Living-Donor Liver Transplantation Timing Under Ambiguous Health State Transition Probabilities

David L. Kaufman

College of Business, University of Michigan-Dearborn

davidlk@umich.edu

Andrew J. Schaefer

Computational and Applied Mathematics, Rice University

andrew.schaefer@rice.edu

Mark S. Roberts

School of Medicine, University of Pittsburgh

robertsm@upmc.edu

June 28, 2017

Abstract: Markov decision process (MDP) models for the optimal time to initial a medical therapy, such as an organ transplantation, require the estimation of health state transition probabilities from physiological data. Such estimation may be a source of probabilistic ambiguity when, for example, some critical health states are seldom visited historically. For MDP models in general, robust dynamic programming has been proposed as an approach to mitigate the effects of ambiguity on optimal decisions. However, very few real-world studies examining the usefulness of robust MDP policies have been reported. We present a robust MDP model for medical therapy initiation in which worst-case transition probabilities are chosen from a set of probability measures constructed using relative entropy bounds. For this model, we prove that therapy is initiated sooner, in additional states, as the ambiguity increases. We apply the methodology to the problem of deciding when to undergo a living-donor liver transplantation, and present the results of a case study using clinical data. We propose a novel implied confidence level measure that maps the robust solutions to historical transplant decisions, and find that in some cases the robust policies are closer to decisions that have been made in practice.

Keywords: Markov decision processes, medical decision-making, robust optimization, dynamic programming

1. Introduction

Physicians often face sequential treatment decisions under conditions of uncertainty. It is perhaps natural that such decisions be modeled as Markov decision processes (MDPs). When there are many possible therapeutic decisions, randomized control trials, the standard method of evaluating different therapeutic policies, are not appropriate for identifying the best policy among all options.

MDPs, on the other hand, allow one to evaluate many treatment strategies with little time, cost, or risk to patients.

One of the most interesting (and vexing) problems encountered in building MDP models for medical therapies, such as an organ transplantation, is the problem of developing quantitative natural history models that represent the progression of a disease over time (e.g., Alagoz et al. 2005). Such models require the estimation of underlying state transition probabilities from physiological data. In some cases, however, clinical data are not available in abundance. There may be several reasons for this. For one, the number of patients with a particular disease and relevant characteristics might be small. Also, clinical data are often very expensive to collect, especially when additional observations and lab tests are needed beyond those of standard care. On another level, even if the amount of data in total is large, it may be the case that some states of disease progression which factor into the decision tradeoffs significantly are seldom visited historically. That is, the data used in estimating transition probabilities may be limited, resulting in transition probability measures that are not certain, but rather are *ambiguous*.

For all data-driven medical decision problems, there is some degree of ambiguity (i.e., epistemic uncertainty) in parameter estimates. Unfortunately, MDP solutions are "quite sensitive to perturbations in the transition probability and ignoring the estimation errors can lead to serious degradation in performance" (Iyengar 2005). These problems could be acute when the transition probabilities describe disease progression. Thus far however, the vast majority of stochastic models of disease progression have implicitly assumed that the parameters governing the underlying stochastic processes are known with certainty. Instead, it may be important to explicitly account for ambiguity in the decision-making process. Several researchers, including Satia and Lave (1973), White and Eldieb (1994), Givan et al. (2000), Bagnell et al. (2001), Iyengar (2005), and Nilim and El Ghaoui (2005) have proposed a max-min optimization framework known as "robust dynamic programming" (robust DP) as a systematic and statistically meaningful way to mitigate the effects of ambiguity in MDPs. But while there has been significant interest in the methodological aspects of robust DP, there are very few reports of real-world studies, in general, examining the merits of using the methodology in practice. Our results appeared in abstract form in Kaufman et al. (2011), which is the first report of an application in health care.

Even though the medical decision-making arena in particular seems ripe for the application of robust DP, without case studies that examine its potential benefits, practitioners are left with basic

questions answered, including:

- Since every real-world decision is made in the face of some level of ambiguity, it is beneficial to use the robust DP framework, or is it simply better to use the traditional MDP approach with maximum likelihood estimates for transition probabilities, which we refer to as the "myopic" model?
- How does one calibrate robust DPs to real-world data?
- Are the "robust" DP solutions actually robust to ambiguity?
- How does one measure performance, to compare the robust and myopic approaches?
- Can clinical decisions observed in the real world be explained, at least in part, by the presence of ambiguity?

We consider the optimal timing of a living-donor liver transplantation. This application has been studied previously, by Alagoz et al. (2004), but we are the first to model ambiguity in the transplant decision process. We use clinical data to calibrate myopic and robust models and compare their policies, and we evaluate performance with a simulation study. We then compare the actual timing decisions of patients who have undergone a living-donor liver transplantation to the optimal robust DP policies. We report a novel implied confidence level measure that maps the robust solutions to historical transplant decisions. Our primary findings are as follows:

- We prove that for our robust DP model it is optimal to transplant sooner, in better states of health, as the level of ambiguity in transition probabilities increases.
- In practice, both robust and myopic solutions may be characterized by a health-state threshold policy. The region between the robust and myopic thresholds may provide a *range* of acceptable transplant health states, ruling out overly conservative actions.
- From a performance standpoint, our simulation study shows that while robust policies may perform better in some replications, the myopic policies actually perform very well on average.
- We find that in some cases (26.7% of the population for a range of confidence levels) the robust policies are closer to transplant decisions that have been made in practice.
- Still, there is a significant portion of the population (53.5% in our study) who transplant even sooner than what the most conservative robust DP solutions suggest.

The rest of the paper proceeds as follows. Next we discuss our motivation and review the related literature. In Sec. 3 we present both myopic and robust models for therapy initiation, and

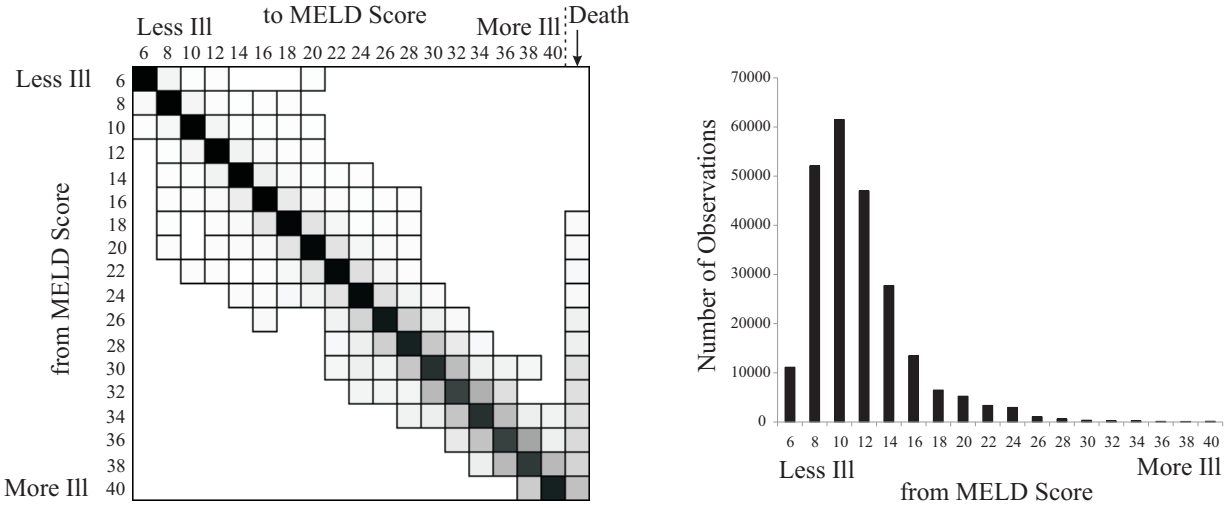
we present structural results. In Sec. 4 we consider the application to living-donor transplantation timing and present our case study comparing the optimal policies of the myopic and robust models and their performance; and, we introduce the implied confidence level that we use to measure the level of conservativeness in historical living-donor transplants. Sec. 5 summarizes our conclusions.

2. Background and Literature Review

We are motivated by our work modeling the decisions faced by patients with end-stage liver disease, for which the only viable treatment is liver transplantation. For this treatment it must be decided when to undergo a liver transplantation by taking into account the balance between the patient's pre-transplant discounted future life expectancy and post-transplant discounted life expectancy. Some patients are fortunate to have the sure option of receiving a portion of a liver donated by a live donor. For other patients there is the possibility of a cadaveric donor organ, for which offers arrive randomly at a rate that depends on the patient's position on a regional registry list. In the United States, cadaveric organ offers are made according to a computerized organ sharing system that is maintained by UNOS (the United Network for Organ Sharing). In addition to maintaining registries, UNOS maintains an extensive patient database that we use to estimate post-transplant life expectancy.

Our focus is on the Living-Donor Model (**LDM**) of Alagoz et al. (2004), an MDP model for the optimal time to undergo a liver transplantation when a patient is only considering a living-donor organ. In the LDM, transplanting right away is typically suboptimal. The total reward, i.e., the pre-transplant plus post-transplant life expectancy, is often greater if the patient waits. This is true despite the fact that post-transplant life expectancy decreases as the patient's health worsens. The post-transplant reward depends on the quality of the donor organ. Typically, when the quality of the donor organ is lower, the incentive to wait is higher (see Figure 3 below). Disease and patient type also affect the optimal policies. For example, older patients should typically wait longer. Similar patterns carry over to MDP models for liver transplantation timing that account for cadaveric liver offers – the models of Alagoz et al. (2007a,b) and Sandıkçı et al. (2008).

Underlying each of these liver transplantation MDP models is the natural history model of Alagoz et al. (2005), a Markov chain model of the pre-transplant progression of an end-stage liver disease. Figure 1(a) depicts the transition probability matrix for the progression of Hepatitis infections, one of five LDM disease groups. The states associated with this matrix are MELD



(a) Transition probability matrix

(b) Number of observations used for estimation

Figure 1: (a) A depiction of a transition probability matrix for the LDM, Disease Group 2 – Hepatitis infections. The rows correspond to the MELD score of the current day, and the columns correspond to the MELD score of the next day. The last column is patient death. The darker squares represent transition probabilities closer to 1. The lighter squares represent transition probabilities closer to 0. (For accentuation, the grayscale is applied to the square root of the probability. Otherwise, the off-diagonal elements are much whiter.) The lack of a square indicates that no such transition was observed historically. (b) The total number of historical observations used to estimate the transition probabilities, by row.

(Model for End-Stage Liver Disease) scores – integer values that represent the state of health of the patient. MELD scores are also used by UNOS to prioritize cadaveric liver allocations. MELD is a function of several laboratory values: total bilirubin, creatinine, and prothrombin time (Malinchoc et al. 2000). These scores are ordered in the range of 6 to 40 MELD, with a higher value indicating that the patient is more ill. In the LDM, MELD scores are aggregated by twos. The transition matrix models the health-state progression for a patient who chooses not to yet initiate therapy. The rows corresponds to the patient's health state on the current day and the columns correspond to the patient's health state on the next day. The last column represents patient death, an absorbing state. The matrix probabilities correspond to daily transitions and are estimated from clinical observations. The data set is from the Thomas E. Starzl Transplantation Institute at the University of Pittsburgh Medical Center, one of the largest liver transplant centers in the world. The observations are the result of applying a cubic spline methodology to clinical scores of a population of patients with similar characteristics to smooth historical data and account for missing values (Alagoz et al. 2005).

Figure 1(a) shows that health is not a skip-free process; end-stage liver diseases exhibit acute exacerbations and recoveries. Most of the probability mass is concentrated along the diagonal:

tomorrow's health state will most likely be the same as today's. The first observed death was at 18 MELD. The probability of death increases as the MELD score increases. This matrix nearly exhibits an *increasing failure rate* (IFR), implying, "The sicker the patient, the more probable the patient will become even sicker" (Alagoz et al. 2004).

Figure 1(b) shows the number of historical observations, by MELD score, that were used to estimate the transition probabilities. The total number of observations across all scores is 233,817. The largest number of observations is for 10 MELD at 61,499. There are fewer observation for 6 MELD, because not all patients present at this low score. The number of observations decreases significantly at the higher MELD scores. For 40 MELD, there are only 142 observations, of which 14 are deaths. On one hand, patients at 40 MELD may not live long, or patients die prior to reaching 40 MELD. This is an example of informed censoring. With so few observations, one cannot be certain that the probability of death at 40 MELD is 14/142.

We consider two models for therapy initiation. The first, the **myopic model**, also sometimes called a "nominal" model, is a traditional DP model that uses point estimates for transition probabilities. The myopic model is a stopping time problem that has been studied previously. The general form of the myopic model has multiple medical decision-making applications (Alagoz et al. 2004, Shechter et al. 2008, Kreke et al. 2008); it is not limited to problems involving liver transplantations. The second model, the **robust model** model, considers ambiguity in transitions. In our case study, we apply these models to the LDM.

Rather than relying on points estimates, robust DP constructs statistically meaningful sets of probability measures. These so-called "uncertainty sets," denoted $\mathcal{P}$, are constructed corresponding to a specified confidence level. The more ambiguity, the larger the $\mathcal{P}$. Robust DP can be viewed as a game between the decision-maker and an adversary called "nature." While the decision-maker seeks to maximize total reward, nature seeks to minimize the decision-maker's reward by selecting worst-case transition probability measures in $\mathcal{P}$.

The choice of uncertainty set construction is important. Perhaps the simplest model for uncertainty sets is a finite collection of transition matrices (Nilim and El Ghaoui 2005). However, it is not clear how to construct finite scenarios that are good representations of statistical uncertainty. Satia and Lave (1973), White and Eldieb (1994), and Givan et al. (2000) instead consider the transition matrix to lie in a given polytope. For general polytopes however, computation can be challenging. An exception is the interval matrix model proposed by Nilim and El Ghaoui (2005) that constrains

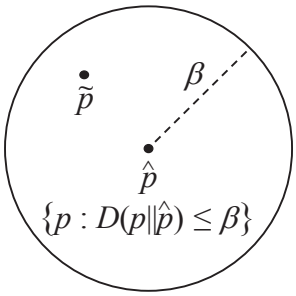


Figure 2: Illustration of uncertainty set construction. A feasible transition probability row (with nonnegative elements that sum to one) is denoted by p. The best (maximum likelihood) estimate from the data is $\hat{p}$. We denote the true, but unknown, transition probability row as $\tilde{p}$, which may or may not be in the uncertainty set. Uncertainty sets are of the form $\mathcal{P} \equiv \{p : D(p||\hat{p}) \leq \beta\}$, where β is calibrated such that $\mathbb{P}\{\tilde{p} \in \mathcal{P}\} = \omega$, for some specified confidence level ω.

the components of a transition matrix row to intervals, plus a constraint that the probabilities sum to one. Still, "polytopic models, especially interval matrices, may be very poor representations of statistical uncertainty and lead to very conservative robust policies" (Nilim and El Ghaoui 2005). Nilim and El Ghaoui (2005) also consider an entropy model, a log-likelihood model, and ellipsoidal methods (using a second-order approximation to a log-likelihood function around the maximum-likelihood estimate). Of these, the most promising in terms of good representations of statistical uncertainty seem to be the related entropy and log-likelihood models, which are both based on so-called relative entropy.

The Kullback-Leibler divergence between two distributions, or *relative entropy*, is "a measure of the 'distance' between two probability mass functions p and q" (Cover and Thomas 2006, p. 9) and is a well-studied concept in information theory. In statistics, it "arises as the exponent in the probability of error in a hypothesis test between two distributions. It is a natural measure of distance between distributions" (Cover and Thomas 2006, p. 11). Though relative entropy, denoted by $D(p||q)$ (defined in (5)), is sometimes referred to as a "distance," it is not a symmetric measure. "The relative entropy $D(p||q)$ is a measure of the inefficiency of assuming that the distribution is q when the true distribution is p" (Cover and Thomas 2006, p. 19).

Bagnell et al. (2001), Iyengar (2005), and Nilim and El Ghaoui (2005) all propose using relative entropy for uncertainty set construction, though only Iyengar (2005) and Nilim and El Ghaoui (2005) propose a method to solve nature's corresponding optimization problem – the so-called "inner problem" (defined in (4)). Uncertainty set construction is illustrated in Figure 2. Let $\hat{p}$ represent the best (maximum likelihood) estimate of the true, but unknown, transition probability measure – a single row of the true transition matrix. An uncertainty set for the row, $\mathcal{P}$, is constructed as the

set of all possible transition probability rows p such that $D(p||\hat{p}) \leq \beta$, where β is an upper bound that depends on the amount of information available in the estimate and a specified confidence level, which we denote by ω (e.g., $\omega = 95\%$). Let $\tilde{p}$ denote the *true* (but unknown) transition probability row. Then, we want to set β such that $\mathbb{P}(\tilde{p} \in \mathcal{P}) = \omega$. As explained in Sec. 3, β should be increasing in ω and depend on both the dimension of $\hat{p}$ and the amount of data available to estimate $\hat{p}$, with β decreasing in the number of data samples. So, β should be small, and hence $\mathcal{P}$ should be small, when the amount of information is large.

The entropy model imposes β as an upper bound on $D(p||\hat{p})$, as just described. Alternatively, the log-likelihood model of Nilim and El Ghaoui (2005) imposes an upper bound on the divergence $D(\hat{p}||\tilde{p})$, using $\tilde{p}$ as the reference distribution (that is, $q = \tilde{p}$ in $D(p||q)$). Instead, the use of $\hat{p}$ as the reference distribution in the entropy model is preferred for several reasons. Perhaps most importantly, for the entropy model, an asymptotic result pointed out by Iyengar (2003) and Ben-Tal et al. (2013) provides an elegant way to calibrate β and construct statistically meaningful uncertainty sets (see Appendix A). Also, the dual problem formulation of nature's inner problem is nicely tractable and computationally easy to implement (equation (7)).

2.1 Literature Review

Prior to initiating a one-time therapy, a patient, or a physician operating on the patient's behalf, faces a sequence of decisions: should the patient start therapy now, or should she wait? Previously studied MDP models for a therapy initiation include the optimal timing of an organ transplant (e.g., David and Yechiali 1985, Ahn and Hornberger 1996, Hornberger and Ahn 1997, Alagoz et al. 2004, 2007a,b, Sandıkçı et al. 2008) and the decisions faced by a physician treating an HIV patient with HAART, a therapy that should not be stopped once started (Shechter et al. 2008). Similarly, physicians treating patients with pneumonia-related sepsis have to decide when to discharge patients in order to maximize their life expectancy (Kreke et al. 2008). Also, patients with Type 2 diabetes must decide when to initiate statins (Denton et al. 2009, Kurt et al. 2011).

Real-world applications require data to estimate MDP model parameters, which introduces bias and variance in the value functions (Mannor et al. 2007). As discussed above, several researchers (Satia and Lave 1973, White and Eldieb 1994, Givan et al. 2000, Bagnell et al. 2001, Iyengar 2005, Nilim and El Ghaoui 2005) have proposed robust DP as a way to mitigate the effects of ambiguity. There are also more recent methodological extensions. Kaufman and Schaefer (2013) introduce a "robust modified policy iteration" (RMPI) algorithm, which we use for our numerical experiments

because (1) RMPI converges faster than value iteration in practice and (2) a stopping condition for ϵ-optimality is provided that does not require the inner problems to be solved exactly in order for convergence to be guaranteed.

In the general robust DP framework, nature's choices of transition probability rows are independent across states and actions. That is, nature's choices are uncoupled. This independence is referred to as the "rectangularity assumption" (Iyengar 2005, Nilim and El Ghaoui 2005). Both Wiesemann et al. (2013) and Mannor et al. (2016) have introduced milder rectangularity conditions by allowing some coupling. It is important to note that our application to therapy initiation is simpler than other applications though, because nature's choices only apply prior to therapy initiation. The decision process stops once therapy is initiated. Hence, there is no need to couple across different actions. On the other hand, one could conceivably couple across states. One interesting idea along these lines is the "k-rectangular uncertainty" of Mannor et al. (2016). For example, it may be possible to constrain nature to deviate from nominal transition probabilities in at most k rows. Mannor et al. (2016) present two stylized examples to demonstrate that such k-rectangular solutions may outperform the traditional robust DP solutions for some parameter settings. Their framework applies to finite horizon problems though, and it remains an open question whether similar results could extend to an infinite-horizon setting like our therapy initiation problem. For other work related to robust DP, including other robust optimization frameworks, we refer the reader to the references in Mannor et al. (2016).

Despite the promise of robust DP, there are few reports of applications. Bagnell et al. (2001) consider the simulated path taken by a robot around an uncertain obstacle and report that the robust policies have some desirable characteristics. They also give a "Mountain-Car" example. Nilim and El Ghaoui (2005) consider the two-dimensional path of an airplane around a storm that moves according to a two-dimensional transition matrix. The technical report of Iyengar (2003) presents some numerical experiments for a stopping time problem with randomly regenerated parameters in which nominal and robust policies are compared under "relative nominal performance" and "relative worst case performance" (the robust DP's max-min objective) measures. In contrast, we evaluate performance by considering expected reward under some true underlying distribution (Sec. 4). Kaufman and Schaefer (2013) present some stylized problems in inventory control to demonstrate the computational efficiencies of RMPI.

An abridged version of our results, without the full case study presented here, appeared in

abstract form in Kaufman et al. (2011), which to the best of our knowledge is the first report of an application of robust DP to a problem in medical decision-making, or health care in general. In the current paper, we present simulation experiments, calibrated using clinical data for the LDM, to study the merits of robust DP. We also present an expanded study comparing the robust DP solutions to actual historical transplantation decisions through an implied confidence level, which is unique to our work. Before presenting our case study, in the next section we present the decision models and provide some structural results.

3. Modeling Framework and Structural Properties

In this section we present the myopic and robust models for the optimal time to initiate a medical therapy. These stopping time models are of a general form, with potentially many applications (e.g., Alagoz et al. 2004, Shechter et al. 2008, Kreke et al. 2008). We are the first to study the robust model. For the myopic model, we present some new sufficient conditions that guarantee that a threshhold policy is optimal (Theorem 1). For the robust model we prove that therapy is initiated sooner as the level of ambiguity increases (Theorem 2).

The Myopic Model

The myopic model is a discrete-time MDP. In every period, a patient who has not yet initiated therapy faces the decision whether to initiate therapy (T) or to wait (W). The state space is $\mathcal{S} = \{1, 2, \ldots, H, H+1\}$, which is ordered. We say that lower states indicate that the patient is "healthier." State $H+1$ is an absorbing state, which is reached either by initiating therapy or by patient death. Once state $H+1$ is reached, the decision process is stopped. Otherwise, the decision process continues. This problem of deciding when to initiate therapy is in the class of *optimal stopping time* problems (c.f., Puterman 1994).

The objective is to maximize expected total discounted reward. Define $\mathcal{S}^W \equiv \mathcal{S} \setminus \{H+1\}$. The feasible actions are $\{T, W\}$ for $s \in \mathcal{S}^W$ and, arbitrarily, $\{T\}$ for state $H+1$. If the patient chooses to initiate therapy, she receives a final reward $r(s, T)$, which may depend on s. This reward $r(s, T)$ equals the expected total discounted future life expectancy of the patient post therapy initiation. If instead the patient chooses to wait, she receives an immediate reward of $r(s, W)$ plus the expected total discounted future reward. For both the myopic and robust models, we do not consider ambiguity in these rewards. For the absorbing state, $r(H+1, T) = 0$.

The myopic model (which is distinguished with the ' $\hat{}$ ' symbol) assumes that states transition

according to probability measures that are known. For an unstopped process with action W, transition probabilities are stored in an $H \times (H+1)$ transition matrix $\hat{P}$ with rows $\hat{p}_s$, $s \in \mathcal{S}^W$. When the decision is to wait, the probability of tansitioning from s in the current time period to s' in the next time period is $\hat{p}_s(s')$. The probability of death is $\hat{p}_s(H+1)$. The time horizon (the number of periods over which the decision-making process may evolve) is infinite. Of course, the life expectancy of a patient will be finite.

Let λ be the discount factor, where $0 < \lambda < 1$. Under the expected total discounted reward criterion, it is well known that there exists an optimal stationary deterministic Markovian policy (Puterman 1994, Theorem 6.2.10) characterized by a decision rule $\hat{d}^*$, where $\hat{d}^*(s)$ is the optimal action for state $s \in \mathcal{S}$. Furthermore, $\hat{d}^*$ is determined by the following recursive relationship, the *optimality equations*:

$$\hat{v}^*(s) = \max\left\{ r(s,T),\ r(s,W) + \lambda \sum_{s' \in \mathcal{S}} \hat{p}_s(s')\hat{v}^*(s') \right\},\ s \in \mathcal{S}^W,$$

$$\hat{v}^*(H+1) = 0.$$

The value function, $\hat{v}^*(s)$, equals the total expected future discounted reward of starting in state s and following an optimal policy. If $\hat{v}^*(s) = r(s,T)$, then it is optimal to initiate therapy in state s, $\hat{d}^*(s) = T$; otherwise, it is optimal to wait, $\hat{d}^*(s) = W$. The value function can be calculated using one of several well-known algorithms: value iteration, policy iteration, modified policy iteration, or linear programming (Puterman 1994).

The Robust Model

The robust model is an infinite-horizon robust MDP. Denote by d a decision rule that prescribes that the decision-maker choose action $d(s)$ in state s, and let $\mathcal{D}$ be the set of feasible decision rules. A deterministic Markovian policy is a sequence of decision rules $\{d_0, d_1, \ldots\}$, where d_t is the decision rule for period t. Under the rectangularity assumption (detailed below) the decision-maker's optimal determinist Markovian policy is also optimal amongst the class of all (allowably) history dependent and randomized control policies (Iyengar 2005, Theorem 3.1). A control policy that employs the same decision rule d in every period, $\{d, d, \ldots\}$, is a stationary deterministic policy. It is assumed that the decision-maker follows such a policy.

Nature also has a policy, for choosing transition probability measures. For general robust DP, the uncertainty sets may depend on both the state s and the action a chosen by the decision-maker. Given (s,a), nature is allowed to choose any transition measure $p \in \mathcal{P}(s,a)$. It is assumed that

nature's choices for a given state-action pair are independent of actions chosen in other states and independent of the history of previously visited states and actions. These independence assumptions are the *rectangularity assumption* (Iyengar 2005, Nilim and El Ghaoui 2005). A stationary policy of nature is a policy that chooses the same probability measure every time the same state-action pair is visited. Denote such a policy by π, and denote the set of all possible stationary policies of nature by Π. Infinite-horizon robust DP is defined by the following optimization problems:

$$w^d(s) = \inf_{\pi \in \Pi} \mathbb{E}_s^{\pi} \left[\sum_{t=0}^{\infty} \lambda^t r(s_t, d(s_t)) \right], \tag{1}$$

$$v^*(s) = \max_{d \in \mathcal{D}} w^d(s), \tag{2}$$

where s_t is the state realized in period t and $\mathbb{E}_s^{\pi}$ denotes expectation under π and initial state $s_0 = s$. This is a sequential game in which the decision-maker first selects d, in (2), and then nature responds by selecting π, in (1). Note that the decision-maker's problem (2) is solved for each initial state, and nature's problem (1) is solved for each initial state and for each feasible decision rule.

One can, without loss of generality, relax the stationarity assumptions by allowing both the decision-maker and nature to vary their decision rules over time. Such a relaxation will nonetheless result in optimal policies that are stationary (Nilim and El Ghaoui 2005, Theorem 4). Moreover, the optimal actions of the robust DP problem are characterized by a set of Bellman-type optimality equations (Iyengar 2005, Theorem 3.2).

The robust model for therapy initiation is similar to the myopic model. In fact, the myopic model is a special case of the robust model, with $\mathcal{P}(s, W) \equiv \{\hat{p}_s\}$. For sure the process does not leave the absorbing state, and $\mathcal{P}(s, T)$ consists of a single probability mass function with all mass on state $H + 1$. So, we drop the dependence on the action; $\mathcal{P}(s) \equiv \mathcal{P}(s, W)$. Given $\mathcal{P}(s)$, $s \in \mathcal{S}^W$, the robust model optimality equations are:

$$v^*(s) = \max \left\{ r(s, T),\ r(s, W) + \lambda \inf_{p \in \mathcal{P}(s)} \sum_{s' \in \mathcal{S}} p(s') v^*(s') \right\},\ s \in \mathcal{S}^W, \tag{3}$$

$$v^*(H + 1) = 0.$$

If $v^*(s) = r(s, T)$, then the optimal action is $d^*(s) = T$; otherwise, the optimal action is $d^*(s) = W$. Define

$$\sigma_{\mathcal{P}}(v) = \inf_{p \in \mathcal{P}} \sum_{s' \in \mathcal{S}} p(s') v(s'). \tag{4}$$

This is the *inner problem*. Note in (3) that $\sigma_{\mathcal{P}(s)}(v^*)$ is solved for each $s \in \mathcal{S}^W$.

As discussed in Sec. 1, we construct uncertainty sets $\mathcal{P}(s)$, $s \in S^W$, using bounds on relative entropy. These uncertainty sets are statistically meaningful, and they are also easily parameterized by the confidence level, $\omega \in (0, 1)$. They have the following form: $\mathcal{P}(s) = \{p \in \mathcal{M}(\mathcal{S}) : D(p||\hat{p}_s) \leq \beta(s)\}$, where $\hat{p}_s$ is a maximum likelihood estimate and $\beta(s)$ is a scalar that depends on both the amount of data available for estimation and ω. We do not vary ω with s.

The Kullback-Leibler distance (or relative entropy distance) between two probability mass functions p and q with sample space $\mathcal{S}$ is defined:

$$D(p||q) = \sum_{s' \in \mathcal{S}} p(s') \log \left(\frac{p(s')}{q(s')} \right). \tag{5}$$

Let $\tilde{p}_s$ denote the true (but unknown) state transition probabilities of which $\hat{p}_s$ is the maximum likelihood estimate, and let N_s be the number of historical samples used for the estimate. That is, if $n_s(s')$ is the number of sampled transitions from s to s', then $N_s = \sum_{s' \in \mathcal{S}} n_s(s')$ and $\hat{p}_s(s') = n_s(s')/N_s$. Let $F^{-1}_{|\mathcal{S}|-1}(\cdot)$ denote the inverse cumulative distribution function for a chi-squared random variable with $|\mathcal{S}| - 1$ degrees of freedom. An asymptotic approximation implies

$$\beta(s) = F^{-1}_{|\mathcal{S}|-1}(\omega)/(2N_s) \tag{6}$$

(Iyengar 2003, Ben-Tal et al. 2013); see Appendix A for details. It is important to note that the uncertainty sets are being constructed separately for each row, one for each $s \in S^W$; the rectangularity assumption is satisfied. Hence, ω is approximately equal to the probability that individual row $\hat{p}_s$ falls within $\mathcal{P}(s)$, as opposed to a collective confidence level for the entire transition matrix. Since we are leveraging an asymptotic result, it is implicitly assumed that there are enough observations to detect any nonzero probabilities. We treat any zero element of $\hat{p}_s$ as a sure zero probability, and then replace $|S|$ with the number of nonzero entries of $\hat{p}_s$.

A duality argument (Iyengar 2003) yields a reformulation of the inner problem:

$$\sigma_{\mathcal{P}(s)}(v) = -\min_{\gamma \geq 0} \left\{ \gamma\beta(s) + \gamma \log \left(\sum_{s' \in \mathcal{S}} \hat{p}_s(s') \exp\left(-\frac{v(s')}{\gamma} \right) \right) \right\}. \tag{7}$$

The function inside this minimization is one-dimensional and convex, and its derivative has a known analytical form (Nilim and El Ghaoui 2005). Therefore, it can be solved numerically using well-known techniques like bisection or Newton's method. To compute ϵ-optimal solutions, we use the Inexact RMPI algorithm under the span stopping criterion of Kaufman and Schaefer (2013).

3.1 Structural Properties

For the myopic model, Alagoz et al. (2004) derive sufficient conditions of the model inputs that guarantee the existence of an optimal decision rule characterized by a *threshold* (i.e., *control limit*): a state $\bar{s}$ such that $\hat{d}^*(s) = W$ for $s < \bar{s}$ and $\hat{d}^*(s) = T$ for $s \geq \bar{s}$. We were able to derive some new sufficient conditions:

Theorem 1 *For the myopic model, suppose that $\hat{P}$ is IFR and*

$$\left[r(s, W) + \lambda \sum_{s' \in \mathcal{S}^W} \hat{p}_s(s') r(s', T) \right] - r(s, T)$$

$$\geq \left[r(s+1, W) + \lambda \sum_{s' \in \mathcal{S}^W} \hat{p}_{s+1}(s') r(s', T) \right] - r(s+1, T), \; s \in \{1, \ldots, H-1\}. \tag{8}$$

Then, there exists an optimal decision rule $\hat{d}^*$ such that $\hat{d}^*(s) = T \Rightarrow \hat{d}^*(s+1) = T$, $s \in \mathcal{S}^W$.

The proof of Theorem 1, which is by backward induction leveraging the convergence of value iteration with a special choice of terminal reward, is presented in Appendix B. These conditions assume that the transition matrix is IFR (Barlow and Proschan 1965), defined in Appendix B: a patient in a relatively higher (less healthy) state has a higher probability of transitioning to a particular state or higher, including death. For the LDM, Alagoz et al. (2004) found that in practice the transition matrices are nearly IFR, not exact; they report a small maximum violation of the IFR assumption. Nonetheless, the optimal myopic policies calibrated to clinical data are of threshold form.

Proposition 1 in Appendix B shows that when the sufficient conditions of Alagoz et al. (2004) hold, our new conditions also hold. Both sets of conditions rely on the IFR property. It is interesting to note though that Theorem 3 of Alagoz et al. (2004) requires $r(s, T)$ to be non-increasing in s, but that is not a requirement of Theorem 1. These new conditions also have a more natural interpretation. Condition (8) is that the difference of total expected reward for waiting exactly one period and then initiating therapy minus the reward for initiating therapy immediately is non-increasing in s. In other words, the marginal benefit of waiting exactly one more period versus initiating therapy immediately decreases as the patient's health becomes worse.

For the robust model, meaningful sufficient conditions guaranteeing an optimal threshold policy are not known. While a robust version of value iteration is valid, analysis for the robust model is complicated by the fact that the transition probability measures (nature's choices) may change from one decision epoch to another. Note that the proof of Theorem 1 makes use of the fact that

the transition measures are held constant, and condition (8) incorporates transition measures that are known a priori.

The next result is more general and does not require any sufficient conditions at all. It says that as the level of ambiguity increases, therapy is initiated sooner, in more states. Moreover, states in which it is optimal to initiate therapy for the myopic model form a subset of states in which it is optimal to initiate therapy for the robust model (Corollary 2). For problem instances where both the robust and myopic models happen to be threshold policies, the robust control threshold is lower than (or equal to) the myopic threshold. That is, patients will initiate therapy in healthier states, i.e., sooner, under the robust model as compared to the myopic model.

Theorem 2 *Consider two instances of the robust model, $i = 1, 2$, with $\hat{P}$ the same for both but with potentially different ambiguity levels $\beta_i \geq 0$ and optimal decision rules d_i^*. Suppose that $\beta_1(s) \leq \beta_2(s)$ for all $s \in \mathcal{S}^W$. Then, if $d_1^*(s) = T$, there exists d_2^* such that $d_2^*(s) = T$. Moreover, if the optimal myopic decision rule satisfies $\hat{d}^*(s) = T$, then there exists d_i^* such that $d_i^*(s) = T$.*

The proof of Theorem 2 is presented in Appendix B. The following corollary follows immediately from the fact that the uncertainty sets grow as ω increases (equation (6)).

Corollary 1 *Consider two instances, $i = 1, 2$, of the robust model with confidence levels $\omega_1 \leq \omega_2$, but all other parameters the same. If $d_1^*(s) = T$, then there exists d_2^* such that $d_2^*(s) = T$.*

Corollary 2 *Consider a myopic model with the same common parameters as the robust model, with myopic transition measures $\hat{p}_s$, $s \in \mathcal{S}^W$. If for a set of states $\mathcal{Y}$ the optimal myopic decision rule $\hat{d}^*$ satisfies $\hat{d}^*(s) = T$ for $s \in \mathcal{Y}$, then there exists an optimal robust decision rule d^* satisfying $d^*(s) = T$ for all $s \in \mathcal{Y}$.*

4. Living-Donor Case Study

Organ transplantation is the only therapy for end-stage liver diseases. Currently in the United States, there are over 14,500 patients on the waiting list to receive a liver transplantion (Organ Procurement and Transplantation Network 2017). For some, a living-donor transplantation may be an option. For others, a cadaveric donor organ may be the only choice. For a living-donor transplantation, a donor who is a match according to some labratory values including blood type, donates a lobe of their liver to the recipient. As shown by Alagoz et al. (2004), the timing of such

a transplantation is an important one. On one hand, the patient is known to be ill and may need a transplantation in the future. On the other hand, they may still get additional functionality out of their own liver prior to transplantation, and it could be very beneficial to wait in order to increase their total (prior plus post tranplantation) life expectancy. The MELD score is a critical measure of current state of health, and patients should not transplant right away if their MELD is low enough. In the LDM, we are restricting attention to the living-donor transplant option. Implicitly then, we are asusming that the patients are only considering the timing of the living-donor tranplantation, and are not delaying their decision in hopes of receiving a poential future cavadeveric option.

Our data are from two sources, the details of which are found in Alagoz et al. (2004). For estimating pre-transplant transition probabilities, we use a smaller (in terms of number of patients) data set from the Thomas E. Starzl Transplantation Institute at the University of Pittsburgh Medical Center. For estimating post-transplant life expectancy, we use a large national data set from UNOS. Our study of the effects of ambiguity on transplant timing decisions is focused on ambiguity in the pre-transplant progression of disease, which relies on the smaller data set. The health state observations are the result of applying a cubic spline methodology to clinical scores of a population of patients with similar characteristics, as reported in Alagoz et al. (2005). Again, the bar chart in Figure 1(b) displays the total number of observations by health state (MELD).

Given a starting health state s, the objective is to maximize total discounted future life expectancy. We set $r(s, W) = 1$ day. The total discounted post-transplant life expectancy, $r(s, T)$, depends on several factors of the patient plus the quality of the donor organ. We consider the seven patient types presented in Alagoz et al. (2004), with patients in decreasing order with respect to their post-transplant discounted life expectancy. The patient types depend on the etiology of end-stage liver disease, the age, race and gender of the patient, and the possible presence of cytomegalovirus in the donor organ. We consider the two largest disease groups. Disease Group (DG) 1 includes primary biliary cirrhosis, primary sclerosing cholongitis, alcoholic liver disease, and autoimmune disorders. DG 2 includes hepatitis C and hepatitis B. As for the quality of the donor organ, details for 14 donor liver types are given in Alagoz et al. (2007a). We consider half of these – types 1, 3, 5, ..., 13, ordered in decreasing quality. We apply a 0.97 annual discount rate (daily = 0.999917). For the robust model, we also need to choose the confidence level ω (equation (6)). We vary ω for the study. Finally, the RMPI algorithm terminates with an ϵ-optimal solution. We set $\epsilon = 0.1$ day. Next, the resulting optimal myopic and robust policies are compared.

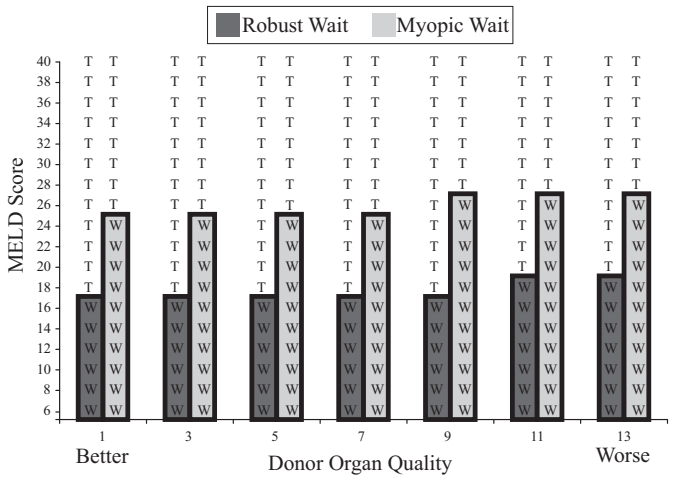


Figure 3: DG 2 – Patient 5's optimal robust policies (left) for $\omega = .95$ and optimal myopic policies (right), for seven separate problem instances – a single patient type but with seven different donor organ qualities, $\{1, 3, \ldots, 13\}$. Liver 1 is the best in the set and Liver 13 is the worst. For each problem instance, the patient only considers a single donor organ, with a known quality. These optimal policies are all threshold policies. For instance, for Liver 7 the optimal myopic threshold is 26 MELD while the optimal robust threshold is 18 MELD, a healthier (lower) state.

4.1 Optimal Policies

Figure 3 presents optimal myopic and robust transplant policies for seven separate problem instances, for a single disease group, DG 2, and single patient type, Patient 5. Each instance corresponds to a different donor organ quality, Livers $1, 3, \ldots, 13$, with Liver 1 being the best donor organ and Liver 13 being the worst donor organ in the set. We find that the resulting optimal myopic policies are indeed threshold policies. For instance, the optimal transplant decision for Liver 7 is to transplant if and only if MELD ≥ 26; the threshold is 26 MELD. In this example, as the donor quality worsens from Liver 7 to Liver 9, the myopic threshold increases to 28 MELD. (Again, Liver 7 and Liver 9 are two separate problem instances, since we assume that the patient is only considering a single living-donor with a fixed donor organ quality.) This increasing threshold trend was originally reported in Alagoz et al. (2004).

Figure 3 displays the robust policies with $\omega = .95$. Recall that this means that the confidence level for each row is 95%: for each s, $\mathbb{P}\{\tilde{p}_s \in \mathcal{P}(s)\} \approx .95$. The transition matrix has 18 rows. Hence, if we instead seek a collective 95% confidence level for the entire transition matrix, we should, under independence of rows, set $\omega = .95^{1/18} = .997154$. We start by setting $\omega = .95$, and later study the impact of varying ω.

We already know, by Corollary 2, that for those states for which transplantation is optimal under the myopic model, transplantation will also be optimal under the robust model. Moreover,

	Liver 1	Liver 3	Liver 5	Liver 7	Liver 9	Liver 11	Liver 13
Patient 1	22	22	22	26	26	26	26
Patient 2	22	22	26	26	26	26	26
Patient 3	26	26	26	26	26	26	28
Patient 4	26	26	26	26	26	28	28
Patient 5	26	26	26	26	28	28	28
Patient 6	28	28	28	30	30	30	30
Patient 7	28	28	30	30	30	30	30

(a) Optimal myopic thresholds

	Liver 1	Liver 3	Liver 5	Liver 7	Liver 9	Liver 11	Liver 13
Patient 1	18	18	18	18	18	18	18
Patient 2	18	18	18	18	18	18	18
Patient 3	18	18	18	18	18	18	20
Patient 4	18	18	18	18	18	18	20
Patient 5	18	18	18	18	18	20	20
Patient 6	20	20	20	20	20	20	20
Patient 7	20	20	20	20	20	20	22

(b) Optimal robust threshold for $\omega = .95$

Table 1: DG 2 optimal myopic and robust policies. The optimal policies are all threshold policies. For instance, for DG 2 – Patient 5 – Liver 7, the optimal myopic threshold is 26 MELD while the optimal myopic threshold is 18 MELD.

we find that the resulting robust policies are indeed threshold polices, with thresholds lower than the myopic thresholds. For instance, Patient 5 with Liver 7 has a robust threshold of 18 MELD, which is lower than the myopic threshold of 26 MELD – a difference of 8 MELD. So, the robust polices do indeed suggest transplanting sooner, in healthier states, than the myopic policies. Note that 18 MELD, the robust threshold, happens to be the lowest MELD score for which there is a historically observed patient death; see Figure 1 (a). The score 18 MELD is the robust threshold for Livers 1 – 9. For the worse donor organs, Livers 11 and 13, the threshold does increase to 20 MELD. A main take-away here is that, instead of a *single* threshold value, a lower robust threshold combined with the higher myopic threshold may instead provide a *range* of acceptable transplant MELD scores.

Table 1 displays the optimal myopic and robust thresholds for all of the DG 2 patient and donor types in our study. The average myopic threshold across all 49 problem instances is 26.78 MELD. The average robust threshold is 18.78 MELD, and the average difference is 8.00 MELD. For DG 1, it turns out that the thresholds are lower, as reported in Table 2. The average myopic threshold is 17.10 MELD, the average robust threshold is 12.78 MELD, and the average difference is 4.33 MELD. Compared to DG 2, the difference between the myopic and robust thresholds is smaller for DG 1. For both disease groups, the myopic and robust thresholds are non-decreasing in both the patient type and the liver type.

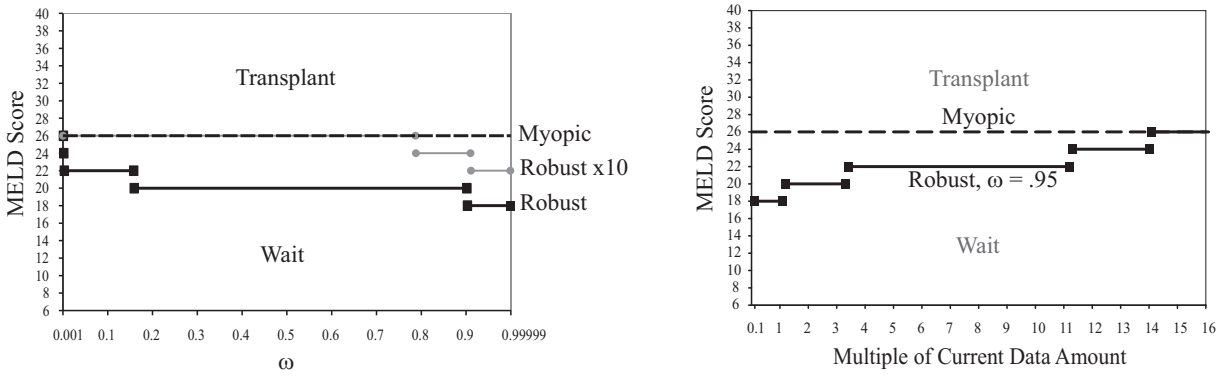
	Liver 1	Liver 3	Liver 5	Liver 7	Liver 9	Liver 11	Liver 13
Patient 1	16	16	16	16	16	16	16
Patient 2	16	16	16	16	16	16	16
Patient 3	16	16	16	16	16	16	16
Patient 4	16	16	16	16	16	16	16
Patient 5	16	16	16	16	16	16	20
Patient 6	16	20	20	20	20	20	20
Patient 7	18	20	20	20	20	20	20

(a) Optimal myopic thresholds

	Liver 1	Liver 3	Liver 5	Liver 7	Liver 9	Liver 11	Liver 13
Patient 1	12	12	12	12	12	12	12
Patient 2	12	12	12	12	12	12	12
Patient 3	12	12	12	12	12	12	12
Patient 4	12	12	12	12	12	12	14
Patient 5	12	12	12	12	12	12	14
Patient 6	14	14	14	14	14	14	16
Patient 7	14	14	14	14	14	16	16

(b) Optimal robust threshold for $\omega = .95$

Table 2: DG 1 optimal myopic and robust policies. The optimal policies are all threshold policies.



(a) Optimal thresholds vs. ω

(b) Optimal thresholds vs. data multiple

Figure 4: (a) DG 2 – Patient 5 – Liver 7 optimal thresholds as a function of ω. "Robust x10" is the optimal robust threshold assuming that transition probability estimates were instead made with 10 times the amount of data (data multiple = 10). (b) The optimal thresholds as a function of a the data multiple; $\omega = .95$.

In Figure 4(a), ω is varied, for DG 2 – Patient 5 – Liver 7. We will explore this particular DG-Patient-Liver combination in greater detail. We know from Corollary 1 that the robust threshold is non-increasing in ω. Here, ω is varied from .001 to .99999. For $\omega = .001$, the robust threshold matches the myopic threshold at 26 MELD. For $\omega = .002$, the robust threshold drops to 24 MELD, and for $\omega = .003$ it further drops to 22 MELD. Hence, for the lowest values of ω, the robust threshold is very sensitive to ω. On the other side, the robust threshold is 18 MELD whether we set $\omega = .95$ or $\omega = .99999$, so the threshold is less sensitive to higher ω.

These results are for the current amount of information, per the histogram in Figure 1(b). We may also consider a counterfactual case that the health state transition matrix had instead

been estimated with more data. For example, for 10 times the amount of data we simply substitute $N_s := 10N_s$ in equation (6), for all s. We refer to this as a *data multiple* of 10. For this hypothetical case, in Figure 4(a), the robust threshold matches the myopic threshold for all $\omega \in (0, .787]$. For higher ω, the threshold drops, but only drops to 22 MELD for $\omega = 0.99999$. Such a sensitivity analysis indicates the potential benefit of investing in a larger, though costly, medical study. Figure 4(b) takes this analysis a step further. For fixed $\omega = .95$, the data multiple is varied. For 1/10th the amount of data, the robust threshold would still be 18 MELD. For a data multiple of 3.4, the robust threshold is 22 MELD. Recall from Figure 4(a) that, for data multiple = 1 (the actual amount of data), 22 MELD is the robust threshold for ω as small as .002. Again from Figure 4(b), for $\omega = .95$, it would take a lot more data, a data multiple of 11.3, to increase the robust threshold above 22 MELD to 24 MELD. It takes a data multiple ≥ 14 for the robust threshold to finally match the myopic threshold. So far we have just explored the form of the optimal policies.

4.2 Performance Evaluation

The question remains open: which perform better, the myopic or the robust solutions? We assume that the ultimate objective is to maximize total discounted life expectancy; but, in practice the true underlying transition probabilities are not known. Rather, we only have the maximum likelihood estimate $\hat{P}$ and ω-confidence uncertainty sets. In order to calculate the true life expectancy, this is not enough.

One approach to evaluating performance is to perform a what-if sensitivity analysis by perturbing $\hat{P}$. We measure the total discounted future life expectancy according to a transition matrix that is a perturbed version of this maximum likelihood estimate, with each row $\hat{p}$ perturbed by shifting some amount $\alpha > 0$ of probability mass to the worst state in that row that has nonzero estimated probability mass. For instance, consider the row for transitions from 14 MELD. Note from Figure 1(a) that the largest historically observed increase is from 14 MELD to 24 MELD. So, 24 MELD is the worst transition state for starting state 14 MELD. In the perturbed matrix, there is α probability more in the element representing a transition to 24 MELD, with the other probabilities in the row decreased proportionally according to the $\hat{p}$. Similarly, the rows for 18 MELD and greater are all perturbed to have α probability more for the transition to death since, for these starting states, death is the worst historical transition state.

Figure 5 demonstrates sensitivity to the perturbation amount α, with α equal for each row. The optimal policy varies with α. We see that the resulting optimal discounted future life expectancy,

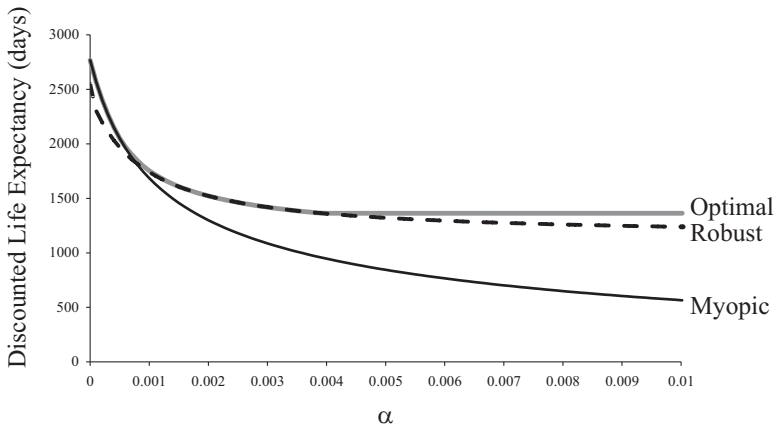


Figure 5: Performance evaluated according to a perturbed matrix. Total discounted life expectancy for DG 2 – Patient 5 – Liver 7 for starting state 6 MELD under the true optimal policy (as measured by the perturbed matrix), the robust policy ($\omega = .95$), and the myopic policy, as a function of the perturbation quantity α. The myopic policy coincides with the optimal policy for $\alpha \in [0, .0005]$. The robust policy coincides with the optimal policy for $\alpha \in [.0015, .0039]$. The robust policy outperforms the myopic policy for $\alpha \geq .0008$.

as measured by the perturbed matrix, is decreasing in α. Figure 5 also shows the performance of the myopic and robust policies. Here, the myopic policy is fixed. The myopic policy is determined using the original maximum likelihood estimate, $\hat{P}$ (unperturbed). The robust policy is determined using uncertainty sets centered around $\hat{P}$. So, the myopic and robust policies do not vary with α. For the figure, ω is held constant at $\omega = .95$. Per Sec. 4.1, the myopic threshold is 26 MELD and the robust threshold is 18 MELD.

While the myopic and robust policies are fixed, performance is measured according to the perturbed transition matrices that are constructed as α varies. We see that that performances of the myopic and robust policies, measured by discounted total life expectancy under the perturbed matrix, are also decreasing in α. When α is very small, below .0008, the myopic performance is closer to the optimal performance, but the robust performance is still close. But for $\alpha > .0008$, the robust performance closely follows the optimal performance, while the myopic performance drops off more dramatically. The plot shows performance for $\alpha \leq .01$. At $\alpha = .01$, the myopic performance is considerably lower than the optimal performance. On the other hand, the robust performance remains very close to optimal. Overall, we find that the discounted life expectancy may be very sensitive to small perturbations in the transition matrix.

Based on this sensitivity analysis, since the robust discounted life expectancy is close to optimal while the myopic performance drops off dramatically as α increases, one may view the robust performance as better overall than the myopic performance. However, we need to be careful in

drawing conclusions; the analysis is limited. We have used the same α for every row; but, as seen in Figure 1(b), the amount of information available varies by row – precisely the reason that, while ω is held constant, robust uncertainty sets are larger when less data are available for estimation. This sensitivity analysis is not performed in a statistically meaningful way. Next, we consider an alternative way to compare the performance of the policies, with a simulation experiment.

4.3 Simulation Experiment

We now present the results of a simulation study that assumes that the true underlying transition matrix is somehow known (to, say, an omniscient oracle). For each replication of the experiment, we compute both myopic and robust solutions using sampled data, and then compare performance using the total discounted life expectancy criterion under the true transition probabilities. We assume that the true underlying transition matrix is indeed the matrix $\hat{P}$ illustrated in Figure 1(a), which is the best estimate available from the clinical data. This matrix is randomly sampled to generate new data, using the inverse transform method (Law and Kelton 1991) to sample the rows' multinomial distributions. As the base case, for each replication, the total number of samples for each row equals the total number of historical observations as presented in Figure 1(b). For each sampled matrix, one per replication, the optimal myopic and robust policies corresponding to the sampled matrix are determined, without direct knowledge of the true underlying matrix. Since the generating process is known (to the oracle) though, the true discounted life expectancies for the resulting myopic and robust policies are easily determined. Also, the true optimal policy is determined, with its optimal total discounted life expectancy being an upper bound for the myopic and robust life expectancies. Note: (1) The simulated transition probability matrices do not depend on constructed uncertainty sets. Hence, the myopic performance is independent of any uncertainty set construction. (2) As is reality, there is no guarantee that true nonzero probability transitions, especially low probability events like patient death in relatively healthy states, will be sampled.

Table 3(a) displays the true optimal total discounted life expectancy for 49 problem instances – the seven Patients combined with the seven Livers for DG 2. The initial health state here is the lowest score, 6 MELD. The corresponding optimal policies (for any starting health state) are characterized by the myopic thresholds previously presented in Table 1(a). Table 3(b) provides the percent of life expectancy loss if instead of taking the optimal action a patient would transplant right away, at 6 MELD. For instance, Patient 5 – Liver 7 has an optimal threshold of 26 MELD and an optimal total discounted life expectancy of 2,767.9 discounted days. If the patient instead

	Liver 1	Liver 3	Liver 5	Liver 7	Liver 9	Liver 11	Liver 13
Patient 1	3520.5	3388.7	3267.7	3187.5	3130.4	3055.8	2947.9
Patient 2	3392.9	3269.1	3171.6	3105.1	3053.3	2985.3	2887.0
Patient 3	3096.0	3024.2	2956.5	2905.6	2865.6	2812.4	2740.3
Patient 4	3004.2	2940.1	2880.1	2834.0	2797.3	2749.7	2687.9
Patient 5	2922.8	2865.3	2810.2	2767.9	2738.9	2697.4	2636.8
Patient 6	2652.9	2610.5	2570.9	2544.3	2521.1	2490.3	2442.8
Patient 7	2617.2	2575.6	2540.3	2515.0	2491.7	2462.3	2416.1

(a) Optimal myopic discounted life expectancy (days) for starting state 6 MELD

	Liver 1	Liver 3	Liver 5	Liver 7	Liver 9	Liver 11	Liver 13
Patient 1	77.83%	74.34%	72.19%	69.54%	67.20%	63.51%	57.40%
Patient 2	75.13%	71.42%	68.81%	65.99%	63.64%	60.09%	54.33%
Patient 3	65.81%	61.47%	58.77%	56.25%	54.25%	51.32%	46.66%
Patient 4	61.46%	57.37%	54.85%	52.58%	50.78%	48.12%	43.79%
Patient 5	57.33%	53.62%	51.33%	49.28%	47.56%	45.14%	41.11%
Patient 6	42.91%	40.40%	38.68%	36.95%	35.50%	33.14%	28.84%
Patient 7	41.07%	38.57%	36.79%	34.99%	33.53%	31.09%	26.66%

(b) Bound: % below optimal life expectancy when transplanting immediately, at 6 MELD

Table 3: DG 2 (a) optimal myopic discounted life expectancies and (b) performance bounds. For instance, the optimal myopic discounted life expectancy for Patient 5 – Liver 7 is 2,767.9 discounted days, following the optimal threshold policy with threshold 26 MELD (Table 1(a)). However, the discounted life expectancy is 49.28% lower (only $2,767.9(1 - .4928) = 1,404$ discounted days) if instead the transplantation occurs immediately, at 6 MELD.

chose to transplant immediately, when at 6 MELD, the life expectancy would drop by 49.28% to 1,403.9 discounted days, which would be a considerable loss. Over all 49 instances for DG 2, the range of loss for the suboptimal action of transplanting immediately is 26.66% – 77.83%, which is substantial. The results for DG 1 are presented in Table C.1, where the range of loss for transplanting immediately is 9.74% – 41.84%. A main take-away here is that it is important to not transplant too soon.

The simulation experiment consisted of 10,000 replications. Table 4 displays the resulting average (over all replications) myopic and robust threshold levels, for the 49 problem instances for DG 2. For instance, Patient 5 – Liver 7 has an average myopic threshold of 27.15 MELD, which happens to be above the true optimal threshold of 26 MELD (Table 1(a)), by amount 1.15 MELD. Since the transition matrices are sampled, there is no guarantee that each of the 10,000 myopic policies is of threshold form. For example, for Patient 5 – Liver 7, of the 10,000 replications, 133 resulted in myopic policies that are not of threshold form. The reported "threshold" for such cases is the lowest state for which transplant is the chosen action.

For the robust model, we again start by setting $\omega = .95$. The average robust threshold is 19.01 MELD, which is lower than the true optimal threshold of 26 MELD, by 6.99 MELD. Performance is

	Liver 1	Liver 3	Liver 5	Liver 7	Liver 9	Liver 11	Liver 13
Patient 1	21.93	22.73	23.56	24.01	24.47	25.26	26.17
Patient 2	22.58	23.52	24.30	24.74	25.18	25.79	26.56
Patient 3	24.78	25.49	26.00	26.29	26.56	26.88	27.52
Patient 4	25.57	26.07	26.42	26.73	26.99	27.30	27.94
Patient 5	26.19	26.49	26.82	27.15	27.40	27.75	28.36
Patient 6	28.14	28.47	28.80	29.17	29.38	29.82	30.37
Patient 7	28.47	28.82	29.15	29.51	29.74	30.15	30.83

(a) Average myopic thresholds (MELD)

	Liver 1	Liver 3	Liver 5	Liver 7	Liver 9	Liver 11	Liver 13
Patient 1	18.02	18.03	18.03	18.04	18.06	18.10	18.45
Patient 2	18.02	18.04	18.04	18.06	18.12	18.20	18.66
Patient 3	18.07	18.19	18.25	18.38	18.59	18.79	19.42
Patient 4	18.17	18.38	18.48	18.67	18.92	19.15	19.67
Patient 5	18.34	18.68	18.81	19.01	19.26	19.47	19.87
Patient 6	19.75	19.99	20.14	20.36	20.55	20.87	21.26
Patient 7	19.92	20.18	20.37	20.60	20.84	21.22	21.92

(b) Average robust thresholds (MELD) for $\omega = .95$

Table 4: DG 2 average robust and myopic thresholds, averaged over 10,000 replications. For instance, for Patient 5 – Liver 7 the true optimal threshold is 26 MELD (Table 1(a)), the average myopic threshold is 27.15 MELD, and the average robust threshold is 19.01 MELD.

measured by total discounted life expectancy, as measured by the true underlying transition matrix. Table 5 presents the average (over all replications) percentage drop in life expectancy as compared to optimal for (a) the myopic policies and (b) the robust policies when $\omega = .95$. For instance, Patient 5 – Liver 7 has an average loss of only 0.12% following the myopic policies. For the robust policies, the average loss for Patient 5 – Liver 7 is 9.93%, which is a significantly bigger average loss, yet still well below the bound of 49.28% (Table 3(b)) for transplanting immediately at MELD 6. The average loss ranges over all problem instances are 0.15% – 0.47% for myopic and 1.98% – 13.33% for robust. On average, the myopic approach outperforms the robust approach. Though the myopic approach outperforms on average, it does not outperform in every replication. Table 6 gives (a) the fraction of replications for which robust strictly outperforms (i.e., beats) myopic, and (b) the fraction of replications that robust either beats or ties myopic. For DG 2, robust either beats or ties myopic in 0% – 2.66% of the replications.

The results for DG 1 are presented in Tables 2(a), C.2, C.3, and C.4 (which are analogous to the DG 2 Tables 1(a), 4 ,5, and 6, respectively). Table C.2 presents the average myopic and robust thresholds for 10,000 replications. Again, for some problem instances the average myopic threshold (Table C.2(a)) is above the true optimal threshold (Table C.2(a)), yet for some problem instances (e.g., Patient 1 Liver 1) the average myopic threshold is below the true optimal threshold. The average robust thresholds (Table C.2(b)) are lower. As compared to DG 2, the DG 1 ranges between

	Liver 1	Liver 3	Liver 5	Liver 7	Liver 9	Liver 11	Liver 13
Patient 1	0.45%	0.45%	0.26%	0.20%	0.35%	0.43%	0.36%
Patient 2	0.47%	0.27%	0.30%	0.39%	0.44%	0.40%	0.27%
Patient 3	0.40%	0.41%	0.34%	0.33%	0.28%	0.21%	0.23%
Patient 4	0.44%	0.34%	0.28%	0.24%	0.16%	0.15%	0.27%
Patient 5	0.35%	0.28%	0.22%	0.12%	0.20%	0.26%	0.24%
Patient 6	0.27%	0.23%	0.15%	0.17%	0.19%	0.21%	0.19%
Patient 7	0.23%	0.15%	0.18%	0.21%	0.22%	0.20%	0.15%

(a) Average myopic difference in life expectancy below optimal (%)

	Liver 1	Liver 3	Liver 5	Liver 7	Liver 9	Liver 11	Liver 13
Patient 1	1.98%	2.75%	3.26%	3.93%	4.90%	6.27%	7.66%
Patient 2	2.63%	3.43%	4.43%	5.31%	6.21%	7.45%	8.44%
Patient 3	5.47%	6.91%	8.02%	8.59%	8.96%	9.65%	9.86%
Patient 4	7.01%	8.15%	9.07%	9.42%	9.58%	10.12%	10.53%
Patient 5	8.31%	9.02%	9.74%	9.93%	10.15%	10.70%	11.20%
Patient 6	11.06%	11.37%	11.89%	12.13%	12.31%	12.59%	13.33%
Patient 7	11.43%	11.68%	12.21%	12.43%	12.45%	12.58%	12.55%

(b) Average robust difference in life expectancy below optimal (%) for $\omega = .95$

Table 5: DG 2 average differences in life expectancy below optimal (%). For instance, for Patient 5 – Liver 7, with optimal discounted life expectancy of 2,767.9 discounted days (Table 3), the myopic solution is only 0.12% below optimal (2,764.4 discounted days) on average, while the robust solution is 9.93% below optimal (2,493.0 discounted days) on average.

	Liver 1	Liver 3	Liver 5	Liver 7	Liver 9	Liver 11	Liver 13
Patient 1	0.99%	1.02%	0.11%	0.21%	0.02%	0.09%	0.01%
Patient 2	1.14%	0.13%	0.21%	0.01%	0.05%		
Patient 3	0.02%			0.01%			
Patient 4		0.01%					
Patient 5	0.01%						
Patient 6							
Patient 7							

(a) Robust beats myopic fraction of replications

	Liver 1	Liver 3	Liver 5	Liver 7	Liver 9	Liver 11	Liver 13
Patient 1	2.66%	1.26%	0.22%	0.27%	0.04%	0.10%	0.02%
Patient 2	1.64%	0.21%	0.25%	0.02%	0.06%		
Patient 3	0.03%			0.01%			
Patient 4		0.01%					
Patient 5	0.01%						
Patient 6							
Patient 7							

(b) Robust beats or ties fraction of replications

Table 6: DG 2 fraction of replications in which the robust policy with $\omega = .95$ (a) strictly outperforms (beats) and (b) performs at least as well as (beats or ties) the myopic policy. Blank cells are zeros.

the myopic and robust thresholds are tighter. The average loss ranges over all problem instances are 0.06% – 0.20% for myopic (Table C.3(a)) and 1.64% – 4.55% for robust (Table C.3(b)). Robust either beats or ties myopic in 0.08% – 10.97% of the replications (Table 6). The main take-away is that while the robust approach (with $\omega = .95$ anyway) does sometimes outperform, it does not outperform very often, and when it does outperform, it does not outperform by much.

ω	.005	.05	.5	.95	.995
True optimal threshold (MELD)	26	26	26	26	26
Avg. myopic threshold (MELD)	27.17	27.17	27.17	27.17	27.17
Avg. robust threshold (MELD)	24.03	22.39	20.39	19.02	18.38
Myopic avg. % below optimal life expec.	0.13%	0.13%	0.13%	0.13%	0.13%
Robust avg. % below optimal life expec.	2.06%	3.95%	7.16%	9.91%	11.23%
Robust beats myopic fraction	31.32%	9.16%	0.06%		
Beats or ties fraction	42.72%	10.56%	0.06%		

(a) DG 2 – Patient 5 – Liver 7

ω	.005	.05	.5	.95	.995
True optimal threshold (MELD)	26.78	26.78	26.78	26.78	26.78
Avg. myopic threshold (MELD)	26.82	26.82	26.82	26.82	26.82
Avg. robust threshold (MELD)	23.51	21.77	20.41	19.11	18.66
Myopic avg. % below optimal life expec.	0.27%	0.27%	0.27%	0.27%	0.27%
Robust avg. % below optimal life expec.	1.89%	3.30%	6.15%	8.75%	9.78%
Robust beats myopic fraction	11.18%	4.68%	0.50%	0.41%	0.30%
Beats or ties fraction	18.02%	5.86%	0.56%	0.53%	0.41%

(b) Averages taken over all 49 DG 2 problem instances

Table 7: Simulation experiment results for various confidence levels (ω), for (a) Patient 5 – Liver 7 and (b) averages over the 49 problem instances. Blank cells are zeros. In (a), for Patient 5 – Liver 7, note that the average (over replications) myopic threshold is 27.17 MELD, which is very close to the reported average myopic threshold of 27.15 MELD in Table 4(a). The difference is due to simulation error; the results in this table used a different simulation run. In (b), the average (over replications) myopic thresholds averaged over the problem instances are close to the true optimal thresholds averaged over the problem instances. However, Table 4(a) shows that there is variance over these problem instances, as the average myopic thresholds are sometimes above and sometimes below the true optimal threshold. Expanded results for all 49 problem instances are reported in Appendix C.

Sensitivity to Confidence Level

Next, we vary ω. We restrict attention to the 49 instances of DG 2. The main table for this section is Table 7. Table 7(a) presents results for Patient 5 – Liver 7, for 5,000 replications, and Table 7(b) presents summary results averaged over the 49 problem instances. Tables C.5, C.6, C.7, and C.8 display performance comparisons for $w = .005$, .05, .5, and .995, respectively, which are the results summarized in Table 7(b). Again, Tables 4, 5, and 6 display results for $\omega = .95$.

As seen in Table 7, for both the myopic and robust solutions, the average loss from the true optimal life expectancy is decreasing in ω. On average, myopic outperforms robust for all ω settings. For Patient 5 – Liver 7, the robust average loss for $\omega = .995$ is 11.23%, up from 9.91% for $\omega = .95$. Still, the average loss is below the bound of 49.28% (Table 3(b)). At the other end, for $\omega = .005$ the robust average loss drops to 2.06%. Recall from Figure 4(a) that robust thresholds may be very sensitive to changes in ω when ω is very small. Note in Table 7(a) that robust beats or ties myopic in 42.72% of the replications. Nonetheless, the myopic solutions perform very well. On average (over replications), the myopic average loss from optimal is only 0.13%. The main take-away is

that the myopic policies perform very well, and robust policies perform well when ω is relatively low. When robust policies outperform myopic policies, the difference is small. In general, setting $\omega = .95$ is conservative and performance can be significantly degraded for $\omega > .95$.

Sensitivity to Data Quantity

While the myopic solutions outperform the robust solutions, on average, under the current amount of information available, we want to explore the impacts of varying the amount of ambiguity. For example, what if instead of the current amount of data, there was significantly less data available – more ambiguity? Tables 8 and 9 explore the impact of varying the data multiple, for DG 2 – Patient 5 – Liver 7. Recall from Sec. 4.1 that a data multiple is a scaling of N_s in (6). The lower the data multiple, the more the amount of ambiguity. Recall from Figure 4(b) that the robust thresholds approach the myopic thresholds as the data multiple increases. From Table 8, we find that the myopic average loss is 1.26% for one-fifth the amount of data, and only .02% for five times the amount of data. At the same time, in Table 9, the robust average loss when $\omega = .95$ is 10.25% for one-fifth the amount of data, and 4.26% for five times the amount of data. As ω is lowered, the robust solutions perform better, but still not as well as the myopic solutions, on average.

The fraction of 5,000 replications in which robust beats or ties myopic is also displayed in Table 9, again for DG 2 – Patient 5 – Liver 7. This fraction increases as ω is lowered and the uncertainty sets shrink. When $\omega = .95$ robust does not ever beat myopic for the current amount of data, data multiple = 1. However, for data multiple = 1/5 (more ambiguity), robust does strictly beat myopic in 1.74% of the replications. Towards the other end of the spectrum, when $\omega = .05$, and the myopic and robust thresholds are closer together, robust beats myopic in 12.94% of replications for data multiple = 1/5, and robust beats myopic in 9.16% of the replications for data multiple = 1. For even lower ω, $\omega = .005$, robust beats myopic in an even higher fraction.

As the data multiple is increased (ambiguity is decreased), and the robust thresholds get closer to the myopic thresholds, the overall tendency is that the robust average loss decreases. However, the trend is not strict. For $\omega = .95$ in particular, we see that the robust average loss in life expectancy actually increases some as the data multiple increases from 1/5 to 1/2, while the average threshold actually decreases slightly from 18.87 MELD to 18.35 MELD. The fact that the average robust threshold decreases as the amount of ambiguity increases may at first blush seem surprising, especially in light of Theorem 2 which says that the robust threshold does not decrease as the level of ambiguity increases. Note however that Theorem 2 applies to fixed maximum likelihood

Data multiple	Avg. myopic threshold (MELD)	Myopic avg. % below optimal life expec.
0.2	27.49	1.26%
0.5	27.30	0.36%
1	27.17	0.13%
2	27.05	0.05%
5	26.94	0.02%

Table 8: Myopic results for DG 2 – Patient 5 – Liver 7 for various data multiples ($\{.2, .5, 1, 2, 5\}$).

ω	.005	.05	.5	.95	.995
Avg. myopic threshold (MELD)	23.20	21.30	19.32	18.87	18.39
Robust avg. % below optimal life expec.	3.78%	6.17%	9.46%	10.25%	11.24%
Robust beats myopic fraction	25.26%	12.94%	3.66%	1.74%	1.16%
Beats or ties fraction	36.48%	15.68%	3.94%	1.78%	1.16%

(a) Data multiple = .2

ω	.005	.05	.5	.95	.995
Avg. myopic threshold (MELD)	23.34	21.63	19.52	18.35	18.28
Robust avg. % below optimal life expec.	3.02%	5.23%	8.93%	11.28%	11.43%
Robust beats myopic fraction	24.58%	7.32%	0.18%	0.04%	0.04%
Beats or ties fraction	31.18%	8.18%	0.20%	0.04%	0.04%

(b) Data multiple = .5

ω	.005	.05	.5	.95	.995
Avg. myopic threshold (MELD)	24.03	22.39	20.39	19.02	18.38
Robust avg. % below optimal life expec.	2.06%	3.95%	7.16%	9.91%	11.23%
Robust beats myopic fraction	31.32%	9.16%	0.06%		
Beats or ties fraction	42.72%	10.56%	0.06%		

(c) Data multiple = 1

ω	.005	.05	.5	.95	.995
Avg. myopic threshold (MELD)	25.18	23.59	21.52	20.13	19.67
Robust avg. % below optimal life expec.	0.84%	2.46%	5.05%	7.63%	8.57%
Robust beats myopic fraction	44.70%	23.08%	0.20%		
Beats or ties fraction	72.90%	29.20%	0.20%		

(d) Data multiple = 2

ω	.005	.05	.5	.95	.995
Avg. myopic threshold (MELD)	25.99	25.80	23.12	21.92	21.28
Robust avg. % below optimal life expec.	0.01%	0.20%	2.93%	4.26%	5.46%
Robust beats myopic fraction	46.90%	46.70%	12.94%		
Beats or ties fraction	99.54%	92.10%	15.52%		

(e) Data multiple = 5

Table 9: Robust results for DG 2 – Patient 5 – Liver 7 for various confidences levels ($\{.005, .05, .5, .95, .995\}$) and data multiples ($\{.2, .5, 1, 2, 5\}$).

estimates, and in the simulation experiment the maximum likelihood estimates instead vary for each replication. In particular, while the true underlying transition probabilities remain fixed, it may be the case, especially for low amounts of data, that some true nonzero transition probabilities that are very close to zero do not actually yield a positive number of samples.

4.4 Implied Confidence Level in Practice

In practice, transplantations are often occurring sooner than what the optimal myopic policies suggest. This was one of the motivating factors for our study of the impact of ambiguity on decisions. In this section, we consider a cohort of 693 patients who underwent a living-donor liver transplantation. We compare their decisions to the myopic and robust policies.

We want to study the level of conservativeness that may exist in practice. For our study, the cohort consists of DG 1 and DG 2 patients, though they are not restricted to the 49 problem instances considered above. Their optimal myopic policies may be quite different, similar to Tables 1 and 2. Therefore, the difference in MELD between the actual transplant MELD score and the myopic threshold might not be a consistent measure of conservativeness across DG-Patient-Liver characteristics. Instead, we propose using the confidence level ω in the robust DP model.

The *implied confidence level* is a value of ω for which the historical transplant MELD score matches the robust threshold. To calculate it, we start with the optimal myopic threshold. If the historical MELD matches the myopic threshold, then the implied confidence level equals zero. Otherwise, if the historical MELD is lower, robust thresholds are calculated, starting with ω close to 0. Then, ω is increased until the historical MELD matches the robust threshold. That is, the implied confidence level is the smallest value of ω such that the robust policy is to transplant at the historical MELD score and to wait at all lower MELD scores. For example, suppose that the characteristics are those of DG 2 – Patient 5 – Liver 7, as in Figure 4(a), and that the historical transplant MELD score is 20 MELD. Given the amount of information available (data multiple = 1), the robust threshold is 20 MELD for $\omega \in [.159, .902]$, which is a large range of ω for the same transplant action. We report that the implied confidence level as the lower bound $\omega = .159$. It is important to note that we do not have evidence that patients in practice actually use MDP models to guide their decisions.

The data is from UNOS, and these patient may have also been waiting for a cadaveric organ. In the end though, the patients in our cohort underwent a living-donor transplantation. We do not have data on whether these patients were also giving serious consideration to waiting for a cadaveric option. It is possible that waiting for a cadaveric organ delayed, but in any case should not sped up, the timing of the living-donor transplantation that was ultimately undertaken (c.f., Alagoz et al. 2007a).

Table 10 presents the implied confidence levels for our cohort. The myopic threshold was

Above Myopic Threshold	16.2%
Myopic Threshold	3.6%
Robust ω Range	Frequency
(0, .001]	0%
(.001, .005]	11.4%
(.005, .01]	0.4%
(.01, .05]	1.2%
(.05, .1]	0.3%
(.1, .3]	0.7%
(.3, .5]	0.3%
(.5, .7]	3.9%
(.7, .9]	5.1%
(.9, .95]	1.7%
(.95, .99]	0.7%
(.99, .995]	0.1%
(.995, .999]	0.9%
(.999, 1]	53.5%
Subtotal:	80.2%

Table 10: Implied ω in practice, for a cohort of 693 patients who underwent a living-donor liver transplantation. Note that the bins for implied ω are not of equal length, with more refinement near the extremes of 0 and 1. For the considered population of patients who underwent a living-donor transplant, 3.6% transplanted at the same MELD score as the myopic threshold, 16.2% transplanted at a MELD score even higher than myopic threshold, and 80.2% transplanted at a MELD score lower than the myopic threshold. A range of implied $\omega \in (0, .999]$ corresponds to 26.7% of the population, while 53.5% of the population transplanted at a MELD score even lower than the robust threshold for $\omega = .999$.

matched in 3.6% of the cases. In 16.2% of the cases, the living-donor transplantation occurred at a MELD score greater than the myopic threshold. Of course, it is possible that some of these patients were waiting for a cadaveric organ. In 80.2% of the cases – a large majority – transplantations were undertaken at MELD scores below those suggested by the myopic policies. The question then is, how conservative are these decisions?

In terms of conservativeness, the decisions are not all equal. Table 10 bins the implied confidence levels of our cohort. The bins are not of equal length; there is more refinement near the extremes of 0 and 1. Recall that the robust thresholds may be very sensitive to ω when ω is low (e.g., Figure 4(a)), and our simulation study suggests that the robust solutions more frequently beat the myopic solutions when ω is low. From Table 10, 13.0% of the cases have an implied confidence level in $(0, 0.05]$, which is not too conservative. As ω increases, the decisions become more conservative. Note that 25.0% of the cases have an implied confidence level in $(0, .95]$; 26.7% have an implied confidence level in $(0, 0.999]$, where $\omega = 0.999$ is larger than a collective 95% confidence level for the entire transition matrix, $\omega = .997154$. An implied confidence level of 0.999 is the most conservative level that we considered. Per our simulation study, such high confidence levels may

be too conservative as they result in significantly degraded performance. It is remarkable then that 53.5% of the cases, with an implied confidence level > 0.999, transplanted even sooner than what the most conservative robust policies suggest. While ambiguity could account for a portion of conservativeness in practice, it may indeed be the case that a majority of transplantations are occurring too soon.

5. Conclusion

Robust DP, which goes back to the 1970's, has been proposed as a methodology for mitigating the effects of ambiguity in MDP solutions. More recent advances that propose constructing uncertainty sets through bounds on relative entropy are both statistically meaningful and computationally tractable. As such, robust DP is a promising framework. Despite this, there is a paucity of real-world studies on the merits of robust DP.

Medical decision-making in particular is an area ripe for study. Unlike other traditional application areas of MDPs, such as highly automated manufacturing systems where data may be readily available and pilot studies may be performed with little risk to people, models for therapy initiation rely on clinical data that might not be abundant. Even when the total amount of data overall is large, information may be limited for some critical states. Informed censoring occurs due to early therapy initiation or patient deaths, and decisions have to be made in the face of ambiguity.

For living-donor transplantation timing, a transplant needs to occur before a patient becomes too ill and the probability of death becomes too big. The most critical health states though are the ones with more ambiguity (see Figure 1). Since end-stage liver diseases exhibit acute exacerbations and recoveries, optimal timing is a challenging decision problem. If one transplants too soon, the loss in total life expectancy (pre-transplant plus post-transplant) may be significant (see Tables 3 and C.1). In practice, we find that many transplants are occurring sooner than what the myopic policies suggest. A natural question is whether ambiguity can account for some of the conservativeness.

For our robust model for therapy initiation, we proved the very general result that therapy is initiated sooner for larger uncertainty sets (Theorem 2), either due to more ambiguity or a higher confidence level. If the myopic and robust thresholds happen to match, then one may be confident in the policy despite the presence of ambiguity. For the LDM though, we find that the robust thresholds are sensitive to ambiguity even at confidence levels close to zero (see Figure 4(a)). Moreover, we find that discounted life expectancies may be very sensitive to small perturbations in

the transition probability matrix (see Figure 5). Due to these sensitivities, there are gaps between the myopic and robust transplant thresholds (see Figures 3 and 4). Instead of a single solution then, these thresholds may provide a range of potentially acceptable transplant MELD scores, ruling out overly conservative actions. This information should be valuable to clinicians.

As for comparing the performance of the myopic and robust policies, in our study of a perturbed transition matrix we found that the robust policy remains close to optimal, while the performance of the myopic policy drops off dramatically (see Figure 5). However, in our simulation experiment, which is better grounded statistically, we found that the myopic policies perform very well, and may actually be considered more "robust" than the robust solutions for the objective of maximizing life expectancy. Still, while robust solutions are more conservative, their performance is significantly better than the performance bounds for transplanting immediately at the lowest MELD score.

Our implied confidence level measure maps historical transplantation timing decisions to robust thresholds. For the cohort in our study of historical living-donor transplantations, we found that the robust policies are closer to transplant decisions that have been made in 26.7% of the cases, over a range of confidence levels. We do not, however, have evidence that ambiguity actually "explains" conservative decisions in practice. We can just say that the robust solutions match decisions made in practice for some cases. On the flip side, there is the remarkable finding that in 53.5% of the cases transplants have occurred even sooner that what is suggested by the most conservative robust solutions. It may be that many transplantations are occurring too soon.

Finally, let us conclude with managerial implications for the LDM. First, the myopic solutions originally proposed by Alagoz et al. (2004) perform quite well. However, if one is even a little bit sensitive to ambiguity, one could consider the robust solutions with $\omega = .05$. Of course, such a low confidence level may not be a traditional choice from a statistical standpoint. Still, the robust thresholds for $\omega = .05$ may already reflect the sensitivity to small perturbations in unknown transition probabilities, with good performance. Interestingly, 13.0% of the historical cases in our study have an implied confidence level in $(0, .05]$. If one is more conservative, then robust solutions with $\omega = .95$ or even $\omega = .95^{1/18} = .997154$ may be warranted. Again, the myopic thresholds together with the robust thresholds provide a range of potentially acceptable transplant MELD scores, ruling out overly conservative actions.

Acknowledgments

This research was partially supported by National Science Foundation Grant DMI-0546960. The data reported here have been supplied by the United Network for Organ Sharing as the contractor for the Organ Procurement and Transplantation Network. The interpretation and reporting of these data are the responsibility of the authors and in no way should be seen as an official policy of or interpretation by the OPTN or the U.S. Government.

References

Ahn, J. H., J. Hornberger. 1996. Involving patients in the cadaveric kidney transplant allocation process: A decision-theoretic perspective. *Manage. Sci.* **42** 629–641.

Alagoz, O., C. L. Bryce, S. M. Shechter, A. J. Schaefer, C.-C. H. Chang, D. C. Angus, M. S. Roberts. 2005. Incorporating biological natural history in simulation models: Empiric estimates of the progression of end-stage liver disease. *Med. Decis. Making* **25** 620–632.

Alagoz, O., L. M. Maillart, A. J. Schaefer, M. S. Roberts. 2004. The optimal timing of living-donor liver transplant timing. *Manage. Sci.* **50**(10) 1420–1430.

Alagoz, O., L. M. Maillart, A. J. Schaefer, M. S. Roberts. 2007a. Choosing among cadaveric and living-donor livers. *Manage. Sci.* **53**(11) 1702–1715.

Alagoz, O., L. M. Maillart, A. J. Schaefer, M. S. Roberts. 2007b. Determining the acceptance of cadaveric livers using an implicit model of the waiting list. *Oper. Res.* **55**(1) 24–36.

Bagnell, J., A. Ng, J. Schneider. 2001. Solving uncertain Markov decision problems. *Technical report, Robotics Inst., Carnegie Melon University* 1–11.

Barlow, R. E., F. Proschan. 1965. *Mathematical Theory of Reliability*. John Wiley and Sons, Inc., New York.

Ben-Tal, A., D. den Hertog, A. De Waegenaere, B. Melenberg, G. Rennen. 2013. Robust solutions of optimization problems affected by uncertain probabilities. *Manage. Sci.* **59**(2) 341–357.

Cover, T. M., J. A. Thomas. 2006. *Elements of Infomration Theory; Second Edition*. John Wiley and Sons, Inc., Hoboken, New Jersey.

David, I., U. Yechiali. 1985. A time-dependent stopping problem with application to live organ transplants. *Oper. Res.* **33** 491504.

Denton, B. T., M. Kurt, N. D. Shah S. C. Bryant, S. A. Smith. 2009. A Markov decision process for optimizing the start time of statin therapy for patients with diabetes. *Med. Decis. Making* **29**(3) 351–367.

Givan, R., S. Leach, T. Dean. 2000. Bounded-parameter Markov decision processes. *Artif. Intell.* **122**(1-2) 71–109.

Hornberger, J., J. H. Ahn. 1997. Deciding eligibility for transplantation when a donor kidney becomes available. *Med. Decis. Making* **17** 160170.

Iyengar, G. 2003. Robust dynamic programming. *Technical report TR-2002-07*, Columbia University 1–29.

Iyengar, G. 2005. Robust dynamic programming. *Math. Oper. Res.* **30**(2) 257–280.

Kaufman, D. L., A. J. Schaefer. 2013. Robust modified policy iteration. *INFORMS J. Comput.* **25**(3) 396–410.

Kaufman, D. L., A. J. Schaefer, M. S. Roberts. 2011. Living-donor liver transplantation timing under ambiguous health state transition probabilities – extended abstract. *Proceedings of the 2011 Manufacturing and Service Operations Management (MSOM) Conference*. http://www-personal.umich.edu/∼davidlk/pubs/robustLiverExtendedAbstract.pdf.

Kreke, J. E., M. D. Bailey, A. J. Schaefer, D. C. Angus, M. S. Roberts. 2008. Modeling hospital discharge policies for patients with pneumonia-related sepsis. *IIE Trans.* **40**(9) 853–860.

Kurt, M., B. T. Denton, A. J. Schaefer, N. D. Shah, S. A. Smith. 2011. The structure of optimal statin initiation policies for patients with Type 2 diabetes. *IIE Trans. Health. Sys. Eng.* **1**(1) 49–65.

Law, A. M., W. D. Kelton. 1991. *Simulation Modeling and Analysis*. 2nd ed. McGraw-Hill, Inc., New York, NY.

Malinchoc, M., P. S. Kamath, F. D. Gordon, C. J. Peine, J. Rank, P. C. terBorg. 2000. A model to predict poor survival in patients undergoing transjugular intrahepatic portosystemic shunts. *Hepatology* **31** 864–871.

Mannor, S., O. Mebel, H. Xu. 2016. Robust MDPs with k-rectagular uncertainty. *Math. Oper. Res.* **41**(4) 1484–1509.

Mannor, S., D. Simester, P. Sun, J. N. Tsitsiklis. 2007. Bias and variance approximation in value function estimates. *Manage. Sci.* **53**(2) 308322.

Nilim, A., L. El Ghaoui. 2005. Robust control of Markov decision processes with uncertain transition matrices. *Oper. Res.* **53**(5) 780–798.

Organ Procurement and Transplantation Network. 2017. National data, June. `https://optn.transplant.hrsa.gov/data`.

Puterman, M. L. 1994. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. Wiley Series in Probability and Mathematical Statistics, John Wiley and Sons, Inc., New York.

Sandıkçı, B., L. M. Maillart, A. J. Schaefer, O. Alagoz, M. S. Roberts. 2008. Estimating the patient's price of privacy in liver transplantation. *Oper. Res.* **56**(6) 1393–1410.

Satia, J. K., R. L. Lave. 1973. Markov decision processes with uncertain transition probabilities. *Oper. Res.* **21**(3) 728–740.

Shechter, S. M., M. D. Bailey, A. J. Schaefer, M. S. Roberts. 2008. The optimal time to initiate HIV therapy under ordered health states. *Oper. Res.* **56**(1) 20–33.

White, C. C., H. K. Eldieb. 1994. Markov decision processes with imprecise transition probabilities. *Oper. Res.* **43** 739–749.

Wiesemann, W., D. Kuhn, B. Rustem. 2013. Robust Markov decision processes. *Math. Oper. Res.* **38**(1) 153–183.

Appendices

A. Uncertainty Set Calibration

For calibrating the uncertainty sets, the following statistical justification appears in (Iyengar 2003, Sec. 4.1); see also (Ben-Tal et al. 2013, Sec. 3.2). Recall that $\tilde{p}_s$ is the true (but unknown) state transition probabilities of which $\hat{p}_s$ is the maximum likelihood estimate, and N_s is the number of historical samples used for the estimate. Let $\chi^2_{|\mathcal{S}|-1}$ denote a chi-squared random variable with $|\mathcal{S}|-1$ degrees of freedom and let $F_{|\mathcal{S}|-1}(\cdot)$ denote its distribution function with inverse $F^{-1}_{|\mathcal{S}|-1}(\cdot)$. Then, as N_s becomes large, there is a convergence in probability:

$$N_s D(\tilde{p}||\hat{p}_s) \xrightarrow{p} \frac{\frac{d^2\phi}{dt^2}(1)}{2}\chi^2_{|\mathcal{S}|-1},$$

where $\phi(t)$ is associated with the more general concept of a ϕ-divergence statistic, and for relative entropy in particular $\phi(t) = t\log t - t + 1$; $\frac{d^2\phi}{dt^2}(1) = 1/\ln(a)$, where a is the base of the logarithm in the relative entropy measure and $\ln(\cdot)$ is the natural logarithm. Throughout, we assume that $a = e$, the natural base. Hence, information is measured in nats rather than bits. So, $\frac{d^2\phi}{dt^2}(1) = 1$. Therefore, $\mathbb{P}\{\tilde{p}_s \in \{p : D(p||\hat{p}_s) \le b\}\} \approx \mathbb{P}\{\chi^2_{|\mathcal{S}|-1} \le 2N_s b\} = F_{|\mathcal{S}|-1}(2N_s b)$. This implies that an (approximate) ω-confidence set for the true transition measure $\tilde{p}_s$ is given by

$$\mathcal{P}(s) \equiv \Big\{p \in \mathcal{M}(\mathcal{S}) : D(p||\hat{p}_s) \le F^{-1}_{|\mathcal{S}|-1}(\omega)/(2N_s)\Big\}.$$

Accordingly, set $\beta(s)$ by equation (6).

B. Proofs

In this section, we prove Theorems 1 and 2. We also relate, in Proposition 1 below, the sufficient conditions of Theorem 1 to the previously known conditions of Alagoz et al. (2004). The sufficient conditions assume that the transition matrix is IFR (Barlow and Proschan 1965):

Definition 1 Suppose that the vectors p_1 and p_2 are probability mass functions on $\mathcal{S}$. The vector p_2 is said to *stochastically dominate* p_1 if, for all $s' \in \mathcal{S}$, $\sum_{j=s'}^{j=H+1} p_2(j) \geq \sum_{j=s'}^{j=H+1} p_1(j)$.

Definition 2 The matrix $\hat{P}$ is said to be *IFR* if $\hat{p}_{s+1}$ stochastically dominates $\hat{p}_s$ for all $s < H$.

The proof of Theorem 1 leverages the convergence of value iteration. To this end, define the following:

$$\hat{v}_t(s) = \max\left\{ r(s,T),\ r(s,W) + \lambda \sum_{s' \in \mathcal{S}} p_s(s') \hat{v}_{t-1}(s') \right\},\ s \in \mathcal{S}^W,$$

$$\hat{v}_t(H+1) = 0,\ t > 0;$$

$$\hat{v}_0(s) = r(s,T),\ s \in \mathcal{S}.$$

By Theorem 6.3.1 of Puterman (1994), $\hat{v}_t(s)$ converges to $\hat{v}^*(s)$ in the limit $t \to \infty$. While the end-of-horizon rewards $\hat{v}_0$ may take on any values over the finite state space, it is key to our proof that $\hat{v}_0(s) = r(s,T)$. The following lemma is well known.

Lemma 1 (Puterman 1994, Lemma 4.7.2) *Suppose that a function $f(s)$, $s \in \mathcal{S}$, is non-increasing (non-decreasing) and that p_1 and p_2 are probability mass functions on $\mathcal{S}$. If p_1 stochastically dominates p_2, then $\sum_{s' \in \mathcal{S}} p_1(s') f(s') \leq (\geq) \sum_{s' \in \mathcal{S}} p_2(s') f(s')$.*

The following lemma drives the main result.

Lemma 2 *For the myopic model, assume that $\hat{P}$ is IFR and that (8) holds. Then, for $s \in \mathcal{S}^W$,*

$$\hat{v}_t(s) - \hat{v}_t(s+1) \geq r(s,T) - r(s+1,T),\ t \geq 0, \tag{B.1}$$

and

$$\hat{v}^*(s) - \hat{v}^*(s+1) \geq r(s,T) - r(s+1,T). \tag{B.2}$$

Proof: The proof is by induction on t. For $t = 0$, expression B.1 holds with equality. Assume that B.1 holds for t and consider $t+1$. Two exhaustive cases are $\hat{v}_{t+1}(s+1) = r(s+1,T)$ and $\hat{v}_{t+1}(s+1) > r(s+1,T)$. For case $\hat{v}_{t+1}(s+1) = r(s+1,T)$, $\hat{v}_{t+1}(s) - \hat{v}_{t+1}(s+1) = \hat{v}_{t+1}(s) - r(s+1,T) \geq r(s,T) - r(s+1,T)$ (since $\hat{v}_{t+1}(s) \geq r(s,T)$); B.1 holds for $t+1$. For the other case,

$\hat{v}_{t+1}(s+1) > r(s+1,T)$,

$$\begin{aligned}\hat{v}_{t+1}(s) - \hat{v}_{t+1}(s+1) &= \hat{v}_{t+1}(s) - r(s+1,W) - \lambda \sum_{s' \in \mathcal{S}} \hat{p}_{s+1}(s')\hat{v}_t(s') \\ &\geq r(s,W) + \lambda \sum_{s' \in \mathcal{S}} \hat{p}_s(s')\hat{v}_t(s') - r(s+1,W) - \lambda \sum_{s' \in \mathcal{S}} \hat{p}_{s+1}(s')\hat{v}_t(s') \\ &= r(s,W) - r(s+1,W) + \lambda \sum_{s' \in \mathcal{S}} (\hat{p}_s(s') - \hat{p}_{s+1}(s'))(\hat{v}_t(s') - r(s',T)) \\ &+ \lambda \sum_{s' \in \mathcal{S}} (\hat{p}_s(s') - \hat{p}_{s+1}(s'))r(s',T).\end{aligned}$$

Since $\hat{P}$ is assumed to be IFR, $\hat{p}_{s+1}$ stochastically dominates $\hat{p}_s$. Also, by the induction assumption the difference $\hat{v}_t(s') - r(s',T)$ is non-increasing in s'. So, it follows from Lemma 1 that $\sum_{s' \in \mathcal{S}} \hat{p}_{s+1}(s')(\hat{v}_t(s') - r(s',T)) \leq \sum_{s' \in \mathcal{S}} \hat{p}_s(s')(\hat{v}_t(s') - r(s',T) \Rightarrow \sum_{s' \in \mathcal{S}} (\hat{p}_s(s') - \hat{p}_{s+1}(s'))(\hat{v}_t(s') - r(s',T)) \geq 0$. Hence,

$$\begin{aligned}\hat{v}_{t+1}(s) - \hat{v}_{t+1}(s+1) &\geq r(s,W) - r(s+1,W) + 0 + \lambda \sum_{s' \in \mathcal{S}} (\hat{p}_s(s') - \hat{p}_{s+1}(s'))r(s',T) \\ &\geq r(s,T) - r(s+1,T),\end{aligned}$$

by condition (8). Expression (B.1) again holds for $t+1$. Therefore by induction, B.1 holds for all t. The convergence of value iteration then implies B.2. □

Proof of Theorem 1: Suppose that action T is optimal for some state $s \in \{1, \ldots, H-1\}$; the result holds otherwise. Then, $\hat{v}^*(s) - r(s,T) = 0$ and by Lemma 2, $0 = \hat{v}^*(s) - r(s,T) \geq \hat{v}^*(s+1) - r(s+1,T)$. Clearly, $\hat{v}^*(s+1) - r(s+1,T) \geq 0$. Therefore, $\hat{v}^*(s+1) - r(s+1,T) = 0$, and T is also optimal for $s+1$. □

Alagoz et al. (2004, Theorem 3) prove that the optimal myopic policy is of threshold form if the following sufficient conditions hold:

$$\hat{P} \text{ is IFR;} \tag{B.3}$$

$$r(s,T) \text{ is non-increasing in } s; \tag{B.4}$$

$$r(s,W) \text{ is non-increasing in } s; \tag{B.5}$$

$$\sum_{s'=j}^{H} \hat{p}_s(s') \leq \sum_{s'=j}^{H} \hat{p}_{s+1}(s'), \text{for } j = s+1, \ldots, H,\ s = 1, \ldots, H; \tag{B.6}$$

$$r(s,T) - r(s+1,T) \leq \lambda[\hat{p}_{s+1}(H+1) - \hat{p}_s(H+1)]r(s,T), \text{for } s = 1, \ldots, H-1. \tag{B.7}$$

Proposition 1 *If Conditions* (B.3)–(B.7) *hold, then the sufficient conditions of Theorem 1, that* $\hat{P}$ *is IFR and condition* (8)*, also hold.*

The proof of Proposition 1 uses the following lemma, which directly follows from the proof of Lemma 1 of Alagoz et al. (2004).

Lemma 3 *Let* $r(s)$ *be a non-increasing function. Then the following hold:*

(a) If $\hat{P}$ *is IFR, then*

$$\sum_{s'=1}^{s}[\hat{p}(s'|s)-\hat{p}(s'|s+1)]r(s') \geq \sum_{s'=1}^{s}[\hat{p}(s'|s)-\hat{p}(s'|s+1)]r(s),\ s=1,\ldots,H-1.$$

(b) If $\hat{P}$ *satisfies condition* (B.6)*, then*

$$\sum_{s''=s+1}^{H}[\hat{p}(s''|s)-\hat{p}(s''|s+1)]r(s'') \geq \sum_{s''=s+1}^{H}[\hat{p}(s''|s)-\hat{p}(s''|s+1)]r(s+1),\ s=1,\ldots,H-1.$$

Proof of Proposition 1: Under condition (B.3), $\hat{P}$ is IFR. It remains to show condition (8). By condition (B.4), condition (B.6), and $\lambda > 0$,

$$0 \leq \lambda\left(\sum_{s''=s+1}^{H}\left[\hat{p}_{s+1}(s'')-\hat{p}_s(s'')\right]\right)[r(s,T)-r(s+1,T)],\ s=1,\ldots,H-1. \tag{B.8}$$

Adding this to condition (B.7) yields

$$\begin{aligned} r(s,T)-r(s+1,T) \leq{}& \lambda\left(\sum_{s''=s+1}^{H}\left[\hat{p}_{s+1}(s'')-\hat{p}_s(s'')\right]\right)[r(s,T)-r(s+1,T)] \\ &+\lambda[\hat{p}_{s+1}(H+1)-\hat{p}_s(H+1)]r(s,T),\ s=1,\ldots,H-1. \end{aligned}$$

Then, for $s=1,\ldots,H-1$,

$$\begin{aligned} &r(s,T)-r(s+1,T) \\ &\quad\leq \lambda\left(\sum_{s''=s+1}^{H}\left[\hat{p}_{s+1}(s'')-\hat{p}_s(s'')\right]\right)r(s,T)-\lambda\left(\sum_{s''=s+1}^{H}\left[\hat{p}_{s+1}(s'')-\hat{p}_s(s'')\right]\right)r(s+1,T) \\ &\qquad+\lambda[\hat{p}_{s+1}(H+1)-\hat{p}_s(H+1)]r(s,T) \\ &\quad=\lambda\Bigg(\left[\left(1-\sum_{s''=s+1}^{H+1}\hat{p}_s(s'')\right)-\left(1-\sum_{s''=s+1}^{H+1}\hat{p}_{s+1}(s'')\right)\right]r(s,T) \\ &\qquad+\left(\sum_{s''=s+1}^{H}\left[\hat{p}_s(s'')-\hat{p}_{s+1}(s'')\right]\right)r(s+1,T)\Bigg) \\ &\quad=\lambda\left(\left(\sum_{s'=1}^{s}\left[\hat{p}_s(s')-\hat{p}_{s+1}(s')\right]\right)r(s,T)+\left(\sum_{s''=s+1}^{H}\left[\hat{p}_s(s'')-\hat{p}_{s+1}(s'')\right]\right)r(s+1,T)\right), \end{aligned}$$

where the last equality follows since total probability is one. Then by Lemma 3,

$$\begin{aligned} r(s,T) - r(s+1,T) &\leq \lambda\Bigg(\left(\sum_{s'=1}^{s} \left[\hat{p}_s(s') - \hat{p}_{s+1}(s')\right] r(s',T) \right) \\ &\quad + \left(\sum_{s''=s+1}^{H} \left[\hat{p}_s(s'') - \hat{p}_{s+1}(s'')\right] r(s'',T) \right) \Bigg) \\ &= \lambda \left(\sum_{s'=1}^{H} \left[\hat{p}_s(s') - \hat{p}_{s+1}(s')\right] r(s',T) \right). \end{aligned}$$

This inequality together with condition (B.5) implies condition (8). □

Proof of Theorem 2: The optimal robust value function for model i, denoted v_i^*, can be computed with robust value iteration (Iyengar 2005). To this end, define

$$\begin{aligned} v_t^i(s) &= \max\left\{r(s,T),\ r(s,W) + \lambda\sigma_{\mathcal{P}_i(s)}(v_{t-1}^i)\right\},\ s \in \mathcal{S}^W, \\ v_t^i(H+1) &= 0,\ t > 0; \\ v_0^i(s) &= 0,\ s \in \mathcal{S}, \end{aligned}$$

where $\mathcal{P}_i(s) = \{p \in \mathcal{M}(\mathcal{S}) : D(p||\hat{p}_s) \leq \beta_i(s)\}$.

We will first show inductively that $v_t^2(s) \leq v_t^1(s)$, $s \in \mathcal{S}^W$. This relationship holds trivially for $t = 0$. Assume that it holds for t. It easily follows that $\sigma_{\mathcal{P}_1(s)}(v_t^2) \leq \sigma_{\mathcal{P}_1(s)}(v_t^1)$. Since $\beta_1(s) \leq \beta_2(s)$, $\mathcal{P}_1(s) \subseteq \mathcal{P}_2(s)$ and it follows that $\sigma_{\mathcal{P}_2(s)}(v_t^2) \leq \sigma_{\mathcal{P}_1(s)}(v_t^2)$. Therefore, $\sigma_{\mathcal{P}_2(s)}(v_t^2) \leq \sigma_{\mathcal{P}_1(s)}(v_t^1)$. It is then straight forward to show that $v_{t+1}^2(s) \leq v_{t+1}^1(s)$; details are omitted. Hence, by induction, $v_t^2(s) \leq v_t^1(s)$ holds for all t; and, by the convergence of robust value iteration as $t \to \infty$, $v_2^*(s) \leq v_1^*(s)$.

Suppose $d_1^*(s) = T$, so that $v_1^*(s) = r(s,T)$. Then, since $r(s,T) = v_1^*(s) \geq v_2^*(s) \geq r(s,T)$, $v_2^*(s) = r(s,T)$ and $d_2^*(s) = T$ is indeed optimal.

Since $\{\hat{p}_s\} \subseteq \mathcal{P}_i(s)$, it follows similarly that $\hat{d}^*(s) = T \Rightarrow d_i^*(s) = T$. □

C. Expanded Simulation Experiment

This section presents expanded simulation results. For discussion, see Sec. 4.3.

Disease Group 1 ($\omega = .95$): Tables C.2, C.3, and C.4 display results for DG 1 for 10,000 replications each, and are analogous to the DG 2 Tables 4 ,5, and 6, respectively.

Various Confidence Levels: Tables C.5, C.6, C.7, and C.8 display DG 2 results for various confidence levels, for $\omega = .005$, .05, .5, and .995, respectively, for 5,000 replications each. Results for $\omega = .95$ are reported in Tables 4(b), 5(b), and 6. For myopic performance, see Table 4(a). Table 7 provides a summary.

	Liver 1	Liver 3	Liver 5	Liver 7	Liver 9	Liver 11	Liver 13
Patient 1	4797.5	4699.7	4617.8	4534.9	4460.6	4337.2	4091.5
Patient 2	4713.6	4604.2	4513.2	4421.6	4339.1	4203.3	3935.2
Patient 3	4412.7	4265.4	4144.6	4024.0	3916.3	3742.8	3408.9
Patient 4	4253.1	4087.5	3952.6	3818.8	3700.8	3510.6	3152.5
Patient 5	4073.7	3889.4	3739.4	3593.6	3465.2	3260.2	2882.5
Patient 6	3044.3	2790.9	2596.3	2415.9	2265.5	2039.0	1662.8
Patient 7	2853.5	2594.7	2398.4	2219.1	2070.5	1848.9	1487.3

(a) Optimal discounted life expectancy (discounted days)

	Liver 1	Liver 3	Liver 5	Liver 7	Liver 9	Liver 11	Liver 13
Patient 1	9.74%	9.92%	9.83%	9.94%	9.98%	10.14%	11.28%
Patient 2	9.74%	9.99%	9.92%	10.08%	10.16%	10.39%	11.79%
Patient 3	9.98%	10.49%	10.55%	10.94%	11.22%	11.83%	14.31%
Patient 4	10.24%	10.91%	11.08%	11.64%	12.05%	12.92%	16.01%
Patient 5	10.65%	11.54%	11.88%	12.62%	13.20%	14.40%	18.53%
Patient 6	15.84%	18.80%	20.56%	23.17%	25.46%	29.32%	37.76%
Patient 7	17.62%	21.15%	23.26%	26.17%	28.71%	32.96%	41.84%

(b) Bound: % below optimal

Table C.1: DG 1 optimal myopic discounted life expectancies and performance bounds for transplanting immediately, at 6 MELD. This table is analogous to DG 2 Table 3.

	Liver 1	Liver 3	Liver 5	Liver 7	Liver 9	Liver 11	Liver 13
Patient 1	15.71	15.62	15.49	15.58	15.60	15.58	16.04
Patient 2	15.68	15.60	15.46	15.58	15.60	15.61	16.17
Patient 3	15.65	15.62	15.50	15.72	15.81	16.04	16.87
Patient 4	15.70	15.69	15.60	15.88	16.05	16.40	17.51
Patient 5	15.77	15.83	15.77	16.14	16.39	16.95	18.38
Patient 6	17.38	18.38	18.64	19.34	19.82	20.12	20.35
Patient 7	17.99	19.09	19.33	19.80	20.08	20.42	21.02

(a) Average myopic thresholds (MELD)

	Liver 1	Liver 3	Liver 5	Liver 7	Liver 9	Liver 11	Liver 13
Patient 1	12.44	12.46	12.42	12.46	12.40	12.45	12.61
Patient 2	12.44	12.47	12.43	12.48	12.42	12.49	12.69
Patient 3	12.47	12.53	12.49	12.59	12.55	12.71	13.19
Patient 4	12.51	12.59	12.57	12.71	12.67	12.94	13.52
Patient 5	12.57	12.70	12.70	12.89	12.89	13.25	13.85
Patient 6	13.61	13.88	13.98	14.15	14.33	14.68	16.07
Patient 7	13.83	14.01	14.10	14.41	14.76	15.34	16.49

(b) Average robust thresholds (MELD) for $\omega = .95$

Table C.2: DG 1 average robust and myopic thresholds, averaged over 10,000 replications. This table is analogous to DG 2 Table 4.

	Liver 1	Liver 3	Liver 5	Liver 7	Liver 9	Liver 11	Liver 13
Patient 1	0.17%	0.17%	0.16%	0.17%	0.17%	0.18%	0.18%
Patient 2	0.17%	0.17%	0.15%	0.16%	0.17%	0.18%	0.17%
Patient 3	0.17%	0.17%	0.16%	0.17%	0.18%	0.20%	0.14%
Patient 4	0.17%	0.18%	0.17%	0.18%	0.19%	0.19%	0.05%
Patient 5	0.17%	0.19%	0.18%	0.18%	0.19%	0.12%	0.16%
Patient 6	0.06%	0.17%	0.19%	0.18%	0.12%	0.10%	0.15%
Patient 7	0.12%	0.20%	0.20%	0.14%	0.10%	0.12%	0.19%

(a) Average myopic difference in life expectancy below optimal (%)

	Liver 1	Liver 3	Liver 5	Liver 7	Liver 9	Liver 11	Liver 13
Patient 1	1.70%	1.66%	1.64%	1.69%	1.71%	1.68%	1.99%
Patient 2	1.70%	1.67%	1.65%	1.71%	1.73%	1.70%	2.06%
Patient 3	1.75%	1.73%	1.73%	1.83%	1.89%	1.87%	2.10%
Patient 4	1.79%	1.78%	1.80%	1.90%	2.00%	1.87%	2.02%
Patient 5	1.85%	1.83%	1.87%	1.95%	2.04%	1.82%	2.31%
Patient 6	1.81%	2.02%	2.27%	3.03%	3.71%	4.42%	4.27%
Patient 7	1.99%	2.58%	2.96%	3.67%	4.12%	4.51%	4.55%

(b) Average robust difference in life expectancy below optimal (%)

Table C.3: DG 1 average differences in life expectancy below optimal (%). This table is analogous to DG 2 Table 5.

	Liver 1	Liver 3	Liver 5	Liver 7	Liver 9	Liver 11	Liver 13
Patient 1	3.63%	3.73%	2.69%	3.32%	3.02%	4.38%	3.04%
Patient 2	3.48%	3.54%	2.61%	3.23%	3.16%	5.00%	3.76%
Patient 3	3.30%	4.10%	2.88%	4.45%	4.90%	2.71%	0.66%
Patient 4	3.73%	4.90%	3.82%	1.97%	2.90%	5.21%	1.05%
Patient 5	4.52%	6.37%	5.32%	3.05%	5.83%	0.41%	
Patient 6	0.44%	0.23%	0.09%	0.23%	0.03%	0.08%	0.28%
Patient 7	0.15%	0.07%	0.17%	0.04%	0.15%	0.09%	3.29%

(a) Robust beats myopic fraction of replications

	Liver 1	Liver 3	Liver 5	Liver 7	Liver 9	Liver 11	Liver 13
Patient 1	6.60%	7.40%	6.42%	6.97%	5.99%	8.37%	5.04%
Patient 2	6.52%	7.28%	6.33%	7.00%	6.28%	9.36%	5.42%
Patient 3	6.50%	8.33%	7.01%	8.37%	7.93%	7.04%	1.51%
Patient 4	6.99%	9.38%	8.26%	5.59%	5.50%	9.15%	1.50%
Patient 5	7.86%	10.97%	9.79%	6.06%	8.00%	3.25%	0.30%
Patient 6	1.19%	0.60%	0.34%	0.31%	0.10%	0.08%	0.59%
Patient 7	0.49%	0.21%	0.23%	0.12%	0.20%	0.10%	4.64%

(b) Robust beats or ties fraction of replications

Table C.4: DG 1 fraction of replications in which the robust policy with $\omega = .95$ (a) strictly outperforms (beats) and (b) performs at least as well as (beats or ties) the myopic policy. Blank cells are zeros. This table is analogous to DG 2 Table 6

	Liver 1	Liver 3	Liver 5	Liver 7	Liver 9	Liver 11	Liver 13
Patient 1	19.30	19.79	20.05	20.39	20.72	21.19	21.87
Patient 2	19.68	20.16	20.53	20.86	21.20	21.73	22.60
Patient 3	20.93	21.47	22.08	22.46	22.83	23.58	24.62
Patient 4	21.55	22.18	22.86	23.25	23.67	24.34	25.25
Patient 5	22.26	22.97	23.67	24.03	24.39	25.02	25.80
Patient 6	25.43	25.86	26.13	26.38	26.63	26.94	27.85
Patient 7	25.79	26.10	26.35	26.66	26.94	27.34	28.25

(a) Average robust threshold

	Liver 1	Liver 3	Liver 5	Liver 7	Liver 9	Liver 11	Liver 13
Patient 1	1.08%	1.17%	1.34%	1.47%	1.79%	2.39%	3.00%
Patient 2	1.25%	1.27%	1.75%	2.04%	2.29%	2.66%	2.92%
Patient 3	2.01%	2.52%	2.72%	2.67%	2.62%	2.35%	1.96%
Patient 4	2.47%	2.72%	2.65%	2.48%	2.28%	1.92%	1.68%
Patient 5	2.65%	2.61%	2.29%	2.06%	2.00%	1.66%	1.37%
Patient 6	1.55%	1.17%	1.11%	1.33%	1.37%	1.57%	1.26%
Patient 7	1.37%	1.14%	1.27%	1.46%	1.44%	1.55%	1.12%

(b) Average robust difference in life expectancy below optimal (%)

	Liver 1	Liver 3	Liver 5	Liver 7	Liver 9	Liver 11	Liver 13
Patient 1	18.56%	28.52%	18.14%	4.74%	5.60%	6.24%	7.20%
Patient 2	23.08%	18.16%	3.26%	4.34%	7.40%	5.44%	12.94%
Patient 3	5.26%	3.30%	5.62%	8.40%	12.62%	23.38%	10.74%
Patient 4	4.54%	6.52%	12.16%	18.00%	24.50%	10.20%	18.58%
Patient 5	7.34%	13.80%	22.22%	31.32%	9.56%	17.62%	7.54%
Patient 6	4.64%	7.06%	11.54%	2.90%	5.00%	8.62%	9.54%
Patient 7	7.92%	11.38%	2.78%	5.18%	10.40%	6.08%	17.92%

(c) Robust beats myopic fraction of replications

	Liver 1	Liver 3	Liver 5	Liver 7	Liver 9	Liver 11	Liver 13
Patient 1	28.18%	35.46%	22.38%	8.26%	9.06%	9.10%	11.14%
Patient 2	30.02%	22.78%	6.54%	7.04%	10.14%	9.20%	19.18%
Patient 3	8.08%	6.82%	10.34%	13.38%	18.64%	35.42%	22.90%
Patient 4	7.26%	11.82%	20.66%	27.00%	34.06%	24.24%	29.16%
Patient 5	11.70%	23.50%	35.12%	42.72%	20.38%	31.12%	16.70%
Patient 6	14.78%	18.00%	20.04%	8.40%	9.86%	13.16%	15.80%
Patient 7	16.64%	19.78%	8.60%	9.46%	14.76%	10.54%	23.86%

(d) Robust beats or ties fraction of replications

Table C.5: DG 2 robust performance for $\omega = .005$. Blank cells are zeros. A summary is presented in Table 7(b).

	Liver 1	Liver 3	Liver 5	Liver 7	Liver 9	Liver 11	Liver 13
Patient 1	18.66	19.09	19.26	19.54	19.90	20.27	20.66
Patient 2	18.95	19.44	19.68	19.96	20.30	20.68	21.09
Patient 3	20.06	20.52	20.90	21.22	21.50	21.99	22.74
Patient 4	20.58	21.01	21.45	21.79	22.09	22.60	23.54
Patient 5	21.10	21.55	22.08	22.39	22.69	23.32	24.32
Patient 6	23.86	24.56	25.14	25.38	25.69	26.08	26.82
Patient 7	24.34	25.00	25.54	25.79	26.02	26.34	27.21

(a) Average robust threshold

	Liver 1	Liver 3	Liver 5	Liver 7	Liver 9	Liver 11	Liver 13
Patient 1	1.51%	1.75%	2.02%	2.25%	2.57%	3.31%	4.17%
Patient 2	1.82%	1.91%	2.53%	2.92%	3.20%	3.77%	4.51%
Patient 3	2.88%	3.51%	3.98%	4.02%	4.09%	4.12%	4.28%
Patient 4	3.51%	3.97%	4.20%	4.10%	4.05%	3.96%	3.90%
Patient 5	3.92%	4.17%	4.09%	3.95%	3.99%	3.77%	3.34%
Patient 6	3.56%	2.93%	2.44%	2.55%	2.45%	2.36%	2.30%
Patient 7	3.27%	2.60%	2.24%	2.40%	2.38%	2.47%	2.14%

(b) Average robust difference in life expectancy below optimal (%)

	Liver 1	Liver 3	Liver 5	Liver 7	Liver 9	Liver 11	Liver 13
Patient 1	11.88%	20.30%	5.10%	2.94%	2.10%	3.92%	2.00%
Patient 2	15.58%	5.58%	2.52%	2.18%	4.46%	1.76%	3.16%
Patient 3	2.80%	0.98%	1.86%	1.88%	3.10%	6.44%	5.70%
Patient 4	1.30%	2.34%	3.22%	4.90%	6.42%	4.92%	12.20%
Patient 5	2.88%	3.72%	6.40%	9.16%	4.44%	11.08%	1.30%
Patient 6	0.86%	1.76%	3.06%	1.00%	2.10%	3.78%	3.32%
Patient 7	1.58%	3.22%	0.96%	2.22%	4.74%	1.56%	3.48%

(c) Robust beats myopic fraction of replications

	Liver 1	Liver 3	Liver 5	Liver 7	Liver 9	Liver 11	Liver 13
Patient 1	15.52%	22.74%	6.60%	3.82%	3.10%	4.24%	2.28%
Patient 2	17.80%	7.26%	3.34%	2.70%	4.94%	2.02%	3.44%
Patient 3	3.36%	1.38%	2.24%	2.24%	3.34%	7.76%	7.20%
Patient 4	1.56%	2.78%	3.74%	5.62%	7.28%	6.84%	14.16%
Patient 5	3.16%	4.36%	7.92%	10.56%	5.74%	13.58%	3.28%
Patient 6	2.94%	5.30%	5.94%	2.52%	3.10%	4.60%	3.68%
Patient 7	3.60%	5.92%	3.24%	3.24%	5.20%	2.12%	3.90%

(d) Robust beats or ties fraction of replications

Table C.6: DG 2 robust performance for $\omega = .05$. Blank cells are zeros. A summary is presented in Table 7(b).

	Liver 1	Liver 3	Liver 5	Liver 7	Liver 9	Liver 11	Liver 13
Patient 1	18.08	18.19	18.23	18.35	18.63	18.88	19.61
Patient 2	18.14	18.33	18.42	18.60	18.93	19.21	19.82
Patient 3	18.71	19.17	19.35	19.63	19.88	20.15	20.39
Patient 4	19.12	19.59	19.78	20.02	20.24	20.51	20.71
Patient 5	19.57	19.95	20.18	20.39	20.58	20.87	21.12
Patient 6	21.23	21.59	22.00	22.28	22.59	23.30	24.72
Patient 7	21.54	21.94	22.43	22.72	23.09	24.00	25.47

(a) Average robust threshold

	Liver 1	Liver 3	Liver 5	Liver 7	Liver 9	Liver 11	Liver 13
Patient 1	1.93%	2.59%	3.05%	3.57%	4.16%	5.15%	5.60%
Patient 2	2.53%	3.11%	3.97%	4.59%	5.02%	5.84%	6.26%
Patient 3	4.61%	5.39%	6.21%	6.41%	6.58%	7.03%	7.88%
Patient 4	5.54%	6.06%	6.74%	6.89%	6.99%	7.33%	8.38%
Patient 5	6.21%	6.64%	7.10%	7.16%	7.42%	7.73%	8.60%
Patient 6	7.79%	7.78%	7.69%	7.73%	7.65%	7.22%	5.77%
Patient 7	7.78%	7.66%	7.56%	7.62%	7.41%	6.56%	4.76%

(b) Average robust difference in life expectancy below optimal (%)

	Liver 1	Liver 3	Liver 5	Liver 7	Liver 9	Liver 11	Liver 13
Patient 1	2.34%	5.06%	0.54%	0.98%	0.20%	0.54%	0.10%
Patient 2	3.88%	0.68%	1.12%	0.28%	0.46%	0.12%	
Patient 3	0.36%	0.04%	0.18%			0.02%	0.10%
Patient 4	0.06%	0.22%			0.04%	0.10%	0.40%
Patient 5	0.36%		0.02%	0.06%	0.10%	0.36%	0.10%
Patient 6	0.02%	0.22%	0.60%	0.08%	0.16%	0.02%	0.20%
Patient 7	0.20%	0.58%	0.16%	0.24%	0.04%	0.02%	0.20%

(c) Robust beats myopic fraction of replications

	Liver 1	Liver 3	Liver 5	Liver 7	Liver 9	Liver 11	Liver 13
Patient 1	3.62%	5.28%	0.68%	1.02%	0.26%	0.56%	0.10%
Patient 2	4.20%	0.84%	1.16%	0.30%	0.48%	0.12%	
Patient 3	0.40%	0.06%	0.18%			0.02%	0.10%
Patient 4	0.06%	0.22%			0.04%	0.10%	0.42%
Patient 5	0.36%		0.02%	0.06%	0.10%	0.38%	0.10%
Patient 6	0.04%	0.22%	0.64%	0.12%	0.16%	0.02%	0.20%
Patient 7	0.20%	0.60%	0.20%	0.24%	0.04%	0.02%	0.20%

(d) Robust beats or ties fraction of replications

Table C.7: DG 2 robust performance for $\omega = .5$. Blank cells are zeros. A summary is presented in Table 7(b).

	Liver 1	Liver 3	Liver 5	Liver 7	Liver 9	Liver 11	Liver 13
Patient 1	17.97	18.02	18.02	18.03	18.04	18.04	18.14
Patient 2	18.00	18.03	18.03	18.03	18.04	18.06	18.21
Patient 3	18.04	18.05	18.07	18.11	18.17	18.26	18.73
Patient 4	18.05	18.11	18.14	18.21	18.33	18.48	19.00
Patient 5	18.09	18.22	18.27	18.38	18.56	18.73	19.28
Patient 6	19.00	19.31	19.44	19.62	19.82	20.05	20.29
Patient 7	19.17	19.51	19.65	19.82	20.05	20.29	20.58

(a) Average robust threshold

	Liver 1	Liver 3	Liver 5	Liver 7	Liver 9	Liver 11	Liver 13
Patient 1	2.04%	2.76%	3.27%	3.94%	4.93%	6.36%	8.22%
Patient 2	2.65%	3.44%	4.44%	5.34%	6.32%	7.67%	9.30%
Patient 3	5.52%	7.12%	8.31%	9.06%	9.74%	10.70%	11.42%
Patient 4	7.20%	8.62%	9.69%	10.31%	10.80%	11.53%	12.10%
Patient 5	8.74%	9.90%	10.81%	11.23%	11.65%	12.36%	12.62%
Patient 6	12.82%	13.03%	13.64%	14.00%	14.20%	14.78%	15.88%
Patient 7	13.23%	13.33%	14.05%	14.48%	14.57%	15.11%	16.08%

(b) Average robust difference in life expectancy below optimal (%)

	Liver 1	Liver 3	Liver 5	Liver 7	Liver 9	Liver 11	Liver 13
Patient 1	0.66%	1.00%	0.10%	0.16%	0.02%	0.02%	
Patient 2	0.98%	0.10%	0.18%		0.02%		
Patient 3	0.02%						
Patient 4							
Patient 5							
Patient 6							
Patient 7							

(c) Robust beats myopic fraction of replications

	Liver 1	Liver 3	Liver 5	Liver 7	Liver 9	Liver 11	Liver 13
Patient 1	1.58%	1.08%	0.14%	0.16%	0.02%	0.02%	
Patient 2	1.14%	0.12%	0.18%		0.02%		
Patient 3	0.02%						
Patient 4							
Patient 5							
Patient 6							
Patient 7							

(d) Robust beats or ties fraction of replications

Table C.8: DG 2 robust performance for $\omega = .995$. Blank cells are zeros. A summary is presented in Table 7(b).