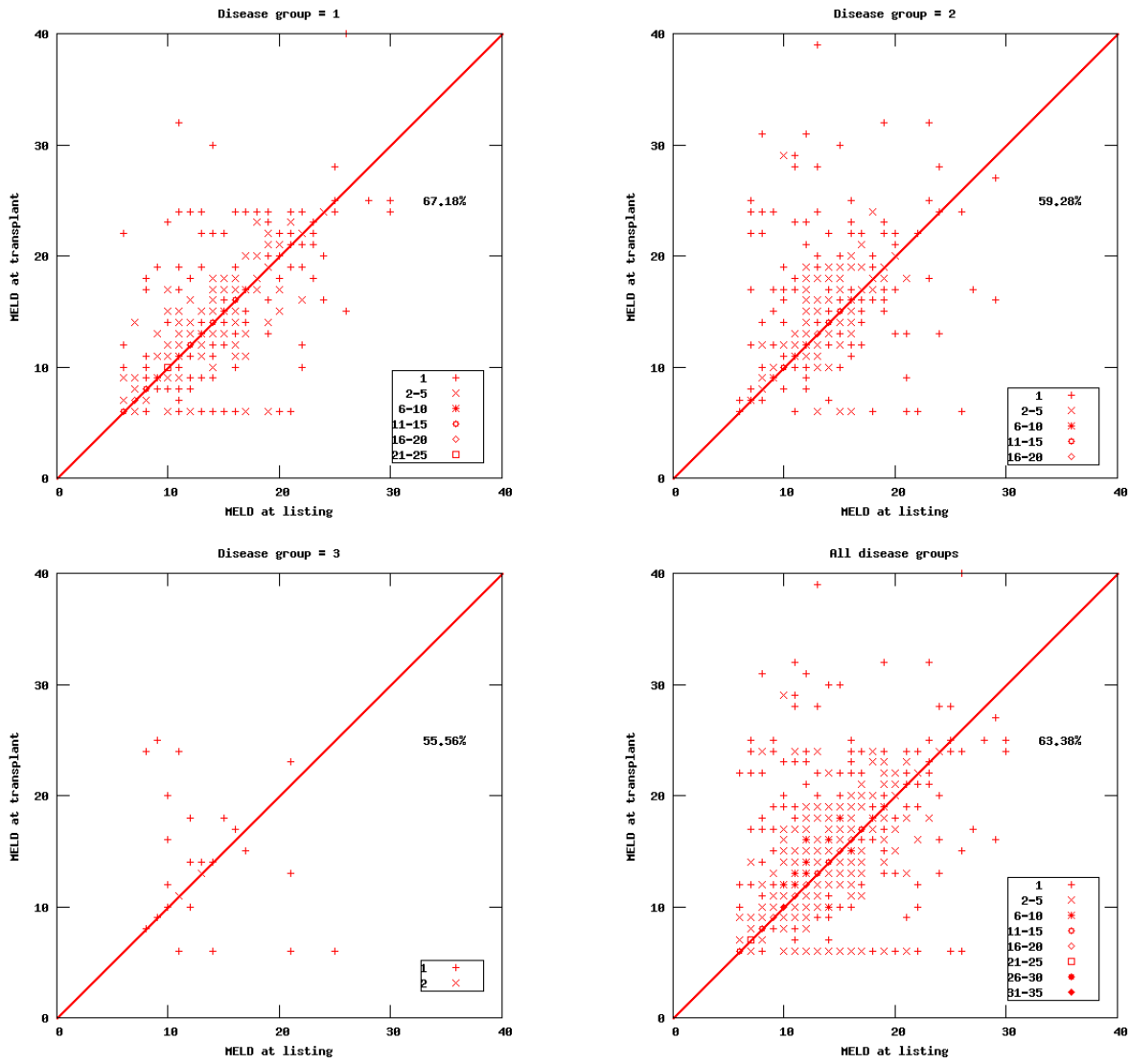


Figure 5.3: Comparison of MELD scores at listing and at transplant

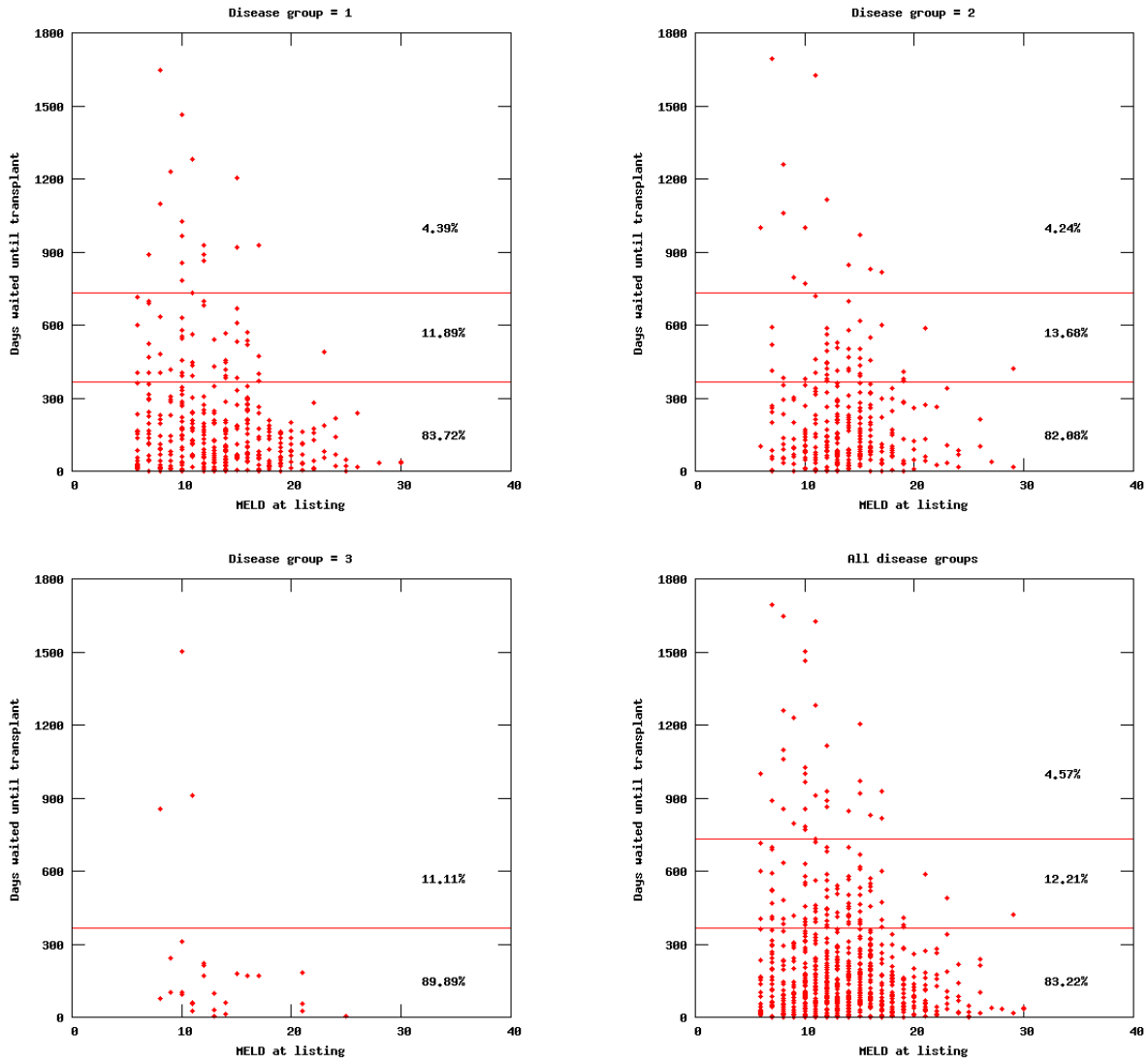


Figure 5.4: Comparison of MELD scores at listing and days until transplant

5.2 STATIONARY MARKOV DECISION PROCESS MODEL

In this section, we review the stationary Markov decision process model of Alagoz et al. [4] for the optimal timing of living-donor liver transplantation problem.

The *optimal timing of living-donor liver transplantation problem* assumes that an end-stage liver disease patient has access to a living-donor, but is ineligible or has decided not to receive cadaveric liver offers. It further assumes that the living-donor is available for transplantation at all times and that the donated liver is of fixed quality $\ell \in \Phi$. The patient occupies a health state $h \in \Omega$. The decision-maker must decide *when* to accept transplantation from the available living-donor so as to maximize her total expected discounted reward. If the patient chooses to wait for one more time period, she accrues an intermediate reward which is a function of her current health status, denoted $r_W(h)$, and faces the same problem at the next time period, provided she lives. If, on the other hand, the patient chooses to transplant in the current decision epoch, then she receives a terminal reward which is a function of her current health status and the quality of the donated liver, denoted $r_T(h, \ell)$, and terminates the process. A potential caveat to this model is the fact that patients simultaneously wait for a cadaveric-donor liver while they are optimizing the timing of their living-donor liver transplantation. Alagoz et al. [5] present a model for such patients, however the data calibration requirements for this model is quite high. Therefore, we restrict ourselves to the simpler model, which we present next.

Alagoz et al. [4] provide the following stationary model for this problem:

$$v(h) = \max \left\{ r_T(h, l); \quad r_W(h) + \lambda \sum_{h' \in \Omega} \mathcal{H}\{h'|h\} \cdot v(h') \right\} \quad \text{for } h \in \Omega, \tag{5.1}$$

where $\lambda \in [0, 1]$ is the discount rate. They analyze the structural properties of this model and solve some example problems. However, they do not present any results comparing the decisions suggested by this model to the actual decisions of patients.

We solved this stationary model for all of the 721 patients in our data set. We compared the resulting optimal policy, for each patient, to her actual decisions and computed the difference in patient's life expectancy associated with these decisions. Figure 5.5 summarizes the results.

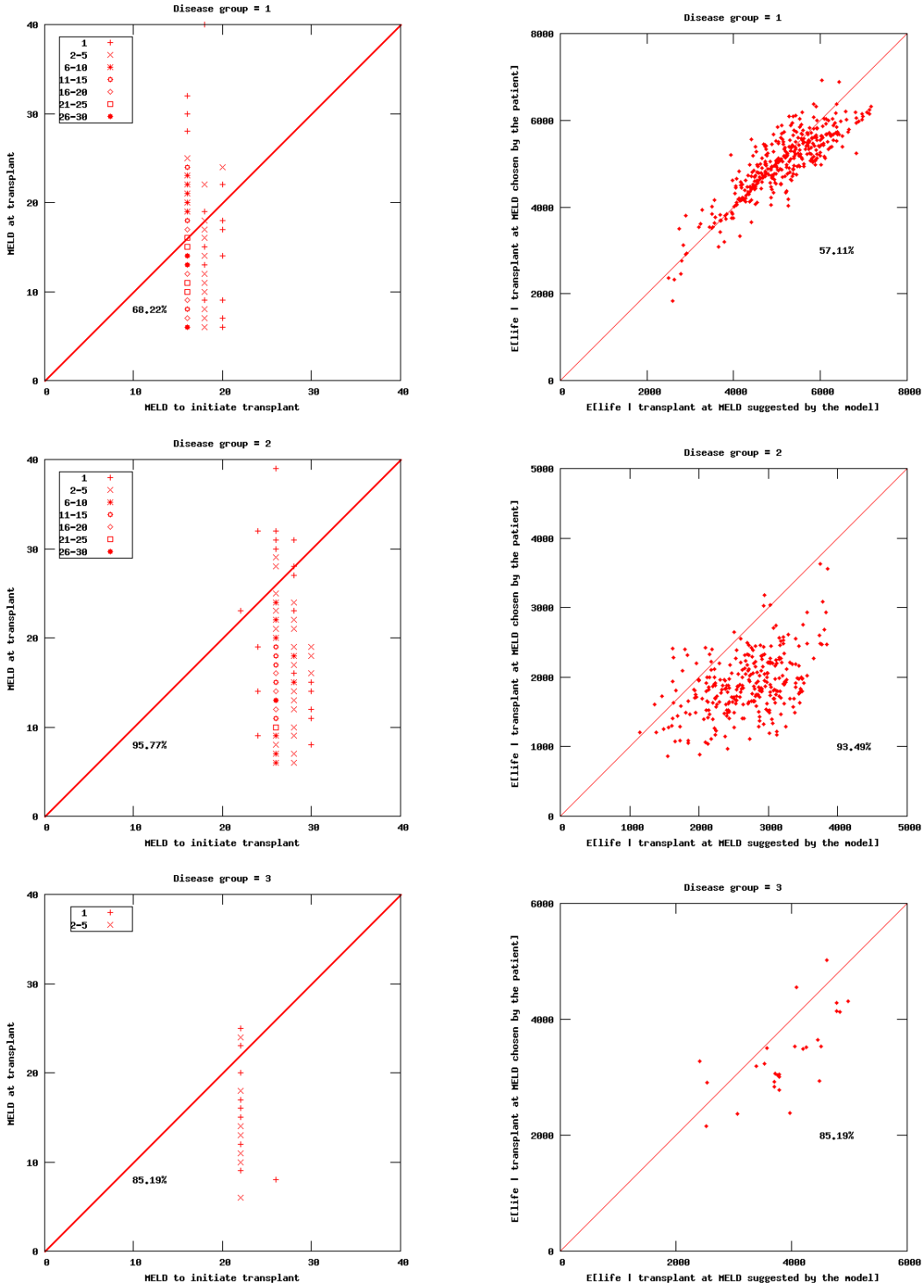


Figure 5.5: Actual decisions compared to decisions suggested by the stationary model

We first investigate if the patient's MELD score choice to receive the graft follows the decision rule suggested by the model (see the plots on the left column of Figure 5.5). The optimal policy suggested by the model is of control-limit type for all of the problems we have solved. That is, it is optimal to wait for all health states up to a threshold health and to transplant for all health states on or above the threshold. Let $MELD^*$ denote this threshold health state for the optimal policy suggested by the model. Also let $MELD_T$ denote the health state in which patient actually received the living-donor liver transplant. Our results show that 264 patients (68.22%) from the disease group 1 cohort actually transplanted at a MELD score lower than what is suggested by the model (see top-left plot in Figure 5.5). These numbers for disease groups 2 and 3 are 294 (95.77%) and 23 (85.19%), respectively. In other words, only a small fraction (less than 5%) of disease group 2 patients follow the decision rule suggested by the model. A slightly larger proportion (about 15%) of disease group 3 patients make decisions in line with what is suggested by the model. A larger (yet still disappointing) proportion (more than 30%) of disease group 1 patients make optimal decisions with respect to maximizing total life expectancy.

Next, we investigate the affect of these apparently suboptimal decisions on patients' expected lives (see the plots on the right column of Figure 5.5). For this purpose, let EL_T denote the expected remaining life of a patient given she has been transplanted in health state $MELD_T$. We compute EL_T by summing the observed pre-transplant life (i.e., actual number of days patient waited from the date she listed in the system until she received her transplant) and the predicted post-transplant life (i.e., $r_T(MELD_T, \ell)$, where ℓ is the fixed liver quality). Similarly, let EL^* denote the expected remaining life of the patient given she follows the optimal policy suggested by the model. We simply take EL^* to be equal to the value of starting in health state $MELD_L$, the MELD score of the patient at the time of listing, and acting optimally as suggested by the model. Our results show that 93.49% (85.19%) of the disease group 2 (3, respectively) cohort experience shorter expected lives as a result of suboptimal decision-making. Furthermore, although 68.22% of the disease group 1 cohort acted suboptimally, only 57.11% (221 patients) of the cohort would have lower expected survival (see top-right plot in Figure 5.5). This perplexing result may be explained by the fact that some patients join the waiting list with relatively higher MELD scores when

Table 5.1: Actual decisions compared to decisions suggested by the stationary model ($MELD_L \leq 12$)

Disease group	Total cases	$MELD_T < MELD^*$		$EL_T < EL^*$	
		Count	Percentage	Count	Percentage
1	181	165	91.16%	160	88.40%
2	115	109	94.78%	113	98.26%
3	15	12	80.00%	14	93.33%

they are really sick, their health improves over time with the help of medications during which they accumulate some waiting time, and they transplant at a healthier MELD score than they were listed. As a result, they accrue much more expected post-transplant life than the model predicts.

We do a simple test of this hypothesis by eliminating all patients with a listing MELD score of more than 12 (i.e., eliminate if $MELD_L > 12$). The new results are summarized in Table 5.1. The results associated with disease groups 2 and 3 have not changed drastically, although they indicate a higher fraction of suboptimal behavior among the restricted cohorts compared, respectively, to the results of the entire disease group 2 and 3 patients. Most importantly, we find that a significantly larger proportion (91.16%) of disease group 1 patients act suboptimally (compare with 68.22%) and, as a result, a significantly larger proportion (88.40%) of this cohort have shorter expected lives (compare with 57.11%). By restricting the cohorts to patients with a listing MELD score of at most 12, the average difference in expected lives increased from 115 days to 412 days for disease group 1 cohort, from 793 days to 1206 days for disease group 2 cohort, and from 536 days to 784 days for disease group 3 cohort. These numbers also show that patients who are performing better (with respect to life expectancy) than the stationary model are those who join the waiting list really sick. These observations motivate us to develop a non-stationary model to represent different natural histories surrounding diagnosis, listing, and chronic disease progression.

5.3 NON-STATIONARY MARKOV DECISION PROCESS MODEL

The non-stationary model introduced in this section is motivated by the results associated with disease group 1 cohort and the discussion surrounding these results in Section 5.2. Some patients join the waiting list when they are really sick and their natural history does not follow a typical chronic progression. Rather, for a certain period, these patients usually improve in health with the help of immediate interventions such as increased use of certain medications. Therefore, the disease progression for these early stages of listing should be estimated separately from the later stages, which requires a non-stationary model. This observation naturally raises the question about the length of this early stage and the number of such stages.

Inspired by the model of Alagoz et al. [4], we provide a more detailed non-stationary MDP model as follows. We divide the entire decision horizon into $M+1$ decision intervals possibly of different lengths. The first M intervals represent the different natural history surrounding diagnosis and listing and the last interval represents the chronic disease progression. Let $\mathcal{H}_m$ denote the health transition probability matrix in interval $m = 1, \ldots, M$, and let $\mathcal{H}$ denote the health transition probability matrix for the last interval $m = M + 1$. Also let $u_n^m(h)$ be the optimal value of state $h \in \Omega$ at stage n for interval $m = 1, \ldots, M$. Similarly, let $v(h)$ be the optimal value of state $h \in \Omega$ for the last interval. Finally, let N_m denote the number of decision epochs in interval $m = 1, \ldots, M$ (i.e., N_m represents the length of each interval).

Given discount rate $\lambda \in [0, 1]$, the optimal solution to the non-stationary model is found by solving the following set of recursive equations:

$$v(h) = \max \left\{ r_T(h, l); \quad r_W(h) + \lambda \sum_{h' \in \Omega} \mathcal{H}\{h'|h\} \cdot v(h') \right\} \quad \text{for } h \in \Omega, \tag{5.2}$$

and, for $m = 1, \ldots, M$ and $n = 1, \ldots, N_m$,

$$u_{N_m}^m(h) = \begin{cases} v(h) & \text{if } m = M, \\ u_0^{m+1}(h) & \text{if } m < M, \end{cases} \quad \text{for } h \in \Omega, \tag{5.3}$$

$$u_{n-1}^m(h) = \max \left\{ r_T(h, l); \quad r_W(h) + \lambda \sum_{h' \in \Omega} \mathcal{H}_m\{h'|h\} \cdot u_n^m(h') \right\} \quad \text{for } h \in \Omega. \tag{5.4}$$

Note that, if $M = 0$ this model reduces to the stationary model of Alagoz et al. [4] presented in Section 5.2. We solve equations (5.2)-(5.4) using backward induction on m: we start with the last interval and proceed our way back to the first interval. Equation (5.2) solves the optimal stopping problem for the last interval. Once the optimal values for this interval are obtained, we take these values to be the terminal rewards of the immediately preceding interval as given in Equation (5.3), change the current interval pointer to the immediately preceding interval, and use these terminal rewards in optimizing the value function for the current interval using Equation (5.4). We iterate in this manner, each time setting the most recent optimal values for the current interval to the terminal rewards of the immediately preceding interval, until we are done with all the intervals.

In summary, we solve M finite horizon MDP models, possibly of different lengths, for the first M intervals and one optimal stopping problem, namely, the living-donor model of Alagoz et al. [4], for the last interval.

For each of the 721 patients in our data set, we solved the non-stationary MDP model several times, each time varying the parameters M and N_m. The choice of these parameters are heavily influenced by the resulting estimates for $\mathcal{H}_m$ and $\mathcal{H}$ matrices. Increasing the value of N_m for any particular interval $m = 1, \ldots, M$ removes off data from all its successive intervals, which may leave insufficient data points for these latter intervals to have meaningful health matrix estimates. The choice of N_M is especially crucial since the number of splines, from which we estimate the health matrices, available at each time point is nonincreasing in time and, therefore, minimum number of splines are available for the last interval $M + 1$. We would like to maximize the number of splines available for the last interval to have a meaningful $\mathcal{H}$ estimate while allowing as much intervals as needed to represent the different stages of the liver disease. While we were able to allow $N_M = 30$ days for disease groups 1 and 2, we could not allow $N_M > 5$ days for disease group 3 as it would produce an $\mathcal{H}$ matrix that would not allow any transitions to non-dead states from most of the health states. Considering all these facts, we have tried several values for M and N_m in consultation with clinicians.

Table 5.2 summarizes the results of solving the non-stationary model for various M and N_m values. Note that $M = 0$ corresponds to the results of the stationary model given in

Table 5.2: Actual decisions compared to decisions suggested by the non-stationary model

Disease group	M	N_m	Total cases	$MELD_T < MELD^*$		$EL_T < EL^*$	
				Count	Percentage	Count	Percentage
1	0	-	387	264	68.22%	221	57.11%
	1	10	387	333	86.05	311	80.36
	1	15	387	333	86.05	312	80.62
	1	20	387	333	86.05	314	81.14
	1	30	387	335	86.56	314	81.14
	2	5, 30	387	324	83.72	225	58.14
	2	10, 30	387	335	86.56	302	78.04
	2	15, 30	387	334	86.30	307	79.33
	2	20, 30	387	335	86.56	313	80.88
	2	25, 30	387	335	86.56	314	81.14
	3	5, 15, 30	387	334	86.30	271	70.03
	3	10, 20, 30	387	337	87.08	307	79.33
2	0	-	307	294	95.77%	287	93.49%
	1	10	307	301	98.05	300	97.72
	1	15	307	302	98.37	305	99.35
	1	20	307	303	98.70	305	99.35
	1	30	307	305	99.35	306	99.67
	2	5, 30	307	304	99.02	305	99.35
	2	10, 30	307	305	99.35	305	99.35
	2	15, 30	307	305	99.35	306	99.67
	2	20, 30	307	305	99.35	306	99.67
	2	25, 30	307	305	99.35	306	99.67
	3	5, 15, 30	307	305	99.35	306	99.67
	3	10, 20, 30	307	305	99.35	305	99.35
3	0	-	27	23	85.19%	23	85.19%
	1	5	27	27	100.00	27	100.00
	2	2, 5	27	27	100.00	27	100.00
	2	3, 5	27	27	100.00	27	100.00

M : number of finite horizon MDPs solved
N_m : end point(s) of the first M intervals
$MELD_T$: MELD score in which patient actually received the transplant
$MELD^*$: MELD score in which model suggests to initiate transplantation
EL_T : expected life given patient transplants at $MELD_T$
EL^* : expected life given patient follows the policy suggested by the model

Section 5.2. The optimal policy suggested by the non-stationary model is of control-limit type for all the M and N_m values we have used. We first compare $MELD_T$, the MELD score in which patient actually received her transplant, to $MELD^*$, the MELD score to initiate transplantation as suggested by the model. We observe that the percentage of disease group 1 patients making suboptimal decisions has increased from 68.22% to about 86% after incorporating non-stationarity into our model. However, changing the number of intervals (M) and the number of decision epochs in each interval (N_m) did not produce significant differences in this percentage. As a result of these suboptimal decisions, compared to the stationary model, the percentage of disease group 1 patients having shorter life expectancies has increased from 57.11% to about 80%. Again different choices of M and N_m did not produce significantly different results with the exception of the configuration $M = 2$, $N_1 = 5$, $N_2 = 30$. Figures 5.6-5.9 present scatter plots of $MELD^*$ versus $MELD_T$ (plots on the left columns) for various M and N_m choices as well as scatter plots of EL^* versus EL_T (plots on the right columns). These results suggest that a significant proportion of patients are transplanting at a MELD score lower than what one should transplant at to maximize life expectancy.

Incorporating non-stationarity causes dramatic changes in the results for disease groups 2 and 3 (see Table 5.2). According to these results, only at most a handful of patients are making their transplant decisions in accordance with the non-stationary model. As a result, almost all of the patients in these disease groups are having shorter life expectancies when compared using the non-stationary model.

Figures 5.10-5.13 display more details about these results for disease group 2 and Figure 5.14 displays similar details for disease group 3. Looking at Figures 5.10-5.13, we notice an odd vertical clustering of points in the scatter plots of expected lives (see plots on the right columns), which becomes more prominent as we increase M and N_M. We further notice that there is a horizontal shift of points to the right in the scatter plots of MELD scores as we increase M and N_m (see plots on the left columns) , implying that the model suggests to wait until patient's MELD score reaches really high values (e.g., above 30). Similarly, looking at Figure 5.14 , we notice unbelievably long life expectancies (e.g., 15,000 days $>$ 40 years) for disease group 3 patients (see plots on the right column).

When we investigate the reason for such behaviors, we find that they are caused by poor transition probability estimates, which dramatically overestimates patient's survival. That is, lack of sufficient data for disease groups 2 and 3 leaves only a few splines (hence, data points) for latter intervals of the non-stationary model. These limited number of splines naturally does not yield clinically plausible transition probability estimates among different health states. In most cases, we do not observe any transition among several health states, which exacerbates the problems when there are no transitions allowed to the death state. This is precisely what is happening in disease groups 2 and 3 as we increase M and N_m. When $M \geq 1$, the $\mathcal{H}$ matrix, the transition probability matrix associated with the last interval of the non-stationary model, significantly overestimates patient survival by not allowing patient death in most health states. As a result, while we solve the optimal stopping problem for the last decision interval, the model makes the patient wait to the longest extend possible and most of the total expected reward is accumulated in this interval, which dominates the results.

Finally, we repeat the same analysis we did at the end of Section 5.2. Since many patients join the system with a high MELD score and improve over time, we restricted our attention to those patients joining the system relatively healthy (i.e., with a listing MELD score of at most 12). Intuitively, such a restriction should yield increased percentages in Table 5.2. Table 5.3 presents these results for patients with a listing MELD score of at most 12. For disease group 1, the percentage of patients acting suboptimally increased from 86% to about 97% and the percentage of patients having shorter expected lives increased from 80% to about 99%. These results show the need of better non-stationary parameter estimates to optimize the decisions of patients who join the system relatively sick and gradually improve over time for a certain period.

5.4 CONCLUSIONS

This chapter is concerned with the issue of assessing whether patients' actual decisions are in line with the decisions suggested by existing stationary and enhanced non-stationary

MDP models. A resolution to this issue would form an important step in identifying true price of privacy, as all of our price of privacy estimates in Chapters 3 and 4 are based on the assumption that patients are making their accept/reject decisions optimally so as to maximize total life expectancy.

We start off with the stationary MDP model of Alagoz et al. [4] and develop a non-stationary MDP model that uses this existing model. We test the performance of these models in an extensive empirical study. In this empirical analysis, we only considered patients transplanted with a living-donor liver. Our numerical results showed significant evidence that patients actual transplant decisions do not follow their optimal policies to maximize life expectancy. More than two-thirds of all patients actually find it attractive to proceed with transplantation at a MELD score lower than their optimal threshold. As a result, patients realize shorter expected total lives. This disparity may be eliminated by considering other objectives such as maximizing quality-adjusted life expectancy or incorporating risk preferences of patient's into the models. It is actually an interesting research question to investigate which objective(s) patients are optimizing when they make their decisions. This and other related issues are left for future research.

The non-stationary model produced clinically plausible results for disease group 1 patients but not so plausible results for patients in disease groups 2 and 3. In other words, while the idea of dividing the entire decision horizon into a few intervals to estimate different natural histories worked well for disease group 1 patients, it backfired for disease groups 2 and 3. This backfire is mainly caused by insufficient data to estimate meaningful transition probabilities among different health states. This limitation can be overcome by accessing more comprehensive data sets.

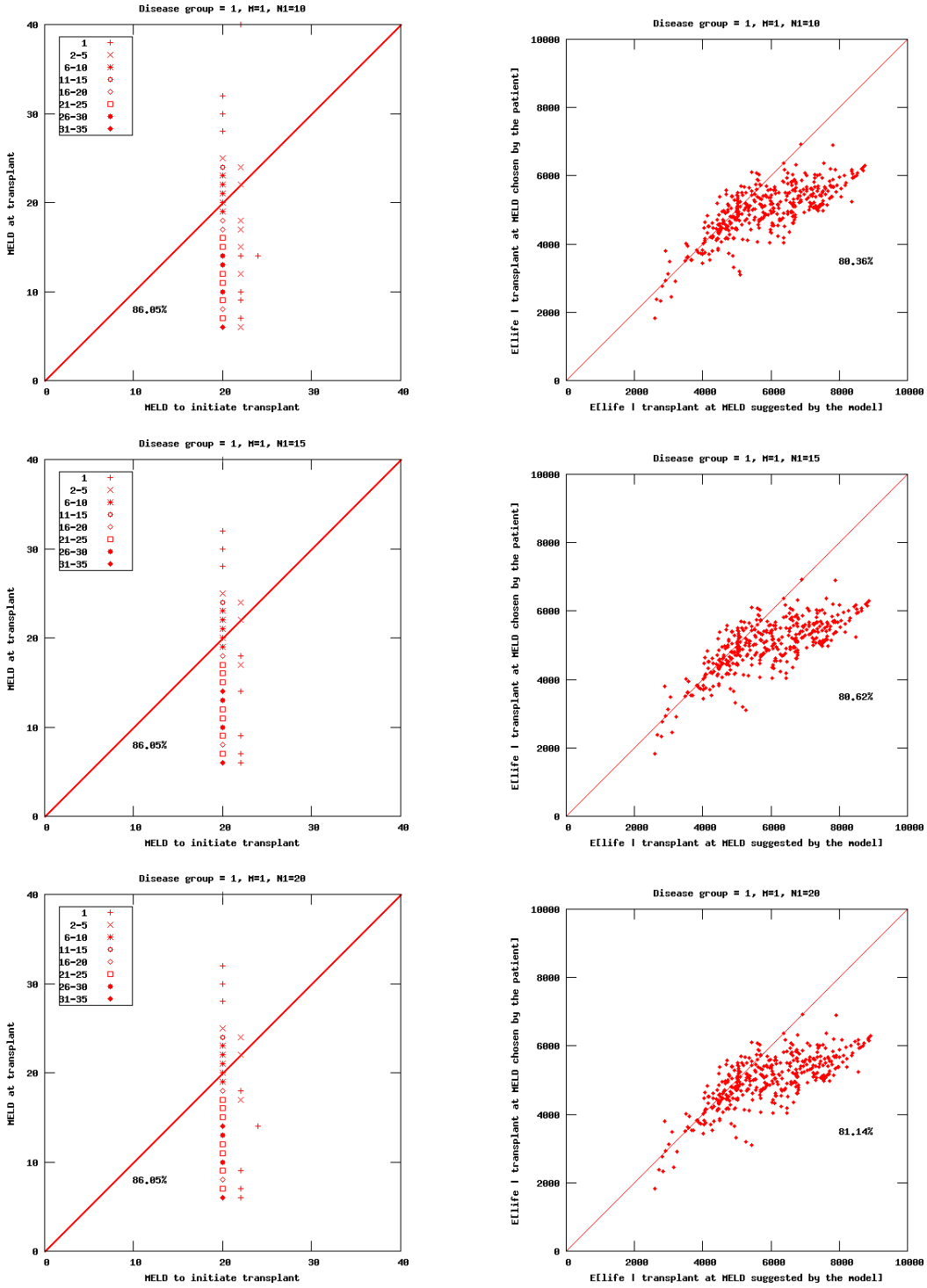


Figure 5.6: Actual decisions compared to decisions suggested by the non-stationary model

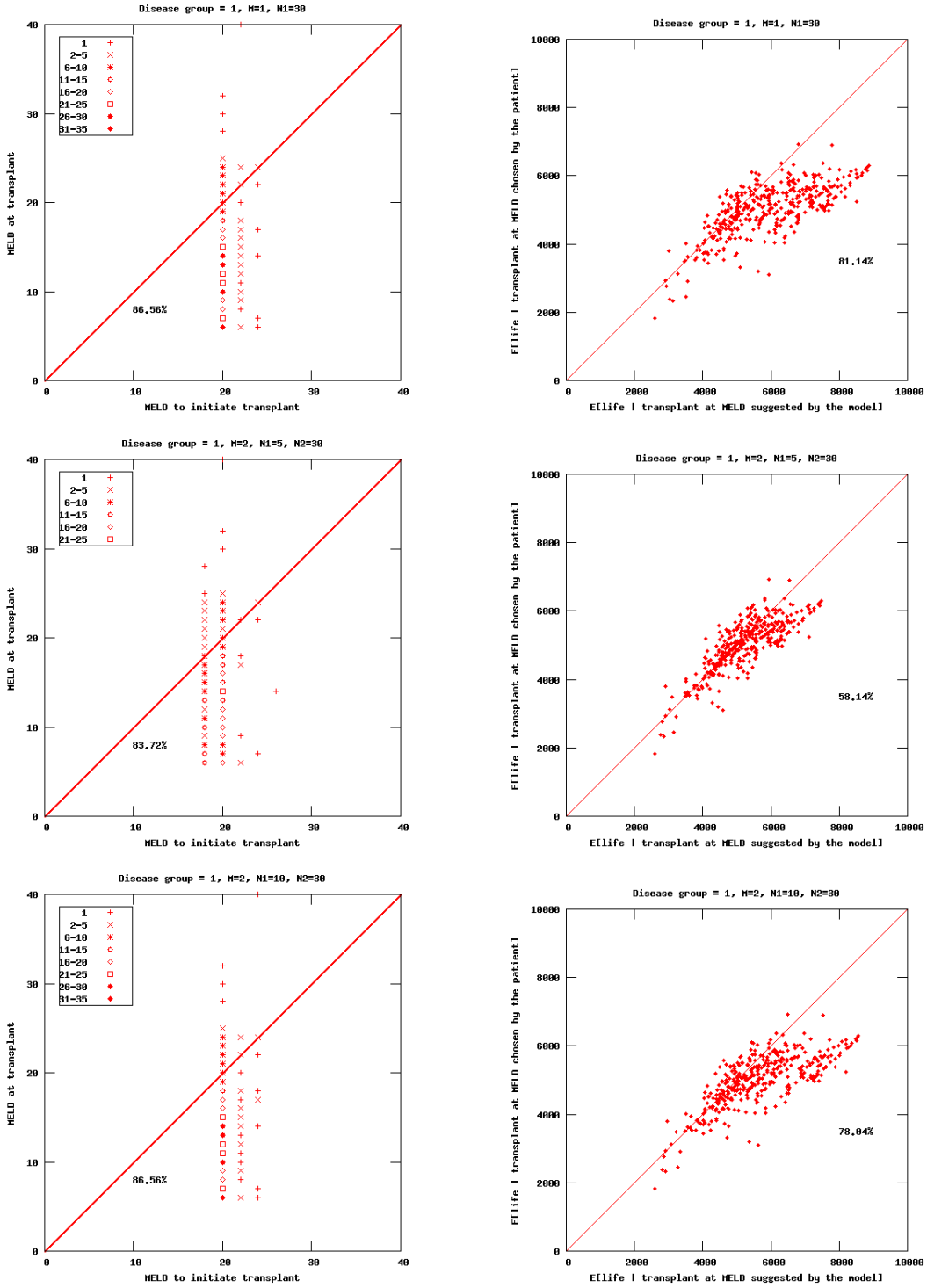


Figure 5.7: Actual decisions compared to decisions suggested by the non-stationary model

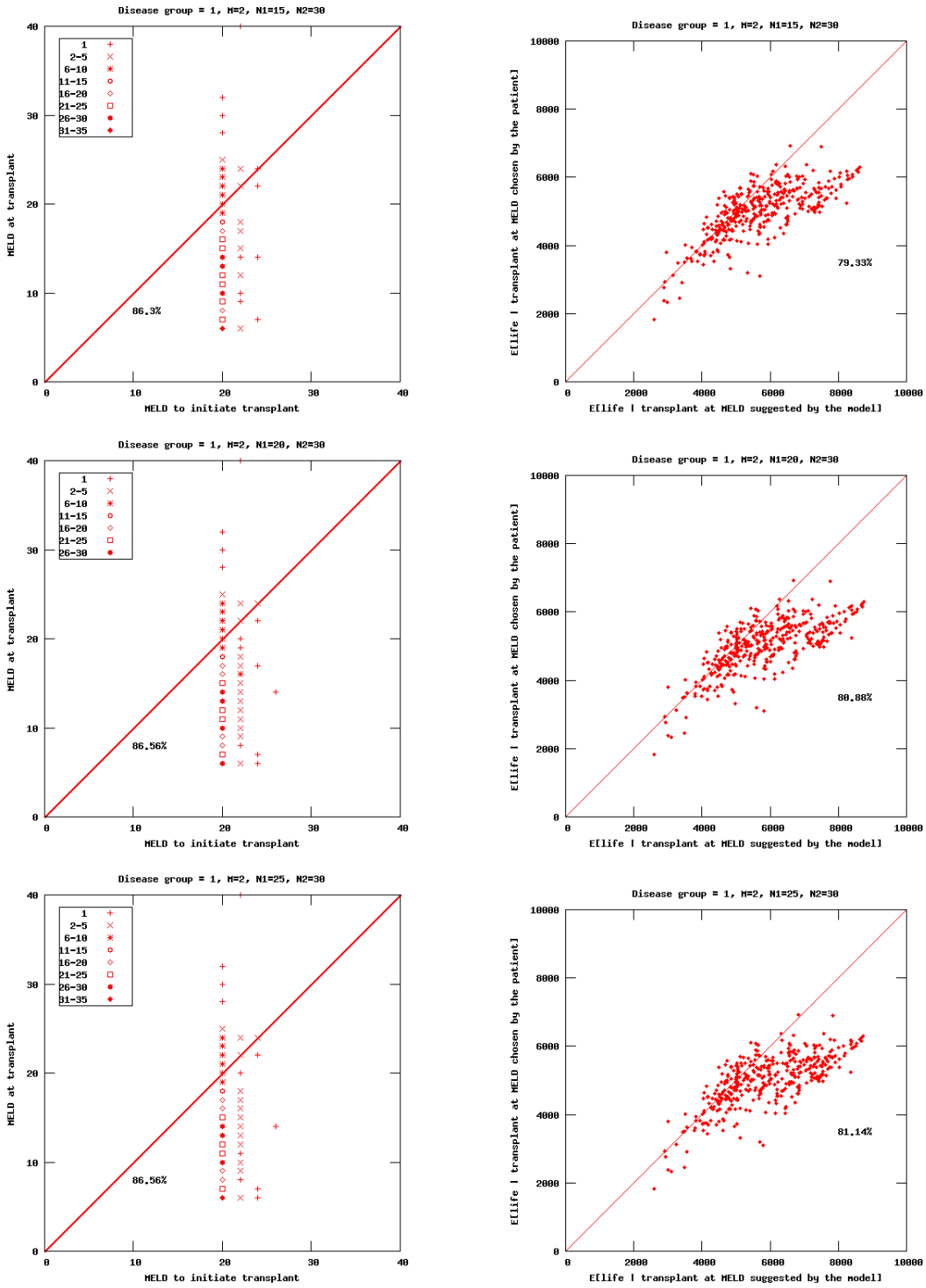


Figure 5.8: Actual decisions compared to decisions suggested by the non-stationary model

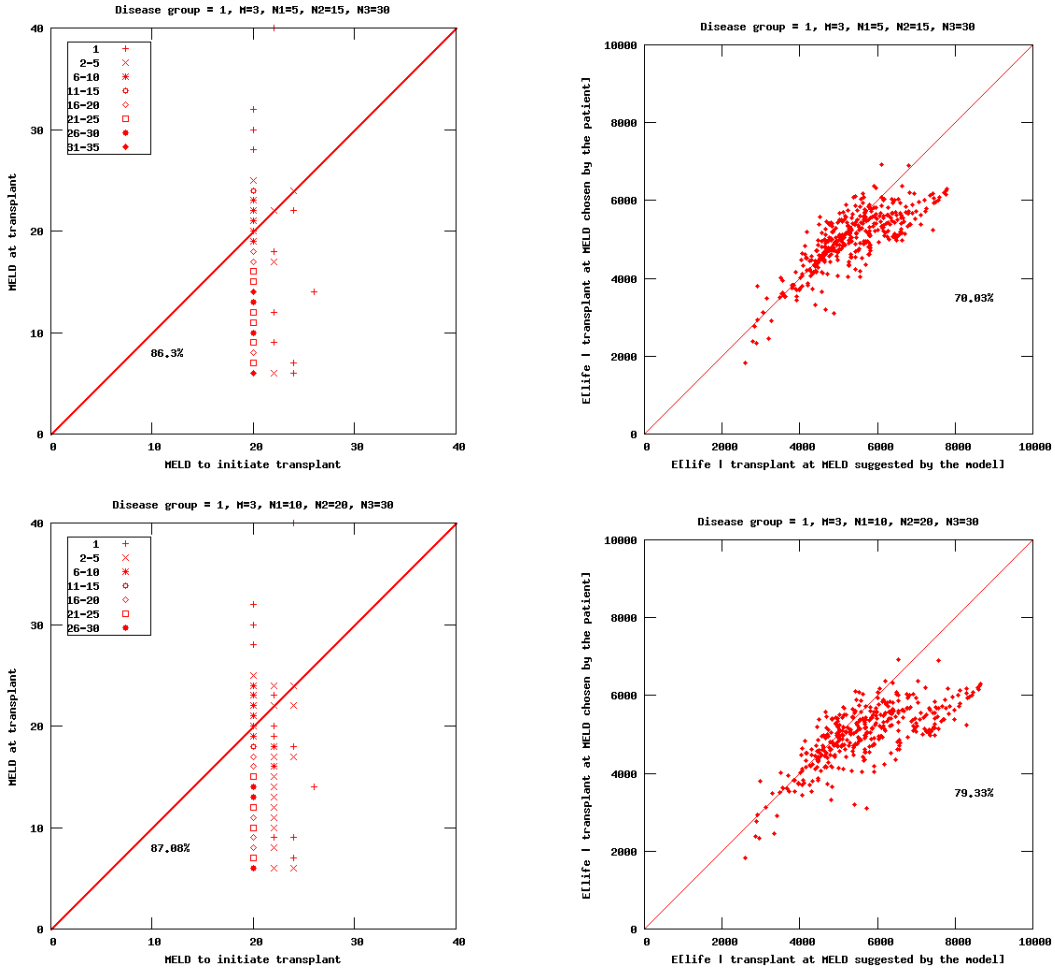


Figure 5.9: Actual decisions compared to decisions suggested by the non-stationary model

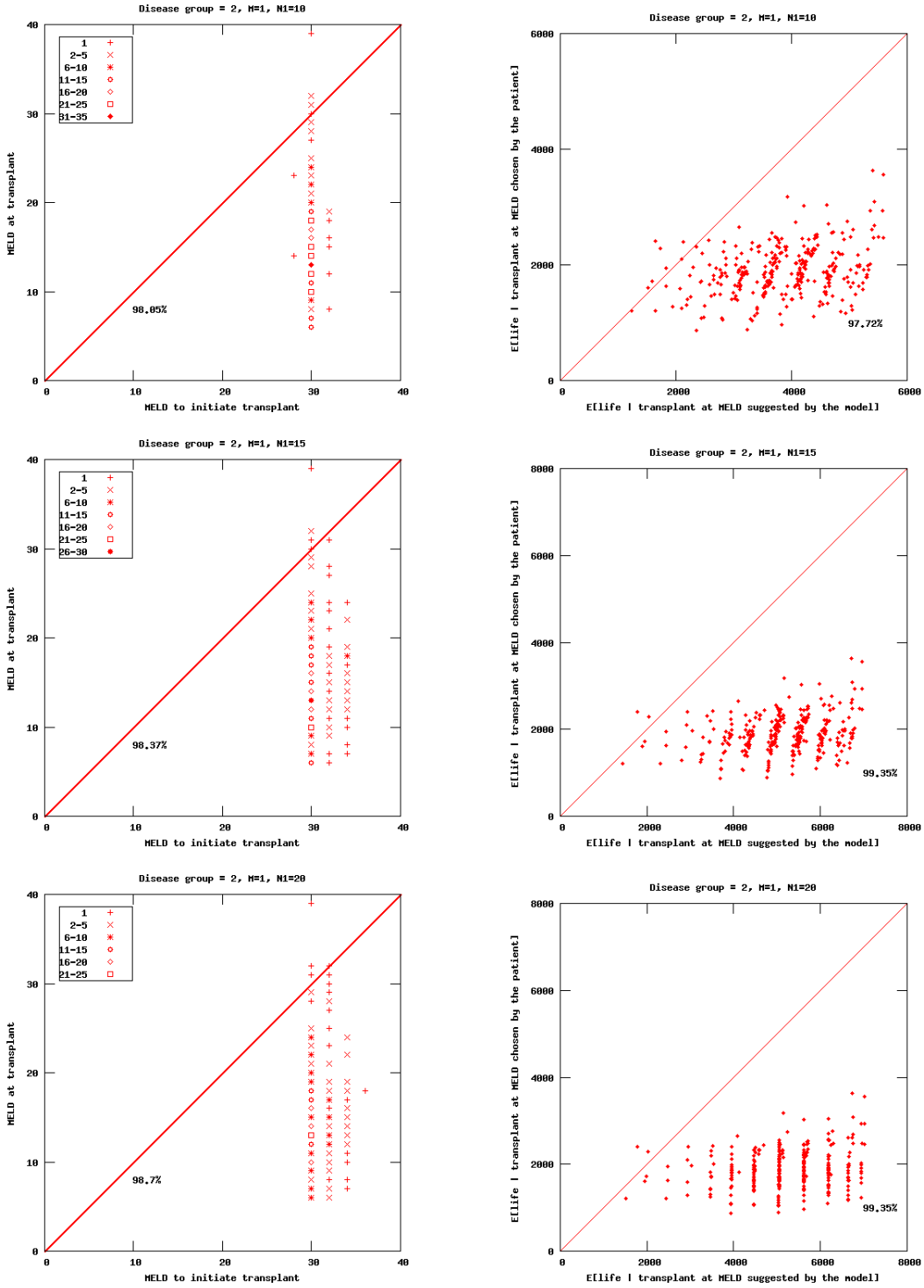


Figure 5.10: Actual decisions compared to decisions suggested by the non-stationary model

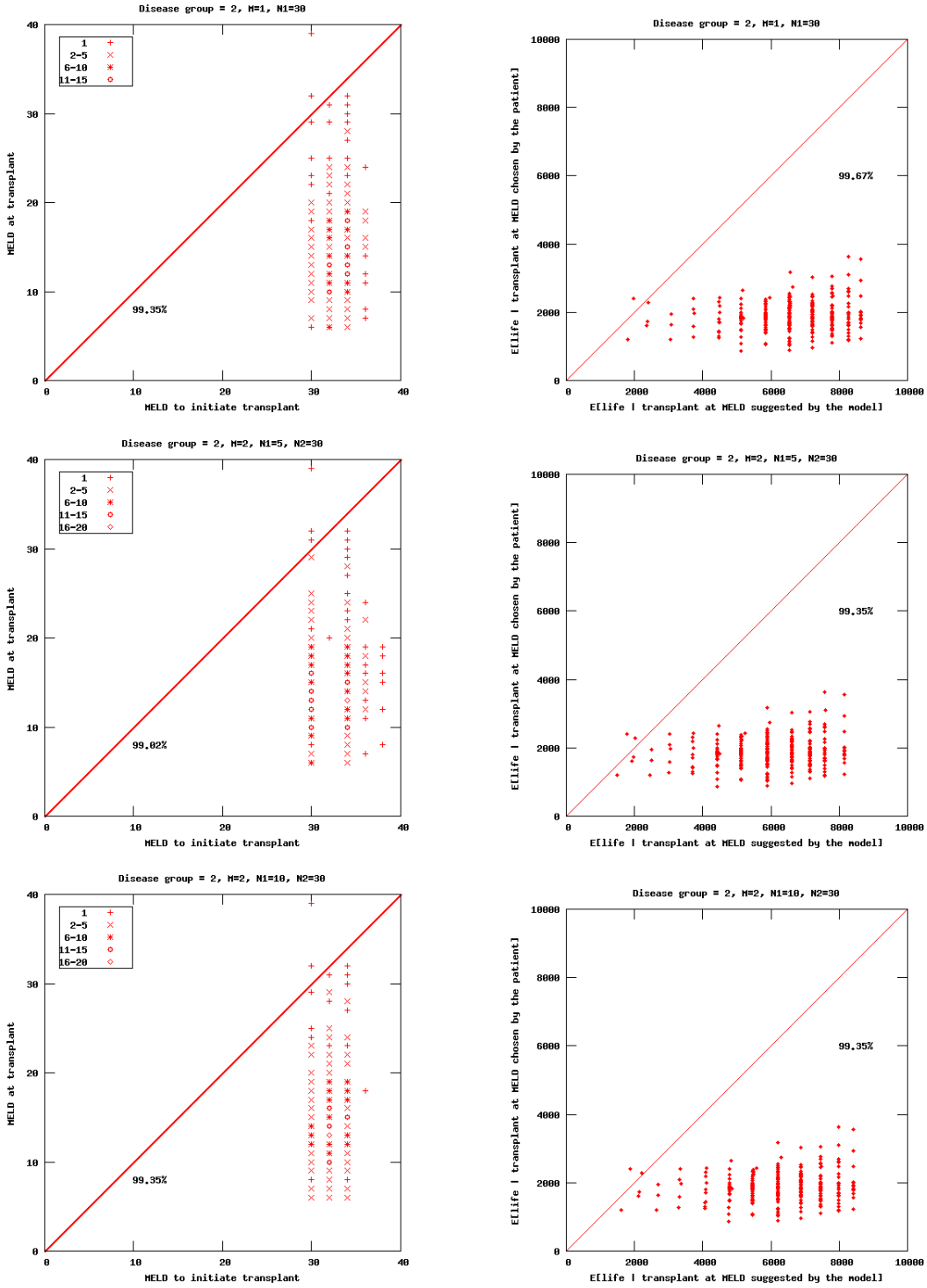


Figure 5.11: Actual decisions compared to decisions suggested by the non-stationary model

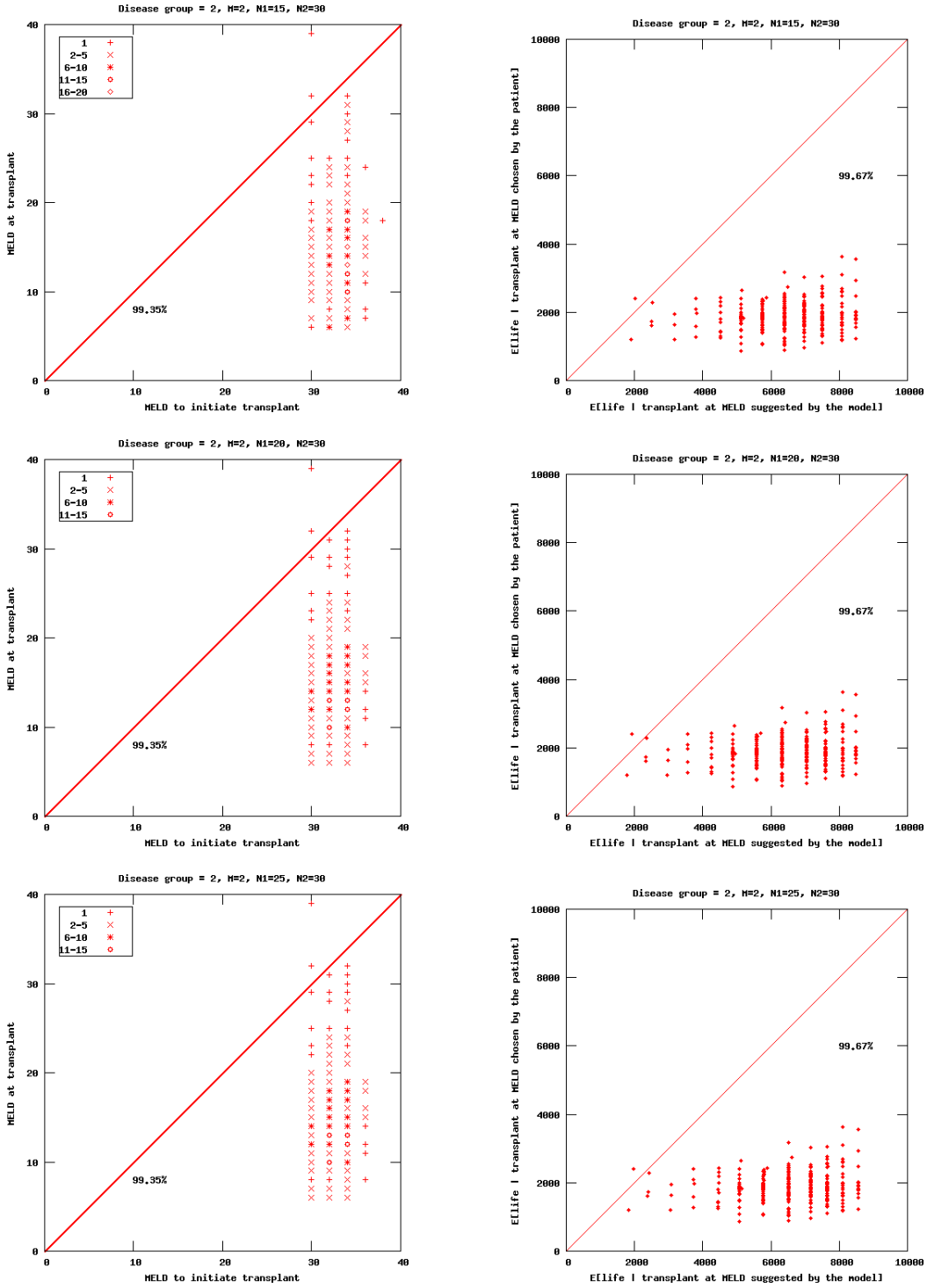


Figure 5.12: Actual decisions compared to decisions suggested by the non-stationary model

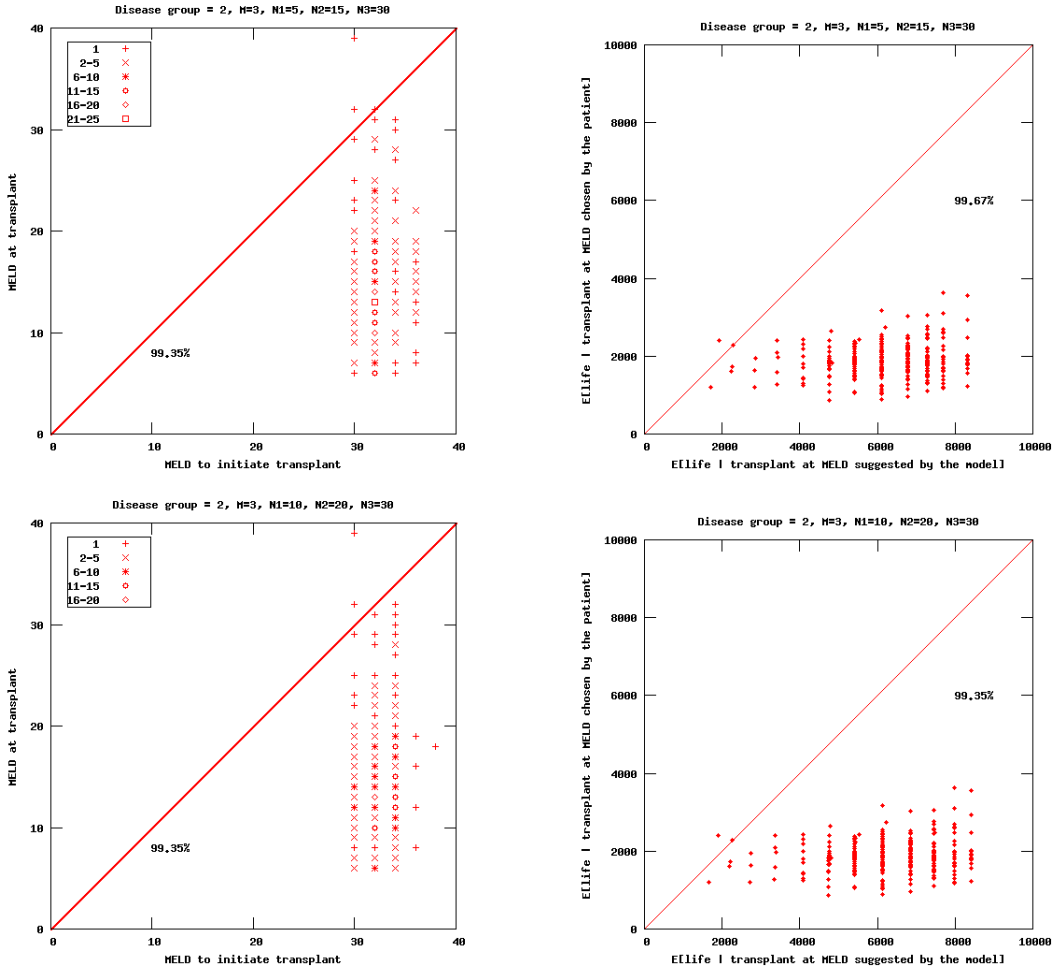


Figure 5.13: Actual decisions compared to decisions suggested by the non-stationary model

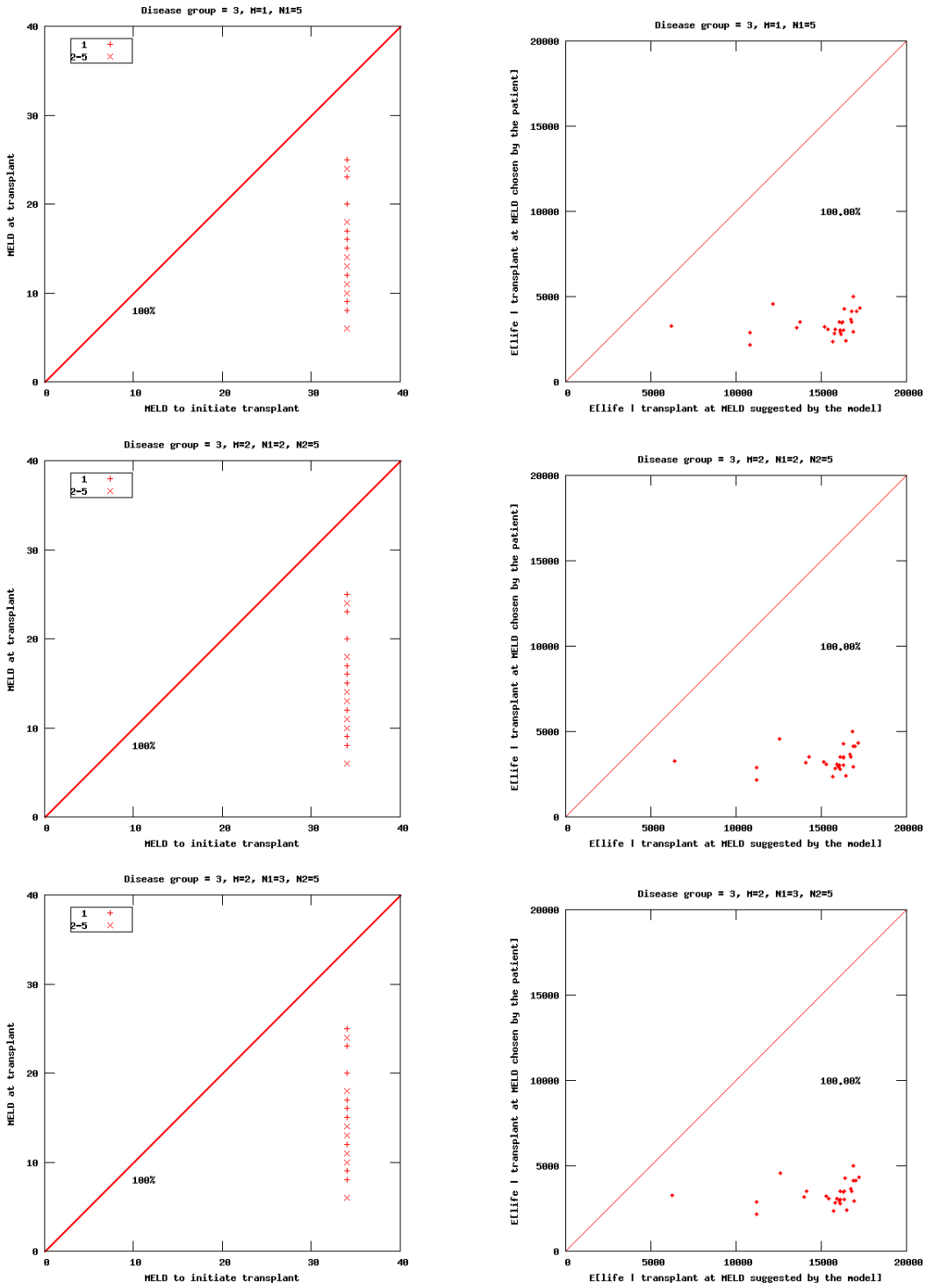


Figure 5.14: Actual decisions compared to decisions suggested by the non-stationary model

Table 5.3: Actual decisions compared to decisions suggested by the non-stationary model ($MELD_L \leq 12$)

Disease group	M	N_m	Total cases	$MELD_T < MELD*$		$EL_T < EL^*$	
				Count	Percentage	Count	Percentage
1	0	-	181	165	91.16%	160	88.40%
	1	10	181	176	97.24	180	99.45
	1	15	181	176	97.24	180	99.45
	1	20	181	176	97.24	180	99.45
	1	30	181	176	97.24	180	99.45
	2	5, 30	181	175	96.69	157	86.74
	2	10, 30	181	176	97.24	176	97.24
	2	15, 30	181	176	97.24	178	98.34
	2	20, 30	181	176	97.24	180	99.45
	2	25, 30	181	176	97.24	180	99.45
	3	5, 15, 30	181	176	97.24	167	92.27
	3	10, 20, 30	181	176	97.24	178	98.34
2	0	-	115	109	94.78%	113	98.26%
	1	10	115	113	98.26	115	100.00
	1	15	115	114	99.13	115	100.00
	1	20	115	114	99.13	115	100.00
	1	30	115	115	100.00	115	100.00
	2	5, 30	115	114	99.13	115	100.00
	2	10, 30	115	115	100.00	115	100.00
	2	15, 30	115	115	100.00	115	100.00
	2	20, 30	115	114	99.13	114	99.13
	2	25, 30	115	115	100.00	115	100.00
	3	5, 15, 30	115	115	100.00	115	100.00
	3	10, 20, 30	115	115	100.00	115	100.00
3	0	-	15	12	80.00%	14	93.33%
	1	5	15	15	100.00	15	100.00
	2	2, 5	15	15	100.00	15	100.00
	2	3, 5	15	15	100.00	15	100.00

M : number of finite horizon MDPs solved
N_m : end point(s) of the first M intervals
$MELD_L$: MELD score in which patient joined the system
$MELD_T$: MELD score in which patient actually received the transplant
$MELD^*$: MELD score in which model suggests to initiate transplantation
EL_T : expected life given patient transplants at $MELD_T$
EL^* : expected life given patient follows the policy suggested by the model

6.0 SUMMARY AND FUTURE RESEARCH

6.1 SUMMARY

This dissertation focuses on the decision problem faced by thousands of end-stage liver disease patients, namely, the problem of optimally deciding which liver offers to accept and which liver offers to reject. There is common agreement in the organ transplantation community that these decisions are significantly affected by the composition of the waiting list, as this composition determines the chances of receiving offers for each patient. The current liver allocation system publishes coarse descriptions of this waiting list, from which patients may obtain their relative chances of receiving offers.

Previous literature [6] incorporates this waiting list information into patient's decision-making implicitly through organ arrival probabilities. In this dissertation, we explicitly incorporate the waiting list information into patient's decision-making. We use Markov decision process models for this purpose. Chapter 3 considers the situation when the waiting list is completely observed by the decision-maker and, hence, she has perfect information about the waiting list. Chapter 4 improves the model presented in Chapter 3 by allowing imperfect information about the waiting list to represent the partially observable nature of this list in the current liver allocation system. After developing the modeling framework for both situations, we examine intuitively appealing and clinically sound structural properties of these models such as monotone value functions and control-limit optimal policies. In proving these structural properties, we introduce a new class of stochastic matrices, namely, the CCD matrices, and investigate its relationship to well-known classes (e.g., IFR and TP_2 matrices) in the literature. Furthermore, for both models, we provide the results of extensive computational experiments parameterized by clinical data. These results indicate

that knowledge of the composition of the waiting list significantly affects the optimal policy. In particular, a patient is much more selective if she knows that she is near the top of the waiting list and becomes gradually less selective as her position deteriorates.

We use both of these models to estimate a patient's price of privacy, which is incurred due to suboptimal decision making from a lack of perfect waiting list information. By comparing the solution of the model provided in Chapter 3 to that of the model provided in [6], we obtain an upper bound on this price. Computational experiments reveal that this bound is typically on the order of 5% of the maximum total discounted life expectancy incurred without explicit consideration of the waiting list information. We also compare the solution of the model provided in Chapter 3 to that of the model provided in Chapter 4 to obtain more precise estimates of the price of privacy. Our grid-based solution methodology for this comparison yields a lower bound on this privacy. Computational experiments reveal that the maximum price of privacy ratio is less than 0.5% over all grid configurations. Although we cannot exactly argue that patients can almost completely eliminate the price of privacy by exploiting the available imperfect waiting list information (because we do not know how tight our grid-based approximation is), we have found significant evidence that the partially observed waiting list can drastically reduce this price.

Finally, we provide a detailed empirical study in Chapter 5 to assess historical liver accept/reject decisions, as our price of privacy estimates in the previous chapters are based on the assumption that patients are maximizing their life expectancy. We provide a non-stationary MDP model for this purpose and compare the decisions suggested by this model to the actual decisions of patients. We find significant evidence that majority of the patients are making suboptimal decisions with respect to the objective of maximizing life expectancy.

6.2 LIMITATIONS AND FUTURE RESEARCH

There are several possible extensions to this dissertation. First, throughout this dissertation, we focused on the objective of maximizing total discounted life expectancy. However, in Chapter 5, we have found significant evidence that patients are actually not acting optimally

with respect to this objective. Consideration of other objectives such as maximizing total quality-adjusted life expectancy might be a better choice and is left for future research. On a related note, it is an interesting research question to identify what objectives patients are optimizing when they are making their accept/reject decisions.

Second, we have assumed risk-neutral patients throughout this dissertation. However, it is true that many patients are indeed risk -averse and some patients may different preferences as to the timing of the transplant operation [34]. Incorporation of patient-specific risk and time preferences into the objective function, which would have enormous data requirements, is also another area for future research.

Third, we have observed in Chapter 5 that the parameter estimates ($\mathcal{H}$ matrix) can dramatically change the optimal policies and the value functions. One potential limitation of our current computational experiments is that the parameter estimates used in these experiments are point estimates. As more data become available, future work should include extensive sensitivity analysis on these point estimates. Also, this limitation can potentially be overcome by incorporating robust dynamic programming techniques [75, 103].

A fourth area for future research is related to the information partitions for the partially observable waiting list model (POWLM). In Chapter 4, we have assumed that the waiting list information is revealed through partitioning the MELD scores into 4 groups (6-10, 11-18, 19-24, and 25-40) as in the current UNOS allocation system. Analyzing this partitioning scheme would be an interesting extension to our current model. How does the price of privacy estimates are affected by changing these partitions? To this end, let p be the number of MELD partitions and partition i include MELD scores in the range $[m_\ell^i, m_u^i]$ for $i = 1, \ldots, p$, where $m_\ell^i \leq m_u^i$, $m_u^i < m_\ell^j$ for $i < j$, and we take $m_\ell^1 = 6$ and $m_u^p = 40$. Finally, let $\Gamma = \{m_\ell^1, m_u^1, m_\ell^2, m_u^2, \ldots, m_\ell^p, m_u^p\}$ be a particular partitioning of the MELD scores and $\phi^i(\Gamma)$ be the maximum total expected discounted life days of patient i obtained through optimizing POWLM with partition Γ. Then, for any given p, the problem of minimizing the price of privacy associated with imperfect waiting list information can be stated as the following nonlinear program:

$$
\begin{aligned}
\max \quad & \sum_i \phi^i(\Gamma) \\
\text{s.t.} \quad & m_\ell^1 = 6 \\
& m_u^p = 40 \\
& m_\ell^j \leq m_u^j \quad \text{for } j = 1, \ldots, p \\
& m_u^j < m_\ell^{j+1} \quad \text{for } j = 1, \ldots, p-1 \\
& m_\ell^j, m_u^j \in \mathbb{Z}.
\end{aligned}
$$

Finally, the public revelation of the waiting list information naturally gives rise to a gaming environment, where each patient, when making their own decisions, has to consider the possible decisions of other patients on the waiting list. Incorporation of this competition between patients requires notoriously difficult stochastic game models. We could also think of extending POWLM to a partially observable stochastic game model as we can imagine such a competition also exists in the current partially observable environment. A comparison of the results of these two stochastic game models would yield the true price of privacy for the entire society. The difficulties associated with characterizing the optimal solutions in these stochastic games may be remedied by a simulation based numerical study as summarized below:

1. Set $n = 0$. Pick an (arbitrary) policy for each patient. Denote this strategy profile as π_0.
2. Set $n = n+1$. Using strategy profile π_{n-1}, simulate the waiting list dynamics to estimate the parameters ($\mathcal{K}_n$ and $\mathcal{L}_n$) of the MDP model provided in Chapter 3.
3. Using $\mathcal{K}_n$ and $\mathcal{L}_n$, find the optimal policies for each patient using the MDP model. Denote the resulting strategy profile as π_n.
4. Repeat steps 2-3 until the stopping criteria is met.

This algorithm may also be used to evaluate a partially observable stochastic game model. For this purpose, we should replace the MDP model in steps (2)-(3) with the POMDP model provided in Chapter 4.

BIBLIOGRAPHY

[1] J.-H. Ahn and J.C. Hornberger. Involving patients in the cadaveric kidney transplant allocation process: A decision-theoretic perspective. *Management Science*, 42(5):629–641, 1996.

[2] O. Alagoz. *Optimal policies for the acceptance of living- and cadaveric-donor livers*. PhD thesis, University of Pittsburgh, 2004.

[3] O. Alagoz, C.L. Bryce, A.J. Schaefer, C.H. Chang, D.C. Angus, and M.S. Roberts. Incorporating biological natural history in simulation models: Empiric estimates of the progression of end-stage liver disease. *Medical Decision Making*, 25(6):620–632, 2005.

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

[5] O. Alagoz, L.M. Maillart, A.J. Schaefer, and M.S. Roberts. Choosing among living-donor and cadaveric livers. *Management Science*, 53(11):1702–1715, 2007.

[6] O. Alagoz, L.M. Maillart, A.J. Schaefer, and M.S. Roberts. Determining the acceptance of cadaveric livers using an implicit model of the waiting list. *Operations Research*, 55(1):24–36, 2007.

[7] S.C. Albright. Structural results for partially observable Markov decision processes. *Operations Research*, 27(5):1041–1053, 1979.

[8] E. Altman. *Constrained Markov Decision Processes*. CRC Press, 1999.

[9] E. Altman. Applications of markov decision processes in communication networks. In E.A. Feinberg and A. Shwartz, editors, *Handbook of Markov Decision Processes: Methods and Applications*, pages 489–536. Kluwer Academic Publishers, 2002.

[10] E. Altman. Applications of dyanmic games in queues. In In A.S. Nowak and K. Szajowski, editors, *Advances in dynamic games: Applications to economics, finance, optimization, and stochastic control*, pages 309–342. Birkhäuser Boston, 2005.

[11] E. Altman and N. Shimkin. Individual equilibrium and learning in processor sharing systems. *Operations Research*, 46(6):776–784, 1998.

[12] K.J. Åström. Optimal control of Markov processes with incomplete state information. *Journal of Mathematical Analysis and Applications*, 10:174–205, 1965.

[13] R.E. Barlow and F. Proschan. *Matematical Theory of Reliability*. John Wiley & Sons, New York, NY, 1965.

[14] H. Behncke. Optimal control of deterministic epidemics. *Optimal Control Applications and Methods*, 21(6):269–285, 2000.

[15] R.E. Bellman. *Dynamic Programming*. Princeton University Press, Princeton, NJ, 1957.

[16] R.S. Bernstein, D.C. Sokal, S.T. Seitz, B. Auvert, J. Stover, and W. Naamara. Simulating the control of a heterosexual HIV epidemic in a severely affected East African city. *Interfaces*, 28(3):101–126, 1998.

[17] D.P. Bertsekas. *Dynamic Programming and Stochastic Control*. Academic Press, New York, NY, 1975.

[18] D.P. Bertsekas. *Dynamic Programming and Optimal Control, Volumes 1 and 2*. Athena Scientific, Belmont, MA, 1995.

[19] D. Blackwell. Discounted dynamic programming. *Annals of Mathematical Statistics*, 36:226–235, 1965.

[20] R. Brafman. A heuristic variable grid solution method for POMDPs. In *Proceedings of the 14th National Conference on Artificial Intelligence (AAAI-97)*, 1997.

[21] S.C. Brailsford, R. Davies, C. Canning, and P. Roderick. Evaluating screening policies for the early detection of retinopathy in patients with non-insulin-dependent diabetes. *Health Care Management Science*, 1:115–124, 1998.

[22] C.C. Branas, E.J. MacKenzie, and C.S. ReVelle. A trauma resource allocation model for ambulances and hospitals. *Health Services Research*, 35(3):489–507, 2000.

[23] M.L. Brandeau, F. Sainfort, and W.P. Pierskalla. *Operations Research and Health Care: A Handbook of Methods and Applications*. Kluwer Academic Publishers, 2004.

[24] R.S. Brown, M.W. Russo, M. Lai, M.L. Shiffman, M.C. Richardson, J.E. Everhart, and J.H. Hoofnagle. A survey of liver transplantation from living adult donors in the United States. *New England Journal of Medicine*, 348(9):818–825, 2003.

[25] A.R. Cassandra. Acting optimally in partially observable stochastic domains. In *Proceedings of the 13th Annual Conference on Uncertainty in Artificial Intelligence (UAI-97)*, pages 472–480, 1997.

[26] A.R. Cassandra. *Exact and approximate algorithms for partially observable Markov decision processes*. PhD thesis, Brown University, 1998.

[27] A.R. Cassandra, L.P. Kaelbling, and M.L. Littman. Acting optimally in partially observable stochastic domains. In *Proceedings of the 12th National Conference on Artificial Intelligence*, pages 1023–1028, 1994.

[28] A.R. Cassandra, M.L. Littman, and N.L. Zhang. Incremental pruning: A simple, fast, exact mehtod for partially observable Markov decision processes. In *Proceedings of the 13th Annual Conference on Uncertainty in Artificial Intelligence (UAI-97)*, 1997.

[29] J.P. Caulkins. Local drug markets' response to focused police enforcement. *Operations Research*, 41(5):848–863, 1993.

[30] J.P. Caulkins. Zero-tolerance policies: Do they inhibit or stimulate illicit drug consumption? *Management Science*, 39(4):458–476, 1993.

[31] M. Chen and R.M. Feldman. Optimal replacement policies with minimal repair and age-dependent costs. *European Journal of Operational Research*, 98(1):75–84, 1997.

[32] F. Cheng and S.P. Sethi. A periodic review inventory model with demand influenced by promotion decisions. *Management Science*, 45(11):1510–1523, 1999.

[33] H.-T. Cheng. *Algorithms for partially observable Markov decision processes*. PhD thesis, University of British Columbia, 1988.

[34] S.H. Chew and J.L. Ho. Hope: An empirical study of attitudes toward the timing of uncertainty resolution. *Journal of Risk and Uncertainty*, 151:949–953, 1994.

[35] National Economic Council. Reforming healthcare for the 21st century, February 2006.

[36] National Research Council. *Understanding risk: Informing decisions in a democratic society*. National Academy Press, 1996.

[37] N. Cuende, B. Miranda, J.F. Canon, G. Garrido, and R. Matesant. Donor characteristics associated with liver graft survival. *Transplantation*, 79(10):1445–1452, 2005.

[38] C. Curran. Adult-to-adult living-donor liver transplantaion: History, current practice, and implications for the future. *Progress in Transplantation*, 15(1):36–42, 2005.

[39] I. David and U. Yechiali. A time-dependent stopping problem with application to live organ transplants. *Operations Research*, 33(3):491–504, 1985.

[40] I. David and U. Yechiali. Sequential assignment match processes with arrivals of candidates and offers. *Probability in the Engineering and Informational Sciences*, 4(4):413–430, 1990.

[41] I. David and U. Yechiali. One-attribute sequential assignment match processes in discrete time. *Operations Research*, 43(5):879–884, 1995.

[42] H. Dawid and G. Feichtinger. Optimal allocation of drug control efforts: A differential game analysis. *Journal of Optimization Theory and Applications*, 91(2):279–397, 1996.

[43] M.A. Dew, G.E. Switzer, J.M. Goycoolea, A. Allen, A. DiMartini, R.L. Kormos, and B.P. Griffith. Does transplantation produce quality of life benefits?: A quantitative analysis of the literature. *Transplantation*, 64(9):1261–1273, 1997.

[44] A. Drake. *Observation of a Markov process through a noisy channel*. PhD thesis, Massachusetts Institute of Technology, 2004.

[45] J.N. Eagle. The optimal search for a moving target when the search path is contsrained. *Operations Research*, 32(5):1107–1115, 1984.

[46] D. Eaton, M. Daskin, D. Simmons, B. Bulloch, and G. Jansma. Determining emergency medical service vehicle deployment in Austin, Texas. *Interfaces*, 15:96–108, 1985.

[47] J.E. Eckles. Optimum maintenance with incomplete information. *Operations Research*, 16(5):1058–1067, 1968.

[48] H. Ellis, M. Jiang, and R.B. Corotis. Inspection, maintenance, and repair with partial observability. *Journal of Infrastructure Systems*, 1(2):92–99, 1995.

[49] E.A. Feinberg and A. Shwartz. *Handbook of Markov Decision Processes: Methods and Applications*. Kluwer Academic Publishers, 2002.

[50] S. Feng, N.P. Goodrich, J.L. Bragg-Gresham, D.M. Dykstra, J.D. Punch, M.A. Debroy, S.M. Greenstein, and R.M. Merion. Characteristics associated with liver graft failure: The concept of a donor risk index. *American Journal of Transplantation*, 6:783–790, 2006.

[51] T.S. Ferguson. *Mathematical Statistics: A Decision Theoretic Approach*. Academic Press, New York, NY, 1967.

[52] J. Filar and K. Vrieze. *Competitive Markov Decision Processes*. Springer, New York, NY, 1996.

[53] Centers for Disease Control and Prevention and The Merck Company Foundation. The state of aging and health in America, 2007. Information available at http://www.cdc.gov/aging and accessed on May 29, 2008.

[54] B.E. Fries. Bibliography of operations research in health-care systems. *Operations Research*, 24(5):801–814, 1976.

[55] D. Fudenberg and J. Tiroli. *Game theory*. MIT Press, 1990.

[56] S. Gavirneni, R. Kapucinski, and S. Tayur. Value of information in capacitated supply chains. *Management Science*, 45(1):16–24, 1999.

[57] L.V. Green and V. Nguyen. Strategies for cutting hospital beds: The impact on patient service. *Health Services Research*, 36:421–442, 2001.

[58] E.A. Hansen. An improved policy iteration algorithm for partially observable MDPs. In *10th Neural Information Processing Systems Conference*, Denver, CO, December 1997.

[59] E.A. Hansen, D.S Bernstein, and S. Zilberstein. Dynamic programming for partially observable stochastic games. In *19th National Conference on Artificial Intelligence (AAAI-04)*, San Jose, CA, July 25-29 2004.

[60] R. Hassin and M. Haviv. *To queue or not to queue: Equilibrium behavior in queueing systems*. Kluwer Academic Publishers, 2003.

[61] M. Hauskrecht. Incremental methods for computing bounds in partially observable Markov decision processes. In *Proceedings of the 14th National Conference on Artificial Intelligence (AAAI-97)*, 1997.

[62] M. Hauskrecht. Value-function approximations for partially observable Markov decision processes. *Journal of Artificial Intelligence Research*, 13:33–94, 2000.

[63] M. Hauskrecht and H. Fraser. Planning treatment of ischemic heart disease with partially observable Markov decision processes. *Artificial Intelligence in Medicine*, 18:221–244, 2000.

[64] P.J.-J. Herings and R.J.A.P. Peeters. Stationary equilibria in stochastic games: Structure, selection, and computation. *Journal of Economic Theory*, 118:32–60, 2004.

[65] D.P. Heyman and M.J. Sobel. *Stochastic Models in Operations Research, Volume II*. Courier Dover Publications, 2003.

[66] HHS/HRSA/HSB/DOT. 2007 OPTN/SRTR annual report 1997-2006, 2007.

[67] D.J. Hockstra. Partially observable Markov decision processes with applications. Technical Report 156, Department of Operations Research, Stanford University, 1973.

[68] J.C. Hornberger and J.-H. Ahn. Deciding eligibility for transplantation when a donor kidney becomes available. *Medical Decision Making*, 17(2):160–170, 1997.

[69] D.H. Howard. Dynamic analysis of liver allocation policies. *Medical Decision Making*, 21:257–266, 2001.

[70] D.H. Howard. Why do transplant surgeons turn down organs? A model of the accept/reject decision. *Journal of Health Economics*, 21:957–969, 2002.

[71] R. Howard. *Dynamic Programming and Markov Processes*. MIT Press, Cambridge, MA, 1960.

[72] C. Hu, W.S. Lovejoy, and S.L. Shafer. Comparison of some suboptimal control policies in medical drug therapy. *Operations Research*, 44(5):696–709, 1993.

[73] IOM. *Organ Procurement and Transplantation: Assessing Current Policies and the Potential Impact of the DHHS Final Rule*. The National Academies Press, Washington, DC, 1999.

[74] J.S. Ivy. A maintenance model for breast cancer detection and treatment, 2002. Submitted for publication.

[75] G.N. Iyengar. Robust dynamic proramming. *Mathematics of Operations Research*, 30(2):257–280, 2005.

[76] T. Jaakkola, S.P. Singh, and M.I Jordan. Montecarlo reinforcement learning in non-markovian decision problems. *Advances in Neural Information Processing Systems*, 7, 1995.

[77] D.A. Jacobs, M.N. Silan, and B.A. Clemson. An analysis of alternative locations and service areas of American Red Cross blood facilities. *Interfaces*, 26:40–50, 1996.

[78] B. Jaumard, F. Semet, and T. Vovor. A generalized linear programming model for nurse scheduling. *European Journal of Operational Research*, 107:1–18, 1998.

[79] P.S. Kamath, R.H. Wiesner, M. Malinchoc, W. Kremers, T.M. Therneau, C.L. Kosberg, G. D'Amico, E.R. Dickson, and W.R Kim. A model to predict survival in patients with end-stage liver disease. *Hepatology*, 33(2):464–470, 2001.

[80] E.P.C. Kao and M. Queyranne. Budgeting costs of nursing in a hospital. *Management Science*, 31(5):608–621, 1985.

[81] E.H. Kaplan, D.L. Craft, and L.M. Wein. Emergency response to a smallpox attack: The case for mass vaccination. *Proceedings of the National Academy of Sciences of the USA*, 99:10395–10440, 2002.

[82] S. Karlin. *Total Positivity*, volume I. Stanford University Press, Stanford, CA, 1968.

[83] D.E. Lane. A partially observable model of decision making by fishermen. *Operations Research*, 37(2):240–254, 1989.

[84] E.K. Lee, T. Fox, and I. Crocker. Integer programming applied to intensity-modulated radiation therapy treatment planning. *Annals of Operations Research*, 119:165–181, 2003.

[85] E.K. Lee and M. Zaider. Mixed integer programming approaches to treatment planning for brachytherapy. *Annal of Operations Research*, 119:147–163, 2003.

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

[87] W.S. Lovejoy. On the convexity of policy regions in partially observed systems. *Operations Research*, 35(4):619–621, 1987.

[88] W.S. Lovejoy. Some monotonicity results for partially observed Markov decision processes. *Operations Research*, 35(5):736–743, 1987.

[89] W.S. Lovejoy. Computationally feasible bounds for partially observed Markov decision processes. *Operations Research*, 39(1):162–175, 1991.

[90] W.S. Lovejoy. A survey of algorithmic methods for partially observed Markov decision processes. *Annals of Operations Research*, 28:47–66, 1991.

[91] W.S. Lovejoy. Suboptimal policies, with bounds, for parameter adaptive decision processes. *Operations Research*, 41(3):583–599, 1993.

[92] P. Magni, S. Quaglini, M. Marchetti, and G. Barosi. Deciding when to intervene: A Markov decision process approach. *International Journal of Medical Informatics*, 60(3):237–253, 2000.

[93] L.M. Maillart. Maintenance policies for systems with condition monitoring and obvious failures. *IIE Transaction*, 38:463–475, 2006.

[94] L.M. Maillart and J.S. Ivy. Assessing dynamic breast cancer screening policies. Working paper, Department of Industial Engineering, University of Pittsburgh, Pittsburgh, PA, 2007.

[95] A. McLennan. The expected number of Nash equlibria of a normal form game. *Econometrica*, 73(1):141–174, 2005.

[96] R.R. Meyer, W.D. D'Souza, M.C. Ferris, and B.R. Thomadsen. MIP models and BB strategies in brachytherapy treatment optimization. *Journal of Global Optimization*, 25:23–42, 2003.

[97] H.E. Miller, W.P. Pierskalla, and G.J. Rath. Nurse scheduling using mathematical programming. *Operations Research*, 24(5):857–871, 1976.

[98] A.M. Minino, M.P. Heron, S.L. Murphy, and K.D. Kochanek. Deaths: Final data for 2004. *National Vital Statistics Reports*, 55(19), 2007.

[99] G.E. Monahan. A survey of partially observable Markov decision processes: Theory, models, and algorithms. *Management Science*, 28(1):1–16, 1982.

[100] D. Moonka, S. Saab, and J. Trotter. Living donor liver transplantation. American Society of Transplantation, Mount Laurel, NJ, May 2007.

[101] J. Neuberger. Should liver transplantation be made available to everyone? *Archives of Internal Medicine*, 163:1881–1883, 2003.

[102] A. Neyman and S. Sorin. *Stochastic Games and Applications*. Kluwer Academic Publishers, 2003.

[103] A. Nilim and L.E. Ghaoui. Robust control of Markov decision processes with uncertain transition matrices. *Operations Research*, 53(5):780–798, 2005.

[104] Commission on Risk Assessment and Risk Management. Framework for environmental health risk assessment. Presidential/Congressional Commission on Risk Assessment and Risk Management, 1997.

[105] OPTN. Transplant milestones in the United States and Canada, 2006. Information available at http://www.optn.org/about/transplantation/history.asp.

[106] OPTN. Organ Procurement and Transplantation Network statistics, 2008. Information available at http://www.optn.org/latestData/viewDataReports.asp and accessed on May 21, 2008.

[107] J.F. Pendergast and W.B. Vogel. A multistage model of hospital bed requirements. *Health Services Research*, 23:381–399, 1988.

[108] S.M. Pollock. A simple model of search for a moving target. *Operations Research*, 18:883–903, 1970.

[109] E. Porteus. Stochastic inventory control. In D.P. Heyman and M.J. Sobel, editors, *Handbook of Operations Research*, pages 605–652. North-Holland, Amsterdam, 1991.

[110] F. Preciado-Walter, R. Rardin, M. Langer, and V. Thai. A coupled column generation, mixed integer approach to optimal planning of intensity modulated radiation therapy for cancer. *Mathematical Programming (Series B)*, 101:319–338, 2004.

[111] A.B. Pritsker. Life and death decisions: Organ transplantation policy analysis. *OR/MS Today*, 25(4):22–28, 1998.

[112] M.L. Puterman. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. John Wiley & Sons, New York, NY, 1994.

[113] H. Raiffa and R.O. Schlaifer. *Applied Statistical Decision Theory*. Harvard University Press, 1961.

[114] J.F. Renz and J.P. Roberts. Long-term complications of living donor liver transplantation. *Liver Transplantation*, 6(6 Suppl. 2):S73–S76, 2000.

[115] A. Richter, M.L. Brandeau, and D.K. Owens. An analysis of optimal resource allocation for HIV prevention among injection drug users and nonusers. *Medical Decisionn Making*, 19:167–179, 1999.

[116] R. Righter. A resource allocation problem in a random environment. *Operations Research*, 37(2):329–338, 1989.

[117] M.S. Roberts, D.C. Angus, C.L. Bryce, Z. Valenta, and L. Weissfeld. Survival after liver transplantation in the United States: A disease-specific analysis of the UNOS database. *Liver Transplantation*, 10(7):886–897, 2004.

[118] H.E. Romeijn, R.K. Ahuja, J.F. Dempsey, and A. Kumar. A new linear programming approach to radiation therapy treatment planning problems. *Operations Research*, 54(2):201–216, 2006.

[119] D. Rosenfield. Markovian deterioration with uncertain information. *Operations Research*, 24(1):141–155, 1976.

[120] S.M. Ross. Quality control under Markovian deterioration. *Management Science*, 17(9):587–596, 1971.

[121] A.E. Roth, T. Sönmez, and M.U. Ünver. Kidney exchange. *Quarterly Journal of Economics*, 119(2):457–488, 2004.

[122] W. Rudin. *Real and Complex Analysis*. McGraw-Hill, New York, NY, 1987.

[123] M. Salizzoni, A. Franchello, F. Zamboni, A. Ricchiuti, D. Cocchis, F. Fop, A. Brunati, and E. Cerutti. Marginal grafts: Finding the correct treatment for fatty livers. *Transplant International*, 16(7):486–493, 2003.

[124] B. Sandıkçı, L.M. Maillart, A.J. Schaefer, O. Alagoz, and M.S. Roberts. Estimating the patient's price of privacy in liver transplantation, 2008. To appear in *Operations Research*.

[125] H.E. Scarf. The optimality of (s, s) policies in the dynamic inventory problem. In K. Arrow, S. Karlin, and P. Suppes, editors, *Studies in the Mathematical Theory of Inventory and Production*, pages 196–202. Stanford University Press, Stanford, CA, 1960.

[126] A.J. Schaefer, M.B. Bailey, S.M. Shechter, and M.S. Roberts. *Operations Research and Health Care: A Handbook of Methods and Applications*, chapter Modeling medical treatment using Markov decision processes, pages 593–612. Kluwer Academic Publishers, 2004.

[127] M. Schal. Markov decision processes in finance and dynamic options. In E.A. Feinberg and A. Shwartz, editors, *Handbook of Markov Decision Processes: Methods and Applications*, pages 461–488. Kluwer Academic Publishers, 2002.

[128] R.F. Serfozo. Monotone optimal policies for Markov decision processes. *Mathematical Programming Studies*, 6:202–215, 1976.

[129] E.C. Sewell and S.H. Jacobson. Using an integer programming model to determine the price of combination vaccines for childhood immunization. *Annals of Operations Research*, 119:261–284, 2003.

[130] S.M. Shechter. *When to initiate, when to switch, and how to sequence HIV therapies: A Markov decision process approach*. PhD thesis, University of Pittsburgh, 2006.

[131] S.M. Shechter, M.D. Bailey, A.J. Schaefer, and M.S. Roberts. The optimal time to initiate HIV therapy under ordered health states. *Operations Research*, 56(1):20–33, 2008.

[132] S.M. Shechter, C.L. Bryce, O. Alagoz, J.E. Kreke, J.E. Stahl, A.J. Schaefer, D.C. Angus, and M.S. Roberts. A clinically based discrete-event simulation of end-stage liver disease and the organ allocation process. *Medical Decision Making*, 25(2):199–209, 2005.

[133] D. Shepard, M. Ferris, G. Olivera, and T. Mackie. Optimizing the delivery of radiation therapy to cancer patients. *SIAM Review*, 41:721–744, 1999.

[134] L.J. Shuman, R.D. Speas, and J.P. Young. *Operations Research in Health Care*. The Johns Hopkins University Press, Baltimore, MD, 1975.

[135] R.D. Smallwod, I. Winstein, and J. Eckles. Quantitative methods in computer-directed teaching systems. Technical report, Department of Engineering Economic Systems, Stanford University, 1967.

[136] R.D. Smallwood. The analysis of economic teaching strategies for a simple learning model. *Journal of Mathematical Psychology*, 8:285–301, 1971.

[137] R.D. Smallwood and E.J. Sondik. The optimal control of partially observable Markov processes over a finite horizon. *Operations Research*, 21(5):1071–1088, 1973.

[138] R.D. Smallwood, E.J. Sondik, and F.L. Offensend. Toward an integrated methodology for the analysis of health-care systems. *Operations Research*, 19(6):1300–1322, 1971.

[139] N.A Smith and R.W. Tromble. Sampling uniformly from the unit simplex. Working paper, Johns Hopkins University, Baltimore, MD, 2004.

[140] M.J. Sobel. Optimal operation of queues. In A.B. Clarke, editor, *Mathematical Methods in Queueing Theory*, pages 233–261. Springer-Verlag, Berlin, 1974.

[141] E.J. Sondik. *The optimal control of partially observable Markov processes*. PhD thesis, Stanford University, 1971.

[142] E.J. Sondik. The optimal control of partially observable Markov processes over the infinite horizon: Discounted costs. *Operations Research*, 26(2):282–304, 1978.

[143] S. Sorin. Discounted stochastic games: The finite case. In In A. Neyman and S. sorin, editors, *Stochastic games and applications*, pages 51–55. Kluwer Academic Publishers, 2003.

[144] S.S. Stidham. Optimal control of admission to a queueing system. *IEEE Transcations on Automatic Control*, AC-30:705–713, 1985.

[145] C.T. Striebel. Sufficient statistics in the optimal control of stochastic systems. *Journal of Mathematical Analysis and Applications*, 12:576–592, 1965.

[146] X. Su and S.A. Zenios. Patient choice in kidney allocation: The role of the queueing discipline. *Manufacturing & Service Operations Management*, 6(4):280–301, 2004.

[147] X. Su and S.A. Zenios. Patient choice in kidney allocation: A sequential stochastic assignment model. *Operations Research*, 53(3):443–455, 2005.

[148] X. Su and S.A. Zenios. Recipient choice can address the efficiency-equity trade-off in kidney transplantation: A mechanism design model. *Management Science*, 52(11):1647–1660, 2006.

[149] Suresh. Sampling from the simplex, 2005. Information available at http://geomblog.blogspot.com/2005/10/sampling-from-simplex.html and accessed on May 23, 2008.

[150] S. Swani, M.L. Puterman, and C.B. Weinberg. Play it again, Sam? Optimal replacement policies for a motion picture exhibitor. *Manufacturing & Service Operations Management*, 3(4):369–386, 2001.

[151] S. Thiebeaux, M.-O. Cordier, O. Jehl, and J.-P. Krivine. Supply restoration in power distribution systems: A case study in integrating model-based diagnosis and repair planning. In *Proceedings of the 12th Annual Conference on Uncertainty in Artificial Intelligence (UAI-1996)*, pages 525–532, 1996.

[152] C.S.R. Toregas, C. ReVelle, and L. Bergman. The location of emergency service facilities. *Operations Research*, 19(6):1363–1373, 1971.

[153] G. Tragler, J.P. Caulkins, and G. Feichtinger. Optimal dynamic allocation of treatment and enforcement in illicit drug control. *Operations Research*, 49(3):352–362, 2001.

[154] J.T. Treharne and C.R. Sox. Adaptive inventory control for nonstationary demand and partial information. *Management Science*, 48(5):607–624, 2002.

[155] UNOS. UNOS Ethics Committee general considerations in assessment for transplant candidacy. Available at http://www.unos.org/resources/bioethics.asp?index=5 and accessed on May 27, 2008.

[156] UNOS. Policy 3.6: Allocation of livers, 2007. Information available at http://www.unos.org/resources/ and accessed on October 7, 2007.

[157] UNOS. View data reports, 2008. Information available at http://www.unos.org/data.

[158] G. Vassilacopoulis. Allocating doctors to shifts in an accident and emergency department. *Journal of the Operational Research Society*, 36:517–523, 1985.

[159] N. Vieille. Stochastic games: Recent results. In In R.J. Aumann and S. Hart, editors, *Handbook of Game Theory with Economic Applications*, pages 1833–1850. Elsevier, 2002.

[160] J. Walrand. Queueing networks. In D.P. Heyman and M.J. Sobel, editors, *Handbook of Operations Research*, pages 519–604. North-Holland, Amsterdam, 1991.

[161] D.M. Warner. Scheduling nursing personnel according to nursing preference: A mathematical programming approach. *Operations Research*, 24(5):842–856, 1976.

[162] E.N. Weiss. Models for determining estimated start times and case ordering in hospital operating rooms. *IIE Transaction*, 22:143–150, 1990.

[163] C.C. White. Optimal diagnostic questionnaires which allow less than truthful responses. *Information and Control*, 32:61–74, 1976.

[164] C.C. White. Procedures for the solution of finite-horizon, partially observed, semi-Markov optimization problem. *Operations Research*, 24(2):348–358, 1976.

[165] C.C. White. A Markov quality control process subject to partial observation. *Management Science*, 23(8):843–852, 1977.

[166] C.C. White. Optimal inspection and repair of a production process subject to deterioration. *Journal of the Operational Research Society*, 29(3):235–243, 1978.

[167] C.C. White. Optimal control-limit strategies for a partially observed replacement problem. *International Journal on Systems Science*, 20:321–331, 1979.

[168] C.C. White. Monotone control laws for noisy, countable-state markov chains. *European Journal of Operational Research*, 5:124–132, 1980.

[169] C.C. White. Structured policy results for a single stage decisionmaking under uncertainty. *IEEE Transactions on Systems, Man & Cybernetics*, 10:891–894, 1980.

[170] C.C. White and W.T. Scherer. Finite-memory suboptimal design for partially observed Markov decision processes. *Operations Research*, 42(3):439–455, 1994.

[171] D.J. White. *Markov Decision Processes*. John Wiley & Sons, New York, NY, 1993.

[172] W. Whitt. A note on the influence of the sample on the posterior distribution. *Journal of The American Statistical Association*, 74(366):424–426, 1979.

[173] W. Whitt. Improving service by informing customers about anticipated delays. *Management Science*, 45(2):192–207, 1999.

[174] R.H. Wiesner, S.V. McDiarmid, P.S. Kamath, E.B. Edwards, M. Malinchoc, W.K. Kremers, R.A. Krom, and W.R Kim. MELD and PELD: Application of survival models to liver allocation. *Liver Transplantation*, 7(7):567–580, 2001.

[175] J.T. Wu, L.M. Wein, and A.S. Perelson. Optimization of influenza vaccine selection. *Operations Research*, 53(3):456–476, 2005.

[176] F. Yokota and K.M. Thompson. Value of information literature: A review of applicatiosn in health risk management. *Medical Decision Making*, 24:287–298, 2004.

[177] S.A. Zenios. Modeling the transplant waiting list: A queuing model with reneging. *Queuing Systems*, 31:239–251, 1999.

[178] S.A. Zenios. Optimal control of a paired-kidney exchange program. *Management Science*, 48(3):328–342, 2002.

[179] S.A. Zenios. Models for kidney allocation. In M.L. Brandeau, F. Sainfort, and W.P. Pierskalla, editors, *Operations research and health care: A handbook of methods and applications*, pages 537–554. Kluwer Academic Publishers, 2004.

[180] S.A. Zenios, G.M. Chertow, and L.M. Wein. Dynamic allocation of kidneys to candidates on the transplant waiting list. *Operations Research*, 48(4):549–569, 2000.

[181] S.A. Zenios, L.M. Wein, and G.M. Chertow. Evidence-based organ allocation. *American Journal of Medicine*, 107(1):52–61, 1999.

[182] W. Zhang. *Algorithms for partially observable Markov decision processes*. PhD thesis, The Hong Kong University of Science and Technology, 2001.

[183] R. Zhou and E.A. Hansen. An improved grid-based approximation algorithm for POMDPs. In *Proceedings of the 17th International Conference on Artificial Intelligence*, 2001.