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

David L. Kaufman
Department of Industrial and Operations Engineering, University of Michigan
1205 Beal Avenue, Ann Arbor, MI 48109, USA
davidlk@umich.edu

Andrew J. Schaefer
Department of Industrial Engineering, University of Pittsburgh
1048 Benedum Hall, Pittsburgh, PA 15261, USA
schaefer@ie.pitt.edu

Mark S. Roberts
Division of General Internal Medicine, School of Medicine, University of Pittsburgh
200 Meyran Avenue, 2nd Floor, Pittsburgh, Pennsylvania 15213, USA
robertsm@upmc.edu

February 22, 2011

Markov decision process 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 statistical ambiguity, when for example some critical health states are seldom visited historically. We account for ambiguity in a robust dynamic programming model for therapy initiation in which worst-case transition probabilities are chosen from a set of probability measures constructed using relative entropy bounds. For a *myopic* model that instead uses point estimates for transition probabilities, we demonstrate, for ordered states, new sufficient conditions that guarantee an optimal control-limit policy. For the *robust* model, we prove that therapy is initiated sooner, in additional states, as the level of ambiguity increases. We apply these models to the problem of deciding when to undergo a living-donor liver transplantation, and we present the results of numerical studies with clinically obtained data. We find that the robust policies are in some cases closer to what is actually being done in practice. Often, however, transplants are occurring even sooner than what is suggested by the most conservative robust policies.

1. Introduction

Medical 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*.

Unfortunately, MDP solutions can be sensitive to errors in the estimation of transition probabilities (c.f., Iyengar 2005, Nilim and El Ghaoui 2005). Therefore, it may be important to account for ambiguity in transition probabilities for MDP models of therapeutic decisions in a systematic and statistically meaningful way. Several researchers, including Satia and Lave (1973), White and Eldieb (1994), 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 way to mitigate the effects of ambiguity in MDPs. While robust DP is a promising methodology, very few applications have been reported. Because physiological data are often not available in abundance, robust DP may be especially relevant in the medical decision-making arena. To our knowledge, this is the first report of a medical application of robust DP.

We consider modeling the decisions faced by patients with end-stage liver disease (ESLD),

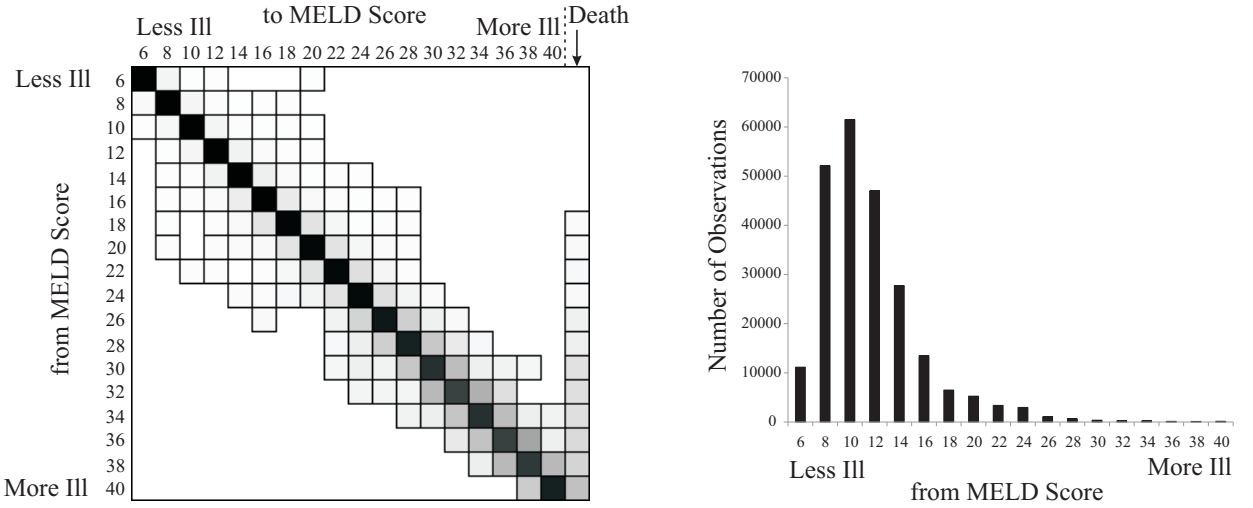
for which the only viable treatment is liver transplantation. For this treatment it must be decided when to undergo a liver transplantation by taking into account the balance between the patient's pre-transplant discounted life expectancy (or alternatively, quality-adjusted life years) and post-transplant discounted life expectancy. Some patients are fortunate to receive a live donated by a living donor.

The Living-Donor Model (**LDM**) of Alagoz et al. (2004) is an MDP model for the optimal time to undergo a liver transplantation when a patient is only considering a living-donor organ. This model demonstrates some interesting results. In particular, transplanting right away is typically suboptimal. The total reward, pre-transplant plus post-transplant life expectancy, can be greater if the patient waits. This is true despite the fact that post-transplant life expectancy decreases as the patient's health worsens.

At the heart of the LDM is a natural history model, a Markov chain model of the pre-transplant progression of an ESLD. 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 (Model for End-Stage Liver Disease) scores, integer values that represent the state of health of the patient. MELD scores are 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, with a higher value indicating that the patient is more ill. The transition matrix models the health state progression for a patient who chooses to not 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.

In Figure 1(a), there are several things worth noting. First, this 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 MELD 18. The probability of death increases as the MELD score increases. As documented by Alagoz et al. (2004), 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, p. 1423).

Figure 1(b) is a histogram of 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 MELD 10 at 61,499. There are fewer observation at MELD 6, perhaps because some patients do not present at such a



(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. 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, corresponding to the rows of the transition matrix.

low score. The number of observations decreases significantly at the higher MELD scores. MELD 40 only has 142 observations. On one hand, patients at MELD 40 might not live long, or patients die prior to reaching MELD 40. On the other hand, transplantations often occur much sooner than this.

If estimates are made with a large total number of observations, how does one know if it is enough? This question is particularly relevant because critical events, like large changes in health or death, may occur with very small probability. Moreover, the quality of estimates may vary significantly by state. There may be ample data for some states, but not for others. For some states, the quality of estimates might not matter. For example, under the optimal policy, some states might only be reachable with a very small probability. Thus, the importance of having good estimates may depend on the optimal actions, which are not known a priori.

Anecdotally, it seems that liver transplantations are occurring sooner (when patients are healthier) than what the MDP models prescribe; we present some evidence in Section 3. There may be several reasons for this. One one hand, transplantations may simply be

occurring too early. That is, following the optimal actions of the decision models could result in better outcomes. On the other hand, the models might not be capturing some of the tradeoffs considered by physicians. Perhaps physicians are inherently conservative. Perhaps they are risk adverse and weigh the possibility of death more. It might be possible too that the models are flawed in some way. We will assume here that the basic model structures are correct, except we will allow for the possibility of ambiguity in the health state transition probabilities of the pre-transplant natural history models. We want study the possible effects of ambiguity on the resulting policies.

Robust DP is a max-min optimization framework for mitigating the effects of ambiguity in transition probabilities. Rather than simply using point estimates, statistically meaningful sets of conditional measures are constructed. These so-called "uncertainty sets," denoted $\mathcal{P}$, can be constructed to correspond to any given level of confidence, and they are easily parameterized. The more ambiguity, the larger $\mathcal{P}$. 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 $\mathcal{P}$. In essence, robust DP introduces an additional degree of freedom into the decision model, while at the same time maintaining computational tractability.

We consider two models. The **myopic model** is a standard DP that uses point estimates for transition probabilities. The myopic model has been studied previously, and it is not limited to liver transplantation problems. The **robust model** model is a robust DP. There are several questions of interest that we address: How do the myopic and robust policies compare structurally? How do the actions change as the level of ambiguity varies? Which perform better, the myopic or the robust policies? How does one measure performance if the true distributions are unknown? Do the robust policies mitigate the effects of ambiguity? How do the robust policies compare to what is being done in practice? We address these questions in the context of the LDM using clinically obtained data.

2. Modeling Framework and Structural Results

In every (discrete) time period, a patient who has not yet initiated therapy (transplantation in the case of LDM) faces the decision whether to initiate therapy (T) or to wait (W). Denote by $\mathcal{S}^W$ the set $\{1, 2, \ldots, H\}$. The state space is $\mathcal{S} \equiv \mathcal{S}^W \cup \{H+1\}$. 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 patient is alive and the decision process continues. The time horizon is infinite. Of course, the life expectancy of a patient will be finite.

The objective is to maximize expected total discounted reward. The discount factor is λ, $0 < \lambda < 1$. 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. For example, $r(s,T)$ could equal the expected total discounted 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; $r(H+1,T) = 0$.

2.1 The Myopic Model

The myopic model assumes that states transition according to probability measures that are known. Transition probabilities are stored in an $H \times (H+1)$ transition matrix $\hat{P}$ with rows $\hat{p}_s$, $s \in \mathcal{S}^W$; e.g., Figure 1(a). When the decision is to wait, the probability of tansitioning from $s \in \mathcal{S}^W$ in the current time period to $s' \in \mathcal{S}$ in the next time period is $\hat{p}_s(s')$.

Under the expected total discounted reward criterion, it is well known that there exists an optimal stationary deterministic Markovian policy, 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 *optimality equations*:

$$\begin{aligned} \hat{v}^*(s) &= \max\left\{ r(s,T),\ r(s,W) + \lambda \sum_{s' \in \mathcal{S}} \hat{p}_s(s') \hat{v}^*(s') \right\},\ s \in \mathcal{S}^W, \\ \hat{v}^*(H+1) &= 0. \end{aligned}$$

Alagoz et al. (2004, Theorem 3) were the first to provide conditions that guarantee the existence of an optimal decision rule characterized by a *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 provide new sufficient conditions that relax those previously known.

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

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

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

2.2 The Robust Model

The robust model is formulated as an infinite-horizon robust DP. For a background, the reader is referred to Iyengar (2005) and Nilim and El Ghaoui (2005). Robust DP can be thought of as a game between the decision-maker and an adversary called "nature." While the decision-maker seeks to maximize total reward by choosing the best policy, nature responds by minimizing the decision-maker's reward by selecting worst-case transition measures. The decision-maker chooses an action $a=d(s)$. If $a=W$, nature responds by choosing any transition measure in a given uncertainty set $\mathcal{P}(s)$. The robust DP optimality equations are:

$$\begin{aligned} v^*(s) &= \max\left\{r(s,T),\ r(s,W)+\lambda\inf_{p\in\mathcal{P}(s)}\sum_{s'\in\mathcal{S}}p(s')v^*(s')\right\},\ s\in\mathcal{S}^W, \quad (2)\\ v^*(H+1) &= 0. \end{aligned}$$

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

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

This is referred to as the *inner problem*. Note in (2) that $\sigma_{\mathcal{P}(s)}(v^*)$ is solved for each $s\in\mathcal{S}^W$. To compute optimal solutions we use the Robust Modified Policy Iteration algorithm of Kaufman and Schaefer (2011).

Several methods for constructing $\mathcal{P}(s)$ have been proposed in the literature. Most significantly, Bagnell et al. (2001), Iyengar (2005), and Nilim and El Ghaoui (2005) suggest sets based on *relative entropy* bounds. The appeal of relative entropy bounds lies in their statistical interpretation. An ω-confidence set for the true transition measure $\tilde{p}_s$ is of the form

$$\begin{aligned} \mathcal{P}(s) &= \{p\in\mathcal{M}(\mathcal{S}):D(p||\hat{p}_s)\leq\beta(s)\},\\ \beta(s) &= F^{-1}_{|\mathcal{S}|-1}(\omega)/(2N_s), \end{aligned}$$

where $\mathcal{M}(\mathcal{S})$ is the set of all probability measures on $\mathcal{S}$, $D(p||\hat{p}_s)$ is the Kullback-Leibler divergence between $\hat{p}_s$ and the true, unknown, measure p, $F^{-1}_{|\mathcal{S}|-1}$ is the inverse of a Chi-squared distribution with $|\mathcal{S}| - 1$ degrees of freedom, and N_s is the number of historical samples used to estimate $\hat{p}_s$. Note that $\beta(s)$ decreases in the confidence parameter ω. For small ω, $\mathcal{P}(s)$ is closer to the point estimate $\hat{p}_s$. As ω increases, $\mathcal{P}(s)$ grows and nature has more choices.

The next next result has the following interpretation: As the level of ambiguity increases, therapy is initiated sooner – in additional 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. In instances where both the robust and the myopic models happen to have optimal control-limit policies, the robust control limit is lower than (or equal to) the myopic control limit. That is, patients will initiate therapy in healthier states 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 $d_i^*(s) = T$.*

3. Numerical Studies and Summary

The data for our studies are from two sources. For estimating pre-transplant transition probabilities, the data set is from the Thomas E. Starzl Transplantation Institute at the University of Pittsburgh Medical Center (UPMC). 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 by Alagoz et al. (2005). We considered $r(s, T)$ to be post-transplant life expectancy, which depends on several factors, including the type of ESLD, the age, race and gender of the patient, and the possible presence of some viruses. The post-transplant rewards also depend on some qualities of the donor organ. Post-transplant rewards are estimated using data from the United Network of Organ Sharing (UNOS).

Since the true underlying transition probabilities are unknown, we had to compare the performance of the robust policies to the myopic policies in simulation studies. The results are not reported here due to space limitations. Our general conclusion from those studies is

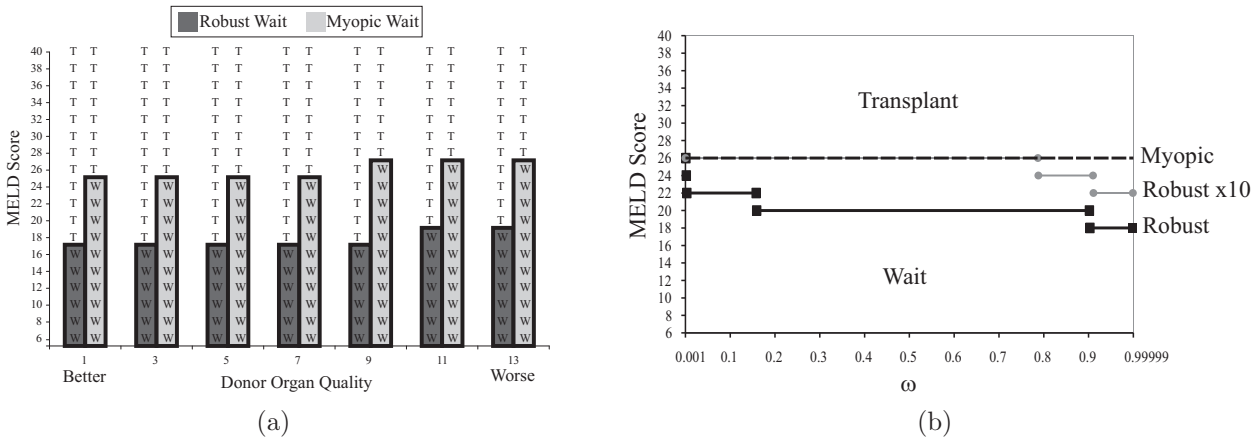


Figure 2: (a) Optimal myopic and robust policies for multiple problem instances for a single patient but with different donor organ qualities; $\omega = .95$. (b) Optimal myopic and robust control limits as a function of ω; organ quality is 7. "Robust x10" is the robust control limit assuming that estimates were instead made with 10 times the amount of data.

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

A question of interest is what the implied ω actually is for transplants that have occurred in practice. We considered a cohort of females in two ESLD disease groups. In total, there were 296 transplants in our sample. Some of these patients may have also been on the cadaveric organ waiting list, which should have only given them an incentive to wait longer. In the end, though, they underwent a living-donor transplantation. Of the 296 patients, only 3 (1.01%) waiting longer than suggested by the optimal myopic solutions. The distribution for the implied ω is presented in Table 1. Note that 72.6% of the patients have an implied ω less that 0.995. For one who believes our models, the majority of transplants occurred too soon.

In summary, our two main findings are:

1. The robust solutions suggest transplanting for a range of health states in which patients are healthier than under the myopic control limit. The robust solutions could be closer to what some patients (and physicians operating on their behalf) are willing to do in practice.
2. A large majority of the patients in our sample transplanted even sooner than what the most conservative robust DP solutions suggest.

ω Range			Frequency
0	–	.005	2.70%
.005	–	.01	0%
.01	–	.05	0%
.05	–	.1	0%
.1	–	.3	2.03%
.3	–	.5	0.68%
.5	–	.7	4.05%
.7	–	.9	6.76%
.9	–	.95	1.35%
.95	–	.99	8.78%
.99	–	.995	0%
.995	<		71.64%

Table 1: Implied ω in practice.

Acknowledgments

This work is partially supported by the National Science Foundation under Grant CMMI-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

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. *Medical Decision 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. *Management Sci.* **50**(10) 1420–1430.

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

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

Kaufman, D. L., A. J. Schaefer. 2011. Robust modified policy iteration. Submitted for publication. Available at http://webster.engr.pitt.edu/~schaefer/papers.htm.

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.

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

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

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