Data-Driven Management of Post-Transplant Medications: An APOMDP Approach

Faculty Research Working Paper Series

Alireza Boloori
Arizona State University

Soroush Saghafian
Harvard Kennedy School

Harini A. A. Chakkera
Mayo Clinic Hospital

Curtiss B. Cook
Mayo Clinic Hospital

August 2017

RWP17-036

Data-Driven Management of Post-Transplant Medications: An APOMDP Approach

Alireza Boloori
Industrial Engineering, School of Computing, Informatics and Decision Systems Engineering, Arizona State University, Tempe, AZ 85281, alireza.boloori@asu.edu

Soroush Saghafian
Harvard Kennedy School, Harvard University, Cambridge, MA 02138, Soroush_Saghafian@hks.harvard.edu

Harini A. Chakkera, Curtiss B. Cook
Divisions of Transplantation and Endocrinology, Mayo Clinic Hospital, Phoenix, AZ 85054, {Harini.Chakkera@mayo.edu, Cook.Curtiss@mayo.edu}

Organ-transplanted patients typically receive high amounts of immunosuppressive drugs (e.g., tacrolimus) as a mechanism to reduce their risk of organ rejection. However, due to the diabetogenic effect of these drugs, this practice exposes them to greater risk of New-Onset Diabetes After Transplant (NODAT), and hence, becoming insulin-dependent. This common conundrum of balancing the risk of organ rejection versus that of NODAT is further complicated due to various factors that create ambiguity in quantifying risks: (1) false-positive and false-negative errors of medical tests, (2) inevitable estimation errors when data sets are used, (3) variability among physicians' attitudes towards ambiguous outcomes, and (4) dynamic and patient risk-profile dependent progression of health conditions. To address these challenges, we propose an ambiguous partially observable Markov decision process (APOMDP) framework, where dynamic optimization with respect to a "cloud" of possible models allows us to make decisions that are robust to misspecifications of risks. We first provide various structural results that facilitate characterizing the optimal policy. Using a clinical data set, we then compare the optimal policy to the current practice as well as some other benchmarks, and discuss various implications for both policy makers and physicians. In particular, our results show that substantial improvements are achievable in two important dimensions: (a) the quality-adjusted life expectancy (QALE) of patients, and (b) medical expenditures.[1]

Key words: Ambiguous POMDP; cloud of models; conservatism level; kidney transplant; immunosuppressive drug; diabetes medication
History: Version: July 23, 2017

1. Introduction

As reported by the United Network of Organ Sharing, 17,878 kidney transplantations were conducted in the U.S. in 2015 (102,082 cases since 2010) (UNOS 2016). According to the Organ Procurement and Transplantation Network (OPTN), the average cumulative probability of 1 to 10-year organ rejection after kidney transplantation is estimated to be 6.35% to 48.7% (OPTN 2011). To reduce the risk of organ rejection post-transplant, physicians typically use an intensive amount of immunosuppressive (a.k.a. anti-rejection) drugs such as tacrolimus, sirolimus, and corticosteroids. However, due to the well-known *diabetogenic effect*, excessive exposure to these drugs may induce New Onset

[1] This work was partially supported by the National Science Foundation through Grant CMMI-1562645 (PI: Saghafian).

Diabetes After Transplantation (NODAT) which refers to incidence of diabetes in a patient with no history of diabetes prior to transplantation (Chakkera et al. 2009).

To clearly illustrate this point, we use a data set of patients who had kidney transplant surgery at our partner hospital, Mayo Clinic, between 1999 and 2006. Figure 1 depicts the empirical cumulative distribution functions (c.d.f.s) of blood glucose level (measured by the HbA1c test) right before and one month after transplantation for patients who had no prior history of diabetes. As can be seen, more than 80% (20%) of patients who undergo transplantation are in danger of becoming pre-diabetic (diabetic), mainly because of intensive amount of immunosuppressive drugs used in practice. Considering the total number of transplantations carried out worldwide, this can account for more than 90,000 new patients per year who are in danger of elevated blood glucose levels.

Elevated blood glucose levels, in turn, increase the risk of organ rejection and may result in re-transplantation, which is a costly medical operation.[2] Although physicians attempt to control the risk of elevated blood glucose levels by putting the patient on diabetes medications (e.g., insulin), this should be coordinated with the intensity of the immunosuppressive drug(s) used, because unnecessary use of such medications is harmful (Kromann et al. 1981). Despite this conundrum faced by physicians, there is currently no clear guideline on how these medications should be simultaneously managed. Our goal in this paper is to address this deficit while taking into account the following issues:

Measurement Errors. Blood glucose levels are measured by test procedures such as *Fasting Plasma Glucose* (FPG) and *Hemoglobin A1c* (HbA1c), which have a wide range of false-positive and false-negative errors (Bennett et al. 2007, Van'T Riet et al. 2010). In addition, the concentration of immunosuppressive drugs is measured in practice through test procedures such as *Abbott Architect* and *Magnetic Immunoassay*, which are similarly error-prone (Bazin et al. 2010, Taguchi et al. 2013).

Estimation Errors. Estimating various parameters (e.g., the probabilistic consequences of various medications on a patient's future health) from data sets is typically subject to errors for a variety of reasons including lack of comprehensive data and data entry errors among others. Furthermore, medication strategies are typically optimized with respect to such estimated parameters. Thus, unless carefully adjusted, they may not represent patients' best medical interest.

Behavioral Attitudes. Physicians have a range of different behavioral attitudes in coordinating the use of immunosuppressive drugs and diabetes medications, resulting in considerable variations among them. In particular, when faced with ambiguity regarding unknown consequences of medication regimens, some show ambiguity-seeking attitudes (low conservatism) and some show ambiguity-aversion attitudes (high conservatism).

[2] Based on billed charges for kidney, liver, pancreas, and heart transplants, the average direct cost for an organ transplant in the U.S. is estimated to be $531,775 (Bentley and Hanson 2011).

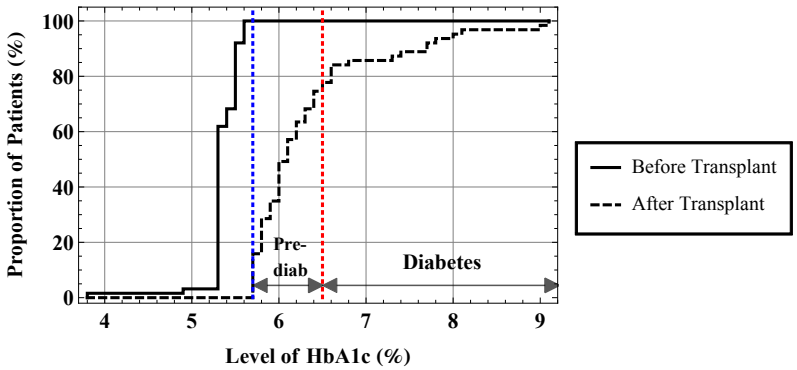

Figure 1 **Empirical c.d.f.s of patients' Hemoglobin A1c (HbA1c) level in our data set: an illustration of the diabetogenic effect of immunosuppressive drugs. The left (right) vertical dotted line shows the threshold for pre-diabetes (diabetes) as defined by American Diabetes Association (ADA 2012).**

Static and Dynamic Risk Factors. Both static/time-invariant (e.g., race and gender) and dynamic/time-variant (e.g., blood pressure and body mass index) risk factors play an important role in effective coordination of post-transplant medication regimens, because they both affect organ rejection and/or diabetes complications.

Ignoring any of the above-mentioned issues can yield suboptimal medication strategies that may harm patients. Thus, in finding a solution for the conundrum discussed earlier, one also needs an approach that allows addressing such issues in an integrated way. To this end, we use an approach termed Ambiguous Partially Observable Markov Decision Process (APOMDP)—an extension of the traditional POMDP approach recently proposed by Saghafian (2017). Utilizing the APOMDP approach allows us to find a dynamically optimal way of coordinating immunosuppressive and diabetes medications during each patient visit while accounting for (1) imperfect state information about the patient's health (caused by measurement errors), (2) model misspecifications (caused by estimation errors), (3) a range of attitudes towards model misspecifications (caused by physicians' behavioral attitudes), and (4) several risk factors (age, gender, race, diabetes history, body mass index (BMI), blood pressure, total cholesterol, high-density lipoprotein (HDL) cholesterol, low-density lipoprotein (LDL) cholesterol, triglyceride, and uric acid). This approach enables us to provide the first analytical study (to the best of our knowledge) that (a) simultaneously analyzes two medical conditions with conflicting risks (i.e., post-transplant organ rejection versus NODAT), and (b) integrates such risks with both static and dynamic patient-dependent characteristics.

Our study contributes to both Operations Research (OR)/Management Science (MS) and medical literature. Our contributions to OR/MS are two-fold:

(1) We demonstrate the use of the APOMDP approach to make *robust* dynamic decisions under both imperfect state information and model misspecifications. Since both imperfect state information and model misspecifications are inevitable in many OR applications including those in medical decision-making, our work sheds light on the advantages of a widely applicable new tool.

Specifically, our approach empowers a decision maker who is facing hidden states to dynamically optimize actions under a variety of possible models (a "cloud" of models as opposed to a single model), and thereby gain robustness to potential model misspecifications without the need to perform sensitivity analyses.

(2) We develop a closed-form expression for the optimal value function (based on a piecewise-linearity and convexity property) to solve our APOMDP formulation optimally for the medical problem under consideration. We also establish other structural properties including (a) an analytical link between a decision maker's ambiguity attitude and the intensity of optimal medication regimens, (b) monotonicity results for the optimal value function and the medication policy, and (c) a lower bound for the optimal value function.

Our study also contributes to the medical literature by presenting various new clinically relevant findings:

(1) We calibrate our APOMDP model based on a clinical data set that we have collected from our partner hospital (Mayo Clinic). Utilizing this data set, we first estimate unobservable disease progression rates, inaccuracies of medical test procedures, and reward-related parameters (e.g., quality of life and residual life expectancy) based on several risk factors. Using these estimations along with our APOMDP approach, we then generate risk-specific medication guidelines for use in practice.

(2) For patients with non-White race, no diabetes history, low-risk levels of cholesterol, HDL, LDL, triglyceride, and uric acid, we find that, under the optimal medication policy, a more conservative physician typically prescribes more intensive regimens of immunosuppressive drugs than a less conservative one. This implies that, for these risk factors, a physician should be more concerned about the risk of organ rejection than the potential risk of NODAT. On the contrary, for patients with male gender, diabetes history, hypertension, and high-risk levels of cholesterol, HDL, and LDL, the result is reversed: for such patients, the physician should be more concerned about the risk associated with diabetes complications than the risk of organ rejection.

(3) Variations in physicians' attitude toward ambiguity will not show a homogeneous pattern with respect to the intensity of the drugs used, if they follow the optimal policy. Thus, the intensification of drug regimens observed in the current practice should not be attributed merely to physicians' behavior toward ambiguity. Instead, our results suggest that lack of adherence to (or knowledge of) the optimal medications might be the main cause of using intensive regimens of drugs in the current practice.

(4) Although the extant medical literature has focused on age and race as predictors of tacrolimus dose variability, our study sheds light on the customization of this monitoring based on risk factors such as age, gender, race, diabetes history, BMI, blood pressure, total cholesterol, and LDL. Specifically, we find that these risk factors make patients significantly vulnerable to the risk of organ rejection.

(5) The diabetogenic effect of tacrolimus—a main immunosuppressive drug—is more likely to influence patients with age≥50, male gender, White race, diabetes history, hypertension, high total cholesterol, low HDL, high triglyceride, and high uric acid. This implies that, when taking high-dose tacrolimus, such patients typically become more dependent on diabetes medications than others.

(6) We compare the performance of the optimal medication policies we obtain from our APOMDP approach with (a) the current medical practice, and (b) medication policies that arise when we use a traditional POMDP approach. We take into account performance measures such as quality-adjusted life expectancy (QALE), medical expenditure, and the intensity of prescribed medications. Our comparison generates the following insights:

- Compared to the current medical practice and based on different risk factors, the medication strategies obtained by our APOMDP approach can improve (per patient per year) the average (a) QALE between 0.66% and 5.16%, and (b) medical expenditures between 4.55% and 11.52%. They also prescribe use of high-dose tacrolimus between 0.95 and 2.75 fewer times per patient per year compared to the current medical practice.
- Cohorts of patients formed by age, non-White race, diabetes history, blood pressure, and HDL cholesterol will benefit the most from our proposed medication strategies compared to the current practice. This is because, for these patients, our proposed medication strategies yield the highest improvements in QALE while incuring the least amount of medical expenditure, providing more *cost-effective* ways of managing medications.
- The prevalent practice of following recommendations derived from approaches that ignore inevitable parameter ambiguities can have dire consequences. For instance, we find that deriving optimal strategies via a traditional POMDP instead of using our APOMDP approach may cause a patient to lose between 1.04 and 4.68 weeks of QALE over the course of first year post-transplant, while imposing between $31 and $214 more medical expenditures per patient to the system during the same time.

In closing this section, we note that although our focus in this paper is on patients who have kidney transplantation, our approach can be extended to liver, heart, and pancreas transplantations.[3] This, in turn, can help to create the first holistic data-driven decision support system (DSS) as a tool to assist physicians in their post-transplant medications management decisions (see Figure 2).

The rest of this paper is organized as follows. In §2, we provide a brief literature review. In §3, we present our APOMDP approach, and in §4, we demonstrate some of its theoretical/structural properties. Our numerical study including our clinical data set and parameter estimations as well as the resulted findings are described in §5. Finally, we conclude the paper in §6, and discuss some avenues for future research.

[3] These four organs account for 88.63% of solid organ transplantations in the U.S. (Bentley and Hanson 2011).

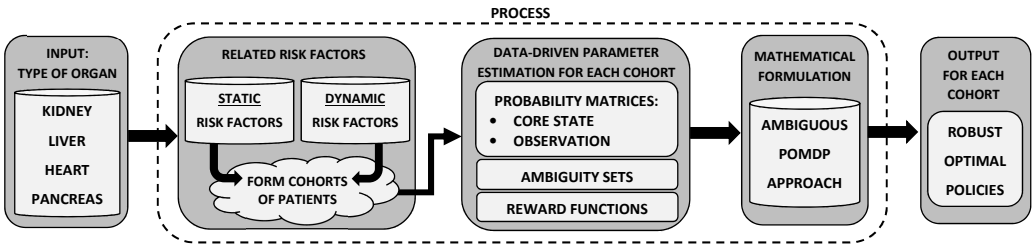

Figure 2 **The multiple-organ data-driven decision support system for post-transplant medications management**

2. Related Studies

We divide the related studies into five categories, and describe each separately below.

Studies on Medical Decision-Making for Diabetes. The main body of literature analyzing diabetes from a decision-making perspective uses Markov Decision Process (MDP) models to focus on optimal initiation time of statin (see, e.g., Denton et al. (2009)), adherence to statin therapy (see, e.g., Mason et al. (2012)), and optimal interval for other diabetes medications (see, e.g., Mason et al. (2014)); see Kurt (2012) for a review. Unlike this stream of research, we (1) address the management of diabetes medications in the presence of an opposing medication (i.e., an immunosuppressive drug), and (2) consider partial observability of health states that arises due to the inevitable measurement errors in medical tests (e.g., FPG and HbA1c). Furthermore, the above studies require incorporating dynamic risk factors as part of the state space definition, which may aggravate the so-called "curse of dimensionality." Instead, our proposed approach directly incorporates such factors into optimal medication strategies.

OR/MS Studies on Pre-Transplant Period. The majority of OR/MS studies on transplantation focus on the pre-transplant period, and typically study mechanisms to facilitate a better match between supply and demand of organs (see, e.g., Alagoz et al. (2004), Su and Zenios (2005), Bertsimas et al. (2013), Sandikci et al. (2013), and Ata et al. (2016)). To the best of our knowledge, our paper is among the first in the OR/MS literature to consider post-transplantation decisions.

Studies on POMDP Applications in Healthcare. In the medical decision-making field, POMDP models have been widely used, but mainly for cancer screening research. Examples include mammography screening in breast cancer (see, e.g., Maillart et al. (2008), and Ayer et al. (2012, 2015)), screening in prostate cancer (see, e.g., Zhang (2011)), and colonoscopy screening in colorectal cancer (see, e.g., Erenay et al. (2014)). Compared to this stream, our proposed APOMDP approach (1) provides optimal policies that are robust to model misspecifications, (2) incorporates physicians' behavorial attitudes toward model misspecifications, and (3) is customized with eleven different risk factors (most of which are time-variant). From the medical decision-making perspective, the latter is particularly a significant improvement, since age and history of screening/treatment are the only dynamic risk factors that have been considered thus far in the extant literature.

Studies on Robust Dynamic Decision-Making. The use of robust dynamic decision-making in healthcare applications can be found in studies such as Kaufman et al. (2011), Zhang (2014),

and Goh et al. (2015), and the references therein. Among theoretical studies addressing robustness in dynamic decision-making, we refer interested readers to those solving MDPs with respect to a worst-case scenario (i.e., those that utilize a *max-min* approach) within the set of possible transition probabilities (see, e.g., Iyengar (2005), Nilim and El Ghaoui (2005), and Xu and Mannor (2012)). However, as noted by Delage and Mannor (2010), generated policies under a max-min approach are often too conservative. To address this, Saghafian (2017) integrates the so-called α-maxmin expected utility (α-MEU) preferences with a sequential decision-making approach, where a controller makes decisions based on a weighted average of both the worst and the best possible outcomes. Moreover, unlike the above-mentioned literature on robust MDPs, the APOMDP approach proposed in Saghafian (2017) allows for making robust decisions under partial observability of system states. This is an important advantage for various applications including our focus in this paper where measurement errors (e.g., due to false positive and false negative errors of medical tests) are inevitable.

Studies from the Medical Literature. Finally, we note that our work is also related to three streams in the medical literature: (1) incorporating the measurement errors of medical tests in decision-making for medication regimens (see, e.g., Bennett et al. (2007) and Van'T Riet et al. (2010)), (2) analyzing the diabetogenic effect of immunosuppressive drugs (see, e.g., Chakkera et al. (2009) and Boloori et al. (2015)), and (3) customizing tacrolimus dose variability based on different risk factors (see, e.g., Yasuda et al. (2008)). Utilizing the APOMDP approach along with our own clinical data set, we contribute to all of these three streams.

3. The Ambiguous POMDP Approach

3.1. The Problem Setting

To gain insights into effective post-transplant medication management strategies, we consider tacrolimus as the primary immunosuppressive drug. We do so because (1) it has been shown that tacrolimus is superior to other immunosuppressive drugs (e.g., cyclosporine) in preventing organ rejection for kidney transplantations (see, e.g., Bowman and Brennan (2008)), and (2) tacrolimus is the main immunosuppressive drug used in our partner hospital: based on our data set, 95% of patients are put on tacrolimus. We also observe from our data set that 94% of patients who are put on diabetes medications post-transplant are prescribed by a fixed-dosage insulin. Therefore, we consider insulin as the main diabetes medication, and to be consistent with practice, assume it can be prescribed only in a fixed dosage (see also Denton et al. (2009), Mason et al. (2012), and Mason et al. (2014) for a similar assumption).

Unlike insulin which is prescribed in a fixed dosage, physicians prescribe tacrolimus based on its lowest concentration in the body of patient known as *trough level* or C_0. A lower (higher) C_0 is known to be associated with a higher (lower) risk of organ rejection (see, e.g., Staatz et al. (2001)). The target therapeutic range of C_0 at our partner hospital is 10-12 mg/dL (month 1 post-transplant),

8-10 mg/dL (month 4 post-transplant), and 6-8 mg/dL (month 12 post-transplant). Thus, we lable any $C_0 \in [4,8)$, $[8,10)$, $[10,14]$ mg/dL as "low," "medium," and "high," respectively. Similarly, we use lables "low," "medium," and "high" to refer to tacrolimus prescription dosages [0.05,0.10], (0.10,0.20], and (0.20,0.25] mg/kg/day, respectively.[4] From the diabetes perspective, blood glucose levels are measured by FPG and HbA1c tests, where a patient with FPG$\geq$126 (100 $\leq$FPG$<$ 126) mg/dL or HbA1c$\geq$6.5% (5.7 $\leq$HbA1c$<$6.5%) is labled as diabetic (pre-diabetic), whereas FPG$<$100 mg/dL or HbA1c$<$5.7% is labled as healthy (ADA 2012).

Based on this premise, we use a discrete-time, finite-horizon ambiguous POMDP (APOMDP) approach, in which, at each patient visit, a decision maker (DM hereafter) who is typically a physician measures the patient's (1) C_0, and (2) blood glucose level. Then, after evaluating whether the patient has a low, medium or high C_0, and whether s/he is diabetic, pre-diabetic, or healthy, the DM needs to make two decisions: (a) whether to use a low, medium or high dosage of tacrolimus, and (b) whether or not to put the patient on insulin. As noted earlier, these decisions need to be made jointly and in an orchestrated way, because of the interactions between tacrolimus and insulin as well as their joint effect on the patient's health state. If prescribed, any medication will be used over the course of one month until the patient's next visit. As a result, the patient's health state with respect to both his/her C_0 level and diabetes condition may move to a new state in the next visit, and this routine continues throughout the planning horizon.

We use an APOMDP approach described below to determine optimal decisions that maximize QALE of a patient with respect to risks of organ rejetion and NODAT complications. Importantly, using an APOMDP instead of a traditional POMDP allows us to (1) gain robustness to estimation errors that arise from the current lack of data, wrong data entries, or differences in experts' opinions, and (2) incorporate the DM's attitude toward ambiguity.[5]

Finally, we use our setting to study unnecessary intensification of prescribed medications. We do so by comparing the effect of using (a) lower dosages of tacrolimus, and (b) insulin versus not using it.

3.2. The Elements of the APOMDP Approach

The elements of our APOMDP approach are as follows. All vectors are considered to be in a column format, and " $'$ " represents the matrix transpose operator.

[4] We do not consider continuous levels of tacrolimus or insulin. This requires using pharmacokinetics/pharmacodynamics (PK/PD) of these drugs as a set of differential equations that can be obtained based on compartmental models. However, lack of granular data necessitates resorting to a single-compartment model (Hu et al. 1996), and such a model is not yet sufficient for modeling PK/PD of tacrolimus and insulin properly (see, e.g., Lehmann and Deutsch (1992) and Jusko et al. (1995)).

[5] We will also show in §5.2.2 that the APOMDP approach results in more cost-effective policies than those obtained via a POMDP.

Decision epochs: Decision epochs correspond to a patient's visits and are denoted by $n = 1, 2, ..., N$, where n represents the number of months post-transplant. We consider one year post-transplant as our planning horizon ($N = 12$), because after the first year the chance of NODAT and the importance of coordinating immunosuppressive and diabetes medications drops drastically (see, e.g., Ghisdal et al. (2012)).

Core state space: $\mathscr{S} = \{\Delta, \nabla\} \cup \mathcal{S}$, where $\mathcal{S} = \{s_i, i = 1, 2, \ldots, 9\}$, and s_i's are described in Table 1. In addition, Δ and ∇ represents "death" and "organ rejection," respectively. We note that ∇ and Δ are fully observable and absorbing states: the decision process in our model ends if either of these two states is reached prior to the end of planning horizon.

Observation state space: $\mathscr{O} = \{\Delta, \nabla\} \cup \mathcal{O}$, where $\mathcal{O} = \{o_i, i = 1, 2, \ldots, 9\}$, and o_i is the observation made by the DM leading him to think that the patient is in the ith core state. For instance, o_1 is the observation that the patient is in s_1: medical tests suggest a low C_0 level ($C_0 \leq 8$ mg/dL) while having organ survival and diabetic conditions (FPG$\geq$126 mg/dL or HbA1c$\geq$6.5%).

Action space: $\mathcal{A} = \{a_i, i = 1, 2, \ldots, 6\}$, where a_i's are described in Table 2. Letting $a \preceq \widehat{a}$ represent the fact that $\widehat{a}$ is not more intensive than a, we have with respect to Table 2: $a_1 \preceq a_2 \preceq a_3$, $a_4 \preceq a_5 \preceq a_6$, $a_1 \preceq a_4$, $a_2 \preceq a_5$, $a_3 \preceq a_6$, and $a_1 \preceq a_6$. Thus, a_1 (a_6) corresponds to administrating the most (least) intensive medication regimen.

Ambiguity set ("cloud" of models): $\mathcal{M} = \{m_1, m_2, ..., m_K\}$, where K is the number of models in the "cloud." Each model in $\mathcal{M}$ represents a different estimation for core state and observation transition probability matrices. In §5.1, we describe how we have used a clinical data set obtained from our partner hospital, Mayo Clinic, to construct this cloud of models.

Core state transition probability: $\mathbf{P}_m = \{\mathbf{P}_m^a : a \in \mathcal{A}\}$, where for each $a \in \mathcal{A}$, $\mathbf{P}_m^a = [p_m^a(j|i)]_{i,j \in \mathscr{S}}$, and $p_m^a(j|i) = Pr\{j|i, a, m\}$ is the probability of moving from state i to state j when taking action a under model $m \in \mathcal{M}$.[6]

Observation probability: $\mathbf{Q}_m = \{\mathbf{Q}_m^a : a \in \mathcal{A}\}$, where for each $a \in \mathcal{A}$, $\mathbf{Q}_m^a = [q_m^a(o|j)]_{j \in \mathscr{S}, o \in \mathscr{O}}$, and $q_m^a(o|j) = Pr\{o|j, a, m\}$ is the probability of observing o under model m and action a when being at core state j.

Information space: $\Pi = \left\{\boldsymbol{\pi} = [\pi_i]_{i \in \mathscr{S}} \in \mathbb{R}^{|\mathscr{S}|} : \sum_{i=1}^{|\mathscr{S}|} \pi_i = 1, \pi_1, \pi_2 \in \{0, 1\}, \pi_3, \ldots, \pi_{11} \in [0, 1]\right\}$, where $\boldsymbol{\pi}$ is an information vector over the state space $\mathscr{S}$. Since Δ (death) and ∇ (organ rejection) are fully observable states, $\boldsymbol{\pi} = [1, \ldots, 0]'$ and $\boldsymbol{\pi} = [0, 1, \ldots, 0]'$ represent death and alive with organ rejection, respectively.

Belief space: In order to distinguish between fully and partially observable states, we define a belief vector $\mathbf{b}$ such that, for any $\boldsymbol{\pi} \neq [1, \ldots, 0]'$ or $\boldsymbol{\pi} \neq [0, 1, \ldots, 0]'$, $\mathbf{b} = [b_i]_{i \in \mathscr{S}} = \boldsymbol{\pi}$ (i.e., DM's belief

[6] A "rectangularity" property for transition probabilities serves as a sufficient condition for "dynamic consistency" in conventional dynamic robust optimization models (e.g., robust MDPs). Although model ambiguity in APOMDPs might compromise the dynamic consistency in general, Saghafian (2017) discuss conditions under which this property can be preserved.

Table 1 Description of core health states

State	Health Condition (Transplant)	Health Condition (Diabetes)
s_1	Organ survival and low C_0	Diabetes (type II)
s_2	Organ survival and medium C_0	Diabetes (type II)
s_3	Organ survival and high C_0	Diabetes (type II)
s_4	Organ survival and low C_0	Pre-diabetes
s_5	Organ survival and medium C_0	Pre-diabetes
s_6	Organ survival and high C_0	Pre-diabetes
s_7	Organ survival and low C_0	Healthy
s_8	Organ survival and medium C_0	Healthy
s_9	Organ survival and high C_0	Healthy

Table 2 Description of actions

Action	Prescription (Transplant)	Prescription (Diabetes)
a_1	Use tacrolimus (with high dose)	Use insulin
a_2	Use tacrolimus (with medium dose)	Use insulin
a_3	Use tacrolimus (with low dose)	Use insulin
a_4	Use tacrolimus (with high dose)	Do not use insulin
a_5	Use tacrolimus (with medium dose)	Do not use insulin
a_6	Use tacrolimus (with low dose)	Do not use insulin

about C_0 and blood glucose levels in an alive patient without an organ rejection). We also let Π_{PO} be the set of all such belief vectors (PO: partially observable).

We use the *Bayes' Rule* in a matrix format to update the elements of the belief vector $\mathbf{b}$ under a model m when action a is taken and observation o is made:

$$B(\mathbf{b}, a, o, m) = \frac{(\mathbf{b}'\mathbf{P}_m^a \mathbf{Q}_m^{a,o})'}{Pr\{o|\mathbf{b}, a, m\}}, \tag{1}$$

where $B(\mathbf{b}, a, o, m) : \Pi_{PO} \times \mathcal{A} \times \mathscr{O} \times \mathcal{M} \to \Pi_{PO}$ is the belief updating operator, $\mathbf{Q}_m^{a,o}$ is the diagonal matrix formed by the column o of $\mathbf{Q}_m^a$, and

$$Pr\{o|\mathbf{b}, a, m\} = \sum_{i \in \mathscr{S}} b_i \sum_{j \in \mathscr{S}} p_m^a(j|i)\, q_m^a(o|j) \tag{2}$$

is the conditional probability that the DM will make observation o given the belief vector $\mathbf{b}$, action a, and model m.

Immediate reward: $\mathbf{r}_n(a) = [r_n(s,a) \geq 0]_{s \in \mathscr{S}}$ for $a \in \mathcal{A}$, where $r_n(s,a)$ is the QALE that a patient accrues when in state $s \in \mathcal{S}$ and taking action a in period $n < N$ (based on experiencing death, an organ rejection, or an organ survival while having different blood glucose levels). Note that a patient experiencing death does not gain any immediate reward (i.e., $r_n(\Delta, a) = 0$) and $0 \leq r_n(\nabla, a) \leq r_n(s,a)$ for all $a \in \mathcal{A}$ and $s \in \mathcal{S}$.

Lump-sum reward: $\mathbf{R}_n = [R_n(s) \geq 0]_{s \in \mathscr{S}}$, where $R_n(s)$ is a lump-sum reward (in QALE) gained by a patient whenever s/he leaves the decision process at state s. This can happen either (1) at the end of the planning horizon ($n = N$), when this value serves as a terminal reward that the patient accrues for his/her remaining lifetime, or (2) during the planning horizon ($n < N$), if s/he experiences a death or an organ rejection, where $R_n(\Delta) = 0$ and $0 \leq R_n(\nabla) \leq R_n(s)$ for all $s \in \mathcal{S}$.

Ambiguity attitude set: $\Lambda = \{\lambda : 0 \leq \lambda \leq 1\}$, where λ represents the DM's *conservatism level*, and captures his/her attitude towards ambiguity.

Discount factor: $\beta \in [0,1]$, which allows us to obtain the present value of QALE gained in future.

Using the elements of the APOMDP approach described above, we now present its optimality equation. For the information vector $\boldsymbol{\pi}$, DM's conservatism level λ, and any period $n \leq N$, we have:

$$V_n(\boldsymbol{\pi}, \lambda) = \begin{cases} R_n(\Delta), & \text{if } \boldsymbol{\pi} = [1, \ldots, 0]', \\ R_n(\nabla), & \text{if } \boldsymbol{\pi} = [0, 1, \ldots, 0]', \\ V_n(\mathbf{b}, \lambda), & \text{otherwise,} \end{cases} \tag{3}$$

where

$$V_n(\mathbf{b}, \lambda) = \begin{cases} \mathbf{b}'\mathbf{R}_N, & \text{if } n = N, \\ \max\limits_{a \in \mathcal{A}} \{U_n(\mathbf{b}, a, \lambda)\}, & \text{if } n < N. \end{cases} \tag{4}$$

In (4), the utility function $U_n(\mathbf{b}, a, \lambda)$ is defined as:

$$U_n(\mathbf{b}, a, \lambda) = \mathbf{b}'\mathbf{r}_n(a) + \lambda \min_{m \in \mathcal{M}} \Big\{ H_n(\mathbf{b}, a, m, \lambda) \Big\} + (1 - \lambda) \max_{m \in \mathcal{M}} \Big\{ H_n(\mathbf{b}, a, m, \lambda) \Big\}, \tag{5}$$

where

$$H_n(\mathbf{b}, a, m, \lambda) = \beta \sum_{o \in \mathcal{O}} Pr\{o|\mathbf{b}, a, m\} V_{n+1}\Big(B(\mathbf{b}, a, o, m), \lambda\Big). \tag{6}$$

The first term in the RHS of (5) represents the expected current "reward" (in QALE) in period n when the belief vector is $\mathbf{b}$, the action is a, the model is m, and the DM's conservatism level is λ. The other terms in the RHS of (5) denote the expected "reward-to-go" for period n, which is calculated as the weighted average of the worst and the best possible expected rewards that can be obtained in future. In (5), as λ increases (decreases), the utility function becomes more (less) dependent on the worst total "reward" that can be achieved in the "cloud" of models. Thus, a higher (lower) λ represents the attitude of a more (less) pessimistic/conservative DM. By varying λ, our framework allows us to capture the behavioral attitudes of physicians, and evaluate their effects on the intensity of medications administered.

Finally, we define the worst (best) model in period n as the minimizer (maximizer) in (6):

$$\underline{m}_n(\mathbf{b}, a, \lambda) = \arg\min_{m \in \mathcal{M}} \Big\{ H_n(\mathbf{b}, a, m, \lambda) \Big\}, \tag{7a}$$

$$\overline{m}_n(\mathbf{b}, a, \lambda) = \arg\max_{m \in \mathcal{M}} \Big\{ H_n(\mathbf{b}, a, m, \lambda) \Big\}, \tag{7b}$$

and, for the ease of notation, we may refer to them as $\underline{m}$ and $\overline{m}$, respectively.

4. Structural Properties

In this section, we establish some structural properties that enable us to solve and analyze our APOMDP approach for the simultaneous management of post-transplant medications. To this end, we start by adopting some of the basic structural properties established for the general class of APOMDPs in Saghafian (2017). However, we also take advantage of the specific structure of our medication management problem, and derive various other results that are particularly useful in our setting.

Piecewise-Linearity and Convexity of Value Function. Unlike traditional POMDPs, it is known that the value function in an APOMDP is not necessarily piecewise-linear and convex (PLC) in the belief vector (Saghafian 2017). This may prevent us from using efficient solution algorithms (similar to those used for POMDPs, since many of them rely on the PLC property of the value function). Thus, to guarantee the PLC property for the value function in our problem, we resort to the definition of a *belief-independent worst-case* (BIWC) member in the cloud of models $\mathcal{M}$:

DEFINITION 1 (SAGHAFIAN 2017). $\underline{m}_n(\mathbf{b}, a, \lambda) \in \mathcal{M}$ defined in (7a) is said to be a BIWC member of $\mathcal{M}$, if it is constant in the belief vector $\mathbf{b}$.

This implies that the model under which a patient receives the least reward (in QALE) is independent of the DM's belief about the patient's health state: irrespective of the DM's belief about a patient's health state, there exists a same set of transition and observation matrices that yield the least total reward (in QALE). The importance of a BIWC member is that, if such a model exists in $\mathcal{M}$, then the optimal value function is PLC in the belief vector $\mathbf{b}$ (see Proposition 2 in Saghafian (2017)), and hence, can be written as:

$$V_n(\mathbf{b}, \lambda) = \max_{\boldsymbol{\psi} \in \Psi_{n,\lambda}} \{\mathbf{b}' \boldsymbol{\psi}\} \quad \forall\, \mathbf{b} \in \Pi_{PO},\, \forall\, \lambda \in \Lambda,\, \forall\, n \leq N, \tag{8}$$

where $\Psi_{n,\lambda}$ is some finite set. Equation (8) is analogous to the use of POMDPs proposed by Smallwood and Sondik (1973). The advantage of (8) is that, to characterize the value function, one only needs to characterize the set $\Psi_{n,\lambda}$.

Although the existence of a BIWC member in the cloud of models $\mathcal{M}$ can be a relatively restrictive assumption, we are able to provide a sufficient condition. We do so by benefiting from the notion of *model informativeness* (as a generalization of Blackwell ordering): if, under an action $a \in \mathcal{A}$, $\mathbf{P}^a_m \mathbf{Q}^a_m = \mathbf{P}^a_{\widehat{m}} \mathbf{Q}^a_{\widehat{m}} \mathbf{W}$ for some stochastic matrix $\mathbf{W}$, then model m is said to be *less informative* than model $\widehat{m}$.[7] It follows that if one model is less informative than the others, then it is a BIWC member in $\mathcal{M}$ (see Proposition 3 in Saghafian (2017)). In §5.1, we will use this fact along with our patient data set to discuss the existence of a BIWC member in our setting.

Assuming that $\mathcal{M}$ is such that it has a BIWC member, we now establish a closed-form analytical representation for the set of $\boldsymbol{\psi}$-vectors, $\Psi_{n,\lambda}$. This, together with (8), enables us to characterize and solve the optimal value function in our problem. All the proofs are provided in Online Appendix A.

[7] For notational simplicity, we suppress the dependency on a, and write $m(a)$ as m for any $m \in \mathcal{M}$.

PROPOSITION 1 **(Representation of $\boldsymbol{\psi}$-Vectors)**. *Suppose $\mathcal{M}$ is such that it has a BIWC member. Let $\underline{m}$ and $\overline{m}$ be the BIWC member and the best-case model of $\mathcal{M}$ defined by (7a) and (7b), respectively. Then, the set of $\boldsymbol{\psi}$-vectors ($\Psi_{n,\lambda}$) in (8) can recursively be obtained as:*

$$\Psi_{N,\lambda} = \{\mathbf{R}_N\} \quad \forall\, \lambda \in \Lambda, \tag{9}$$

$$\Psi_{n,\lambda} = \left\{ \boldsymbol{\psi} \in \mathbb{R}^{|\mathscr{S}|} : \boldsymbol{\psi} = \mathbf{r}_n(a) + \lambda \left(\beta \sum_{o \in \mathscr{O}} \mathbf{P}^a_{\underline{m}} \mathbf{Q}^{a,o}_{\underline{m}} \boldsymbol{\psi}^{(\mathbf{b},a,o)}_{\underline{m}} \right) + (1-\lambda) \left(\beta \sum_{o \in \mathscr{O}} \mathbf{P}^a_{\overline{m}} \mathbf{Q}^{a,o}_{\overline{m}} \boldsymbol{\psi}^{(\mathbf{b},a,o)}_{\overline{m}} \right), \right.$$
$$\left. a \in \mathcal{A},\, \boldsymbol{\psi}^{(\mathbf{b},a,o)}_{\underline{m}}, \boldsymbol{\psi}^{(\mathbf{b},a,o)}_{\overline{m}} \in \Psi_{n+1,\lambda} \right\} \quad \forall\, \lambda \in \Lambda, \forall\, n < N, \tag{10}$$

where

$$\boldsymbol{\psi}^{(\mathbf{b},a,o)}_m = \underset{\boldsymbol{\psi} \in \Psi_{n+1,\lambda}}{\arg\max} \{\mathbf{b}' \mathbf{P}^a_m \mathbf{Q}^{a,o}_m \boldsymbol{\psi}\} \quad \forall\, \mathbf{b} \in \Pi_{PO},\, \forall\, a \in \mathcal{A},\, \forall\, m \in \mathcal{M},\, \forall\, o \in \mathscr{O}. \tag{11}$$

The characterization of the set of $\boldsymbol{\psi}$-vectors in Proposition 1 depends on identifying both models $\underline{m}$ and $\overline{m}$. Although $\underline{m}$ can be obtained in the ambiguity set $\mathcal{M}$ without the need for solving the APOMDP model (see our discussion above), $\overline{m}$ cannot be identified a priori. To address this, we present the following alternative approach for characterizing the $\boldsymbol{\psi}$-vectors.

PROPOSITION 2 **(Alternative Representation of $\boldsymbol{\psi}$-Vectors)**. *Let $\underline{m}$ be a BIWC member of $\mathcal{M}$, $\boldsymbol{\psi}^{(\mathbf{b},a,o)}_m$ be defined by (11), and let*

$$\widetilde{\Psi}_{N,\lambda} = \{\mathbf{R}_N\} \quad \forall\, \lambda \in \Lambda, \tag{12}$$

$$\widetilde{\Psi}_{n,\lambda} = \left\{ \widetilde{\boldsymbol{\psi}} \in \mathbb{R}^{|\mathscr{S}|} : \widetilde{\boldsymbol{\psi}} = \mathbf{r}_n(a) + \lambda \left(\beta \sum_{o \in \mathscr{O}} \mathbf{P}^a_{\underline{m}} \mathbf{Q}^{a,o}_{\underline{m}} \widetilde{\boldsymbol{\psi}}^{(\mathbf{b},a,o)}_{\underline{m}} \right) + (1-\lambda) \left(\beta \sum_{o \in \mathscr{O}} \mathbf{P}^a_m \mathbf{Q}^{a,o}_m \widetilde{\boldsymbol{\psi}}^{(\mathbf{b},a,o)}_m \right), \right.$$
$$\left. a \in \mathcal{A},\, m \in \mathcal{M} \setminus \{\underline{m}\},\, \widetilde{\boldsymbol{\psi}}^{(\mathbf{b},a,o)}_{\underline{m}}, \widetilde{\boldsymbol{\psi}}^{(\mathbf{b},a,o)}_m \in \widetilde{\Psi}_{n+1,\lambda} \right\} \quad \forall\, \lambda \in \Lambda, \forall\, n < N. \tag{13}$$

Then, $\Psi_{n,\lambda}$ in (10) can be obtained from $\widetilde{\Psi}_{n,\lambda}$ in (13) by applying the Monahan's algorithm (Monahan 1982).

Proposition 2 implies that, even if we consider all models in $\mathcal{M} \setminus \{\underline{m}\}$, by the Monahan's algorithm, we will be able to shrink the set of $\boldsymbol{\psi}$-vectors to those attributed only to $\underline{m}$ and $\overline{m}$. Propositions 1 and 2 enable us to characterize our APOMDP value function in an efficient way: we can first find the $\boldsymbol{\psi}$-vectors, and then replacing them in (8). This is especially advantageous when used in the large-scale data-driven DSS introduced in §1 (see Figure 2), since the DM needs to promptly decide about medication regimens during the patient's visit.

Effect of the DM's Conservatism Level on Drug Intensification. As noted earlier, the DM's conservatism (i.e., ambiguity attitudes) may affect the intensification of medications regimens. To study this phenomenon, we start by defining the notion of *decreasing differences* in the DM's utility function.

DEFINITION 2 (VAN ZANDT 2002). Let $a, \widehat{a} \in \mathcal{A}$ and $\lambda, \widehat{\lambda} \in \Lambda$ be such that $a \preceq \widehat{a}$ and $\lambda \leq \widehat{\lambda}$, where $\preceq$ represents the intensive ordering defined in §3.2. Then, the utility function $U_n(\mathbf{b}, a, \lambda)$ defined in (5) is said to have decreasing differences in (a, λ) if

$$U_n(\mathbf{b}, \widehat{a}, \widehat{\lambda}) - U_n(\mathbf{b}, a, \widehat{\lambda}) \leq U_n(\mathbf{b}, \widehat{a}, \lambda) - U_n(\mathbf{b}, a, \lambda). \tag{14}$$

Inequality (14) implies that the marginal utility achieved (in QALE) when the DM prescribes a less intensive medication regimen (i.e., $\widehat{a}$ instead of a) is lower when the DM has a higher conservatism level (i.e., $\widehat{\lambda}$ instead of λ). When the inequality in (14) is reversed, the utility function is said to have *increasing differences*.

THEOREM 1 **(Effect of λ on Drug Intensification)**. *Let* $a_n^*(\mathbf{b}, \lambda) = \arg\max_{a \in \mathcal{A}} \{U_n(\mathbf{b}, a, \lambda)\}$ *be the optimal medication action for period n. Then, if the utility function in (5) has*
(i) decreasing differences in (a, λ), $\lambda \leq \widehat{\lambda}$ *yields* $a_n^*(\mathbf{b}, \lambda) \succeq a_n^*(\mathbf{b}, \widehat{\lambda})$, *or*
(ii) increasing differences in (a, λ), $\lambda \leq \widehat{\lambda}$ *yields* $a_n^*(\mathbf{b}, \lambda) \preceq a_n^*(\mathbf{b}, \widehat{\lambda})$.

Theorem 1(i) provides a condition under which a higher level of conservatism results in a more intensive medication regimen. This can happen because the DM should be more concerned about the risk of organ rejection than potential diabetes complications, and hence, s/he should put more emphasis on using higher dosages of tacrolimus. Similarly, it can happen when the DM should try to diminish diabetes complications by putting the patient on insulin. On the other hand, part (ii) provides a condition under which a higher level of conservatism results in a less intensive medication regimen. This can happen because the DM should be more concerned about the risk of NODAT than an organ rejection, and because of its diabetogenic effect, s/he should avoid prescribing high dosages of tacrolimus. Similarly, it can happen because the DM should prefer not to use insulin. In our numerical experiments in §5.2 we will study whether and how the optimal medications regimens fit either of these two scenarios depending on the patient's risk factors.

To shed more light on Theorem 1, in Online Appendix C we present conditions under which the utility function defined in (5) has either increasing or decreasing differences (see Conditions 2, 3, and 4 in Online Appendix C as well as Lemma EC.1 in Online Appendix B). We also test their validity using our data set, and discuss whether and when such conditions hold.

Monotonicity of the Optimal Medication Policy. In Online Appendix B, we discuss the monotonicity of the value function in the belief vector. We do so by presenting sufficient conditions under which a better health belief implies a higher QALE. Here, we shed light on the monotonicity of the optimal policy. When the optimal policy is monotone, a simple *control-limit* policy becomes optimal, making the complex search for an optimal medication policy a much simpler task. Furthermore, as we will discuss, the control-limit policy provides an easy-to-implement guideline for the medical practice. To establish the monotonicity of the optimal policy, we need the following condition.

CONDITION 1. *For any model $m \in \mathcal{M}$ and any pair of actions a and $\widehat{a}$ such that $a \preceq \widehat{a}$: letting $\widetilde{\mathbf{P}}_m^a(\mathbf{b}) = \beta \sum_{o \in \mathscr{O}} \mathbf{P}_m^a \mathbf{Q}_m^{a,o} \boldsymbol{\psi}_m^{(\mathbf{b},a,o)}$ where $\boldsymbol{\psi}_m^{(\mathbf{b},a,o)}$ is defined in (11), vector $\widetilde{\mathbf{P}}_m^{\widehat{a}-a}(\mathbf{b}) = \widetilde{\mathbf{P}}_m^{\widehat{a}}(\mathbf{b}) - \widetilde{\mathbf{P}}_m^{a}(\mathbf{b})$ is nondecreasing in its elements.*

Condition 1 implies that, when taking a less intensive medication regimen compared to a more intensive one, the resulted difference between the reward to-go is nondecreasing along core health states. In Online Appendix C, we discuss this condition in more detail, and numerically evaluate its validity based on our clinical data set.

THEOREM 2 **(Monotone Optimal Medication Policy)**. *Suppose that the ambiguity set $\mathcal{M}$ has a BIWC member, and let $a_n^*(\mathbf{b},\lambda) = \arg\max_{a \in \mathcal{A}} \{U_n(\mathbf{b},a,\lambda)\}$ be the optimal medication action for period n. Then, under Condition 1, $a_n^*(\mathbf{b},\lambda)$ is TP_2-nondecreasing in $\mathbf{b}$ for all $\lambda \in \Lambda$ and $n < N$.*[8]

To understand Theorem 2, consider two patients, patients 1 and 2, where patient 2 is believed to be in a better health condition than patient 1. By Theorem 2, if the optimal medication policy for patient 1 is "tacrolimus: low dosage" and "no insulin," then patient 2 should be prescribed with the same regimen. On the other hand, if patient 2 is optimally prescribed by "tacrolimus: high dosage" and "insulin," then patient 1 must follow the same prescription. In general, Theorem 2 states that under some conditions, a patient who is in a better health condition should not be prescribed with a more intensive medication regimen than another patient who is in a worse condition. Hence, it transfers the typically complex search for an optimal medication policy to a much simpler monotonic search. In particular, under the condition of Theorem 2, the optimal policy will be of control-limit (or switching-curve to be more precise) type, where we only need to impose limits on the belief state, and change the action as we pass the limits. This provides an easy-to-implement guideline for use in practice. In our numerical results in §5.2.1, we show that these control-limit type policies are sustained for various cohorts of patients.

Bounds for the Value Function. For our numerical experiments, we solve our APOMDP model optimally based on Proposition 2. However, the time complexity of finding an optimal policy (at any period n and for any conservatism level λ) is $O\left(|\mathcal{M}||\mathcal{A}||\mathscr{S}||\Psi_{n+1,\lambda}|^{|\mathscr{O}|}\right)$ (see, e.g., Hauskrecht (2000)).[9] Although we alleviate this effect by implementing the Monahan's algorithm (Monahan 1982) to eliminate dominated $\boldsymbol{\psi}$-vectors, to further streamline computational burdens, we now develop a bound for the value function in (4). This, in turn, provides an efficient way of approximating the value function. We let $J_n(\mathbf{b},\lambda)$ be the approximate value function, and $a^{*,J}(\mathbf{b},\lambda)$ be its corresponding action (denoted by a^J for the ease of notation). In the optimal value function $V_n(\mathbf{b},\lambda)$, the DM returns the expected future reward based on his/her updated belief about the

[8] See, e.g., §1 of Karlin and Rinott (1980) or Definition 5 of Saghafian (2017) for the notion of TP_2-nondecreasing functions.

[9] See also Papadimitriou and Tsitsiklis (1987) for a discussion about the time and space complexities of (PO)MDPs.

patient's health state (i.e., expected reward-to-go). However, in the approximate value function $J_n(\mathbf{b},\lambda)$, the DM first obtains his/her expected belief (over all updated belief vectors), and then returns the reward based on the expected belief. Based on this premise, we have:

$$J_n(\mathbf{b},\lambda) = \mathbf{b}'\mathbf{r}_n(a^J) + \lambda \min_{m\in\mathcal{M}} \left\{\beta J_{n+1}\left(\mathbf{b}'\mathbf{P}_m^{a^J},\lambda\right)\right\} + (1-\lambda)\max_{m\in\mathcal{M}} \left\{\beta J_{n+1}\left(\mathbf{b}'\mathbf{P}_m^{a^J},\lambda\right)\right\}, \tag{15}$$

where we obtain $\mathbf{b}'\mathbf{P}_m^{a^J}$ from $\sum_{o\in\mathscr{O}} Pr\{o|\mathbf{b},a^J,m\}\, B(\mathbf{b},a^J,o,m)$ by following the Bayesian update in (1) and the fact that $\sum_{o\in\mathscr{O}} \mathbf{Q}_m^{a,o} = \mathbb{I}$, where $\mathbb{I}$ is an identity matrix.

The following proposition shows that, under some conditions, the optimal value function $V_n(\mathbf{b},\lambda)$ is tightly bounded from below by $J_n(\mathbf{b},\lambda)$.

PROPOSITION 3 **(Performance Bound)**. *Suppose (i) the ambiguity set $\mathcal{M}$ has a BIWC member, (ii) $|p_m^a(j|i) - p_{\widehat{m}}^a(j|i)| \le \eta$ for some $\eta \ge 0$ ($\forall\, a\in\mathcal{A}$, $\forall\, m,\widehat{m}\in\mathcal{M}$, $\forall\, i,j\in\mathscr{S}$), (iii) $\overline{r}$ is the maximum possible reward in each period, and (iv) based on Lemma EC.2 (Online Appendix B),*

$$\epsilon_{n+1} = \epsilon_q \sum_{l=0}^{N-n} \beta^l + \epsilon_r \beta^{N-n-1} \quad \forall\, n < N, \tag{16}$$

where ϵ_q and ϵ_r are upper bounds for the quality of life and lump-sum reward, respectively. Then, we have:

$$V_n(\mathbf{b},\lambda) - J_n(\mathbf{b},\lambda) \le \min\left\{\frac{\beta\,\eta\,\epsilon_{n+1}\,|\mathscr{S}|}{1-\beta}, \overline{r}\sum_{l=0}^{N-1}\beta^l\right\} \quad \forall\,\mathbf{b}\in\Pi_{PO}, \forall\,\lambda\in\Lambda, \forall\, n<N. \tag{17}$$

Proposition 3 implies that, when the DM follows a^J instead of the optimal policy a^*, the reward loss (in QALE) will be less than or equal to the RHS of (17). We note that, under the following conditions, $J_n(\mathbf{b},\lambda)$ converges to $V_n(\mathbf{b},\lambda)$, making the performance bound in (17) completely tight: (1) when transition probabilities under different models get closer to each other (i.e., different models in the cloud of models $\mathcal{M}$ become more similar), η approaches 0. (2) When $\beta\in[0,1)$ and the time horizon is lengthened, ϵ_{n+1} asymptotically approaches $\frac{\epsilon_q}{1-\beta}$, which, in turn, approaches 0 as a patient's health status gets aggravated (see our discussion for Lemma EC.2 in Online Appendix B). (3) When β approaches 0 (i.e., the DM decides upon medications regimens in a myopic approach). Also, when β approaches 1, the performance bound in (17) approaches $N\overline{r}$ which is small when N or $\overline{r}$ is small. In general, the bound in (17) is advantageous for the DM, because it enables him/her to obtain a near-optimal performance.

5. Numerical Experiments

In this section, we first explain our clinical data set, the main risk factors affecting NODAT patients, the estimation of the set of transition and observation probability matrices using our data set, and the mechanism used to estimate reward functions (in QALE). We then perform various numerical experiments using our data set, and shed light on various implications for researchers, practitioners, and policy makers.

5.1. Data and Parameter Estimation

The Clinical Data Set. The clinical data set we use in this study contains information of 407 patients who had a kidney transplant operation over a period of seven years (1999–2006) at our partner hospital. The information pertains each patient's visit at months 1, 4, and 12 post-transplant and includes the following attributes: (1) demographic (e.g., age, race, gender, etc.), (2) clinical (e.g., blood pressure, body mass index (BMI), cholesterol level, etc.), (3) immunosuppressive drugs (e.g., tacrolimus) and diabetes medications (e.g., insulin) prescribed by physicians, and (4) results of medical tests (FPG, HbA1c, and Architect). Further details about our data set can be found in our earlier study (Boloori et al. 2015).

Interpolation and Imputation. Since the length of each decision epoch in our framework is one month but our data set only includes information at months 1, 4, and 12 post-transplant, we employ the *cubic spline interpolation* method (see, e.g., Alagoz et al. (2005)) to simulate the natural clinical history of patients for months 1 to 12 post-transplant. Prior to that, to replace missing values in the data entries, we employ *multiple imputations by chained equations* (MICE) by the R computing package (see, e.g., Buuren and Groothuis-Oudshoorn (2011) for more details).

Risk Factors. As noted earlier, our goal is to derive robust optimal medication policies based on different risk factors. Table 3 summarizes the main risk factors affecting NODAT patients, where each risk factor is considered to be *low* or *high*. In this table: (1) age is classified based on a 50-year-old threshold, making an almost equal percentage of patients in each age category.[10] (2) Non-White race includes Hispanic, Black, and Native Americans. (3) Diabetes history refers to the existence of diabetes prior to the time of transplant.[11] (4) The thresholds for classifying risk factors (except for age, gender, race, and blood pressure) as low/high is based on MedPlus (2015). (5) Blood pressure is defined as "low" for patients with systolic and diastolic blood pressure of "<120" and "<80" mm Hg, respectively, whereas it is defined as "high" when at least one of these conditions is violated (AHA 2015).

Remark 1 (Risk Profile). Table 3 suggest that there are as many as $2^{11} = 2,048$ risk profiles. Instead of analyzing each and every such risk profile, we consider $2 \times 11 = 22$ cohorts of patients by changing one risk factor at a time. We do so because (1) in our setting, not only core state transition and observation probabilities but also reward functions depend on risk factors. To the best of our knowledge, there is no study in the medical literature that reports on reward functions based on a patient's risk profile. (2) In order to estimate transition and observation probabilities for each of the 2^{11} risk profiles, one would require data of about $10,000$ patients (i.e., more than

[10] The median age of patients in our data set is 53 years, and 40% of patients are below 50. We consider age as a static risk factor (i.e., invariable in time), because the planning time horizon for our problem is one year.

[11] Among 407 patients, there were 115 patients (28%) with the history of diabetes before or at the time of transplant. Considering patients with or without diabetes history allows us to prioritize those who might be more vulnerable to the diabetogenic effect of tacrolimus.

Table 3 Description of main risk factors and their levels (see also Boloori et al. (2015))

Risk Factor (Abbreviation)	Unit	Low Level	High Level	Static (S)/Dynamic (D)
Age	Years	<50	$\geq$ 50	S
Gender	—	Female	Male	S
Race	—	White	non-White	S
Diabetes history (Diab Hist)	—	No	Yes	S
Body mass index (BMI)	kg/m^2	<30 (non-obese)	$\geq$30 (obese)	D
Blood pressure (BP)	—	Normal	Hypertension	D
Total cholesterol (Chol)	mg/dL	<200	$\geq$200	D
High-density lipoportein (HDL)	mg/dL	$\geq$40	<40	D
Low-density lipoportein (LDL)	mg/dL	<130	$\geq$130	D
Triglyceride (TG)	mg/dL	<150	$\geq$150	D
Uric acid (UA)	mg/dL	<7.3	$\geq$7.3	D

half of all kidney transplantations in the U.S. in 2015 (UNOS 2016)), which is much larger than the number of patients seen at our partner hospital. (3) As we will see, our approach of considering 22 cohorts of patients is strong enough to detect the impact of each risk factor on optimal prescription of medications.

Remark 2 (Dynamic Risk Factors). The importance of considering dynamic risk factors (see the last column of Table 3) in our study is that they can change the underlying dynamics of a patient's health state. For example, an increase in cholesterol level may increase the risk of moving to a diabetes state. To this end, we customize our optimal policies based on dynamic risk factors in the following manner. As mentioned in Remark 1, we estimate transition and observation probabilities as well as reward functions for each cohort of patients. This enables us to obtain the set of non-dominated $\boldsymbol{\psi}$-vectors for any cohort at any period (see Propositions (1)-(2) in §4). Next, whenever a dynamic risk factor level changes at any period, it implies that the patient alternates his/her cohort (and thus the corresponding set of non-dominated $\boldsymbol{\psi}$-vectors) at that period. Due to the fact that such $\boldsymbol{\psi}$-vectors are obtained based on the number of periods remaining until the end of time horizon, this change of cohort indicates that the patient may be prescribed by different medications in the next period (see also Ayer et al. (2012) for a related discussion).[12]

Finally, while it is apparent that our optimal medical actions will affect the transitions of the core health states, we also aim to observe whether or not there exist significant associations between using medications (tacrolimus and insulin) and high levels of dynamic risk factors (e.g., effect of medications used on gaining weight, and hence, increasing BMI). To this end, Table 4 shows the results of a set of logistic regression analyses for dynamic risk factors. Based on these results, except for high-dose tacrolimus (under obesity) and insulin (under high UA), we could not establish any statistically significant association between using medications and increases in any of the dynamic risk factors.

Estimation of Transition Probability Matrices and Cloud Construction. The steps we have taken to estimate core state and observation probability matrices and to construct the cloud of

[12] As mentioned in §2, such dynamic risk factors are not defined as parts of the core state space, because doing so unnecessarily aggravates the "curse of dimensionality" in our model.

Table 4 Logistic regression results (numbers in parenthesis represent odds ratio and P-value, respectively; values in bold represent statistically significant numbers at 95% confidence level)

Medication	Dynamic Risk Factor						
	BMI	BP	Chol	HDL	LDL	TG	UA
Tacr: med-dose	(0.37,0.118)	(0.75,0.395)	(0.94,0.081)	(0.99,0.554)	(0.95,0.136)	(1.15,0.267)	(0.91,0.605)
Tacr: high-dose	(0.60,**0.001**)	(0.94,0.680)	(1.18,0.241)	(1.05,0.798)	(0.97,0.878)	(1.12,0.422)	(0.84,0.259)
Insulin	(1.47,0.095)	(1.05,0.865)	(0.97,0.885)	(1.64,0.052)	(0.91,0.749)	(1.04,0.848)	(0.57,**0.043**)
(Tacr: med-dose,insulin)	(1.13,0.488)	(1.04,0.708)	(0.77,0.419)	(0.76,0.391)	(0.96,0.715)	(1.00,0.540)	(1.03,0.442)
(Tacr: high-dose,insulin)	(1.30,0.366)	(1.21,0.550)	(0.92,0.756)	(0.61,0.126)	(0.98,0.957)	(1.02,0.932)	(1.12,0.731)

models are provided in detail in Table 5. In these steps: (1) we use the *Baum-Welch* (BW) algorithm (Welch 2003) to obtain *point* estimations for core state and observation probability matrices (steps 5–11 in Table 5). We also iterate this algorithm for 1,000 times to account for the inevitable variability caused by considering random initial transition probability matrices, and then, obtain the average outputs over all iterations. (2) Despite 1,000 iterations, the resulted point estimates for core state and observation probability matrices are not reliable. Thus, we construct an ambiguity set as a cloud of probabilistic models (steps 12–21 in Table 5). We do so by using the *Kullback-Leibler* (KL) divergence criterion (a.k.a. *relative entropy*):

$$d_{KL}\big(\mathbf{v}_1||\mathbf{P}^a_{BW}(i)\big)=\sum_{j\in\mathscr{S}}\mathbf{v}_1(j)\log_2\Big(\frac{\mathbf{v}_1(j)}{p^a_{BW}(j|i)}\Big)\ \text{for}\ \mathbf{v}_1\in\mathbf{\Theta}_1\quad\forall\, i\in\mathscr{S},\forall\, a\in\mathcal{A}, \tag{18a}$$

$$d_{KL}\big(\mathbf{v}_2||\mathbf{Q}^a_{BW}(i)\big)=\sum_{o\in\mathscr{O}}\mathbf{v}_2(o)\log_2\Big(\frac{\mathbf{v}_2(o)}{q^a_{BW}(o|i)}\Big)\ \text{for}\ \mathbf{v}_2\in\mathbf{\Theta}_2\quad\forall\, i\in\mathscr{S},\forall\, a\in\mathcal{A}, \tag{18b}$$

where for each $a\in\mathcal{A}$, $\mathbf{P}^a_{BW}=[p^a_{BW}(j|i)]_{i,j\in\mathscr{S}}$ and $\mathbf{Q}^a_{BW}=[q^a_{BW}(o|j)]_{j\in\mathscr{S},o\in\mathscr{O}}$ are the point estimates returned by the BW algorithm. Of note, for a row corresponding to an absorbing state, we do not apply the KL divergence, and consider a unit row vector instead. Moreover, we select the KL divergence bound ϵ so as to ensure that the cloud of models has a BIWC member (see Definition 1). In Online Appendix E, we discuss scenarios where the model informativeness condition (and thus the existence of a BIWC member) is satisfied in our setting.

Remark 3 (Exclusion of Death). Among the 407 patients in our clinical data set and throughout one year post-transplant, a total of 6 deaths occurred. Since (1) the point estimate for the transition probability matrix is obtained based on a sequence of actions and observations (see step 6 in Table 5), and (2) each risk factor is dichotomized into low and high risk levels, the above-mentioned death rate results in a very low transition probability from any core state to death (i.e., a probability less than 0.0027). Therefore, we do not incorporate "death" into our numerical experiments, and hence, adopt $\mathscr{S}=\{\nabla\}\cup\mathcal{S}$ and $\mathscr{O}=\{\nabla\}\cup\mathcal{O}$.

Estimation of Immediate and Lump-Sum Rewards. As introduced in §3, the immediate reward, $r_n(s,a)$, represents the QALE that a patient receives in period n based on core health state $s\in\mathscr{S}$, and the action taken $a\in\mathcal{A}$. We obtain these rewards based on the *quality-of-life* (*qol*), which is a score in $[0,1]$, where 0 (1) represents death (full health). Let a core health state be

Table 5 A pseudocode for estimating transition probability matrices and constructing the cloud of models

1:	input 1: $\lvert\mathcal{M}\rvert = 3$ // size of the ambiguity set/cloud of models
2:	input 2: $\epsilon \geq 0$ // overall Kullback-Leibler (KL) divergence bound
3:	input 3: sets of data samples $\mathbf{\Theta}_1 \in \mathbb{R}_+^{\lvert\mathscr{S}\rvert}$ and $\mathbf{\Theta}_2 \in \mathbb{R}_+^{\lvert\mathscr{O}\rvert}$ // (randomly generated vectors with elements summing to one)
4:	input 4: initial observation probability matrix (see Online Appendix D)
5:	**for each** cohort of patients in Table 3
6:	input 5: sequence of observations (tacrolimus C_0 and blood glucose levels) and actions (prescribed medications) from our data set (tailored based on each cohort)
7:	**for** i = 1 to 1,000
8:	input 6: an initial core state transition probability matrix (randomly generated)
9:	**do** *Baum-Welch* algorithm (based on inputs 4–6)
10:	**return** core state transition and observation probability matrices
11:	**return** point estimate: average of transition and observation probability matrices // $\mathbf{P}_{BW} = \{\mathbf{P}^a_{BW} : a \in \mathcal{A}\}$ and $\mathbf{Q}_{BW} = \{\mathbf{Q}^a_{BW} : a \in \mathcal{A}\}$
12:	**for** $a = 1$ to $\lvert\mathcal{A}\rvert$
13:	**while** the model informativeness condition is not met for probability sets $\mathbf{P}_m$ and $\mathbf{Q}_m$ (for all $m \in \mathcal{M}$)
14:	**for** $m = 1$ to $\lvert\mathcal{M}\rvert$
15:	**for** $i = 2$ to $\lvert\mathscr{S}\rvert$
16:	$\mathbf{\Theta}_1(i) = \{\mathbf{v}_1 : \mathbf{v}_1 \in \mathbf{\Theta}_1$ and $d_{KL}\big(\mathbf{v}_1 \lVert \mathbf{P}^a_{BW}(i)\big) \leq \epsilon\}$ // see (18a) for $d_{KL}(.\lVert.)$
17:	**do** select a vector $\mathbf{v}_1 \in \mathbf{\Theta}_1(i)$ and $\mathbf{P}^a_m(i) = \mathbf{v}_1$
18:	**for** $i' = 2$ to $\lvert\mathscr{S}\rvert$
19:	$\mathbf{\Theta}_2(i') = \{\mathbf{v}_2 : \mathbf{v}_2 \in \mathbf{\Theta}_2$ and $d_{KL}\big(\mathbf{v}_2 \lVert \mathbf{Q}^a_{BW}(i')\big) \leq \epsilon\}$ // see (18b) for $d_{KL}(.\lVert.)$
20:	**do** select a vector $\mathbf{v}_2 \in \mathbf{\Theta}_2(i')$ and $\mathbf{Q}^a_m(i') = \mathbf{v}_2$
21:	**return** probability sets $\mathbf{P}_m$, $\mathbf{Q}_m$ (for all $m \in \mathcal{M}$)

dichotomized into transplant and diabetes-related states: s^T and s^D, and $r_n(s^T, a)$ and $r_n(s^D, a)$ be the corresponding immediate rewards for these health states, respectively. Also, let $\langle x, y \rangle$ denote the average of two real numbers x and y. Then, we have for all $a \in \mathcal{A}$ and $n \leq N-1$: $r_n(s, a) = \big\langle r_n(s^T, a), r_n(s^D, a) \big\rangle$, where

$$r_n(s^T, a) = \begin{cases} qol(\text{organ rejection})/12, & \text{if } s^T = \text{Organ rejection}, \\ qol(\text{organ survival})/12, & \text{if } s^T = \text{Organ survival (different } C_0\text{'s)}, \end{cases} \tag{19a}$$

$$r_n(s^D, a) = \begin{cases} qol(\text{diabetes})/12, & \text{if } s^D = \text{Diabetic}, \\ qol(\text{pre-diabetes})/12, & \text{if } s^D = \text{Pre-diabetic}, \\ qol(\text{healthy})/12, & \text{if } s^D = \text{Healthy}. \end{cases} \tag{19b}$$

In (19a)-(19b), we note that the length of each period in our problem is one month, and thus, the corresponding *qol* scores are converted to a monthly basis (i.e., divided by 12).[13]

Furthermore, the lump-sum reward denoted by $R_n(s)$ is the QALE that a patient receives based on the core state s whenever s/he leaves the decision process (e.g., organ rejection or at the end of time horizon). Let $RLE(s, n) \geq 0$ be the *residual life expectancy* score (i.e., the expected remaining life years at any point of time) attributed to core state s in period n. Following Sassi (2006), we assume:

$$R_n(s) = \frac{qol(s)\big(1 - e^{-r\, RLE(s,n)}\big)}{r} \quad \forall\, s \in \mathscr{S}, \forall\, n \leq N, \tag{20}$$

[13] Note that equations (19a)-(19b) imply that the immediate reward is nondecreasing in the health state $s \in \mathscr{S}$. Thus, Condition 2(i) in §EC.3 is met in our numerical setting.

where r is a discount rate which accounts for degradation of the core health state over the remaining lifetime of a patient. In (20), $qol(s) = \langle qol(s^T), qol(s^D) \rangle$, and $RLE(s,n) = \langle RLE(s^T,n), RLE(s^D,n) \rangle$, where $RLE(s^T,n)$ and $RLE(s^D,n)$ are defined similar to (19a)-(19b).[14] Further details about estimating the required parameters (e.g., *qol* and *RLE* scores) can be found in Online Appendices F and J, where we also perform various sensitivity analyses regarding these parameters.

5.2. Numerical Results, Guidelines, and Policy Implications

In this section, we present our numerical results including the robust optimal medication policies for different cohorts of patients (§5.2.1) and comparison of our optimal policies with other policies including the one currently used in practice (§5.2.2). These results shed light on important implications for policy makers as well as individual physicians and patients which we will discuss in what follows.

5.2.1. Robust Optimal Medication Policies. We provide optimal policies obtained from our APOMDP approach separately for 22 cohorts of patients based on the risk factors described in Table 3. For each of these cohorts, we consider 3 different values for the DM's conservatism level (i.e., $\lambda \in \{0.0, 0.5, 1.0\}$) and 3 models for the ambiguity set (i.e., $|\mathcal{M}| = 3$). Also, due to the wide variety of parameters involved (e.g., risk factors and conservatism levels), we do not illustrate all the optimal medication regimens obtained at each period under all combinations of these parameters. Instead, we focus on the results for the first period in 12-month, 6-month, and 2-month horizon problems. We illustrate the results graphically in Figure 3 here and in Figure EC.1 in Online Appendix G. In these figures, each 2-simplex represents a cut of the belief space under a specific concentration of tacrolimus in the body of the patient. For example, a 2-simplex under "Low C_0" indicates $b_2, b_5, b_8 \neq 0$ and $b_1, b_3, b_4, b_6, b_7, b_9, b_{10} = 0$ (i.e., the patient is believed to have organ survival with low C_0, while the exact diabetes status is unknown). Although we calculate optimal medications over the full belief space Π_{PO}, which is an 8-simplex, we choose these cuts for the sake of illustration. This also helps us to understand the interaction of two medications under different risks of organ rejection and diabetes complications. Below, we summarize the main observations from our results.

Observation 1 **(Role of Ambiguity Attitudes in Medication Intensification).** *Increasing the conservatism level, λ, results in using (i) more intensive medication regimens for patients with non-White race, no diabetes history, low-risk levels of Chol, HDL, LDL, TG, and UA, age<50, age≥50, and BMI (both non-obese and obese), and (ii) less intensive medication regimens for patients with male gender, diabetes history, hypertension, and high-risk levels of BP, Chol, HDL, and LDL. Also, increasing λ does not change the intensity of medication regimens for patients with White race, female gender, no blood pressure, and high-risk levels of TG and UA.*

[14] Therefore, Condition 2(ii) in §EC.3 is also met in our medical problem.

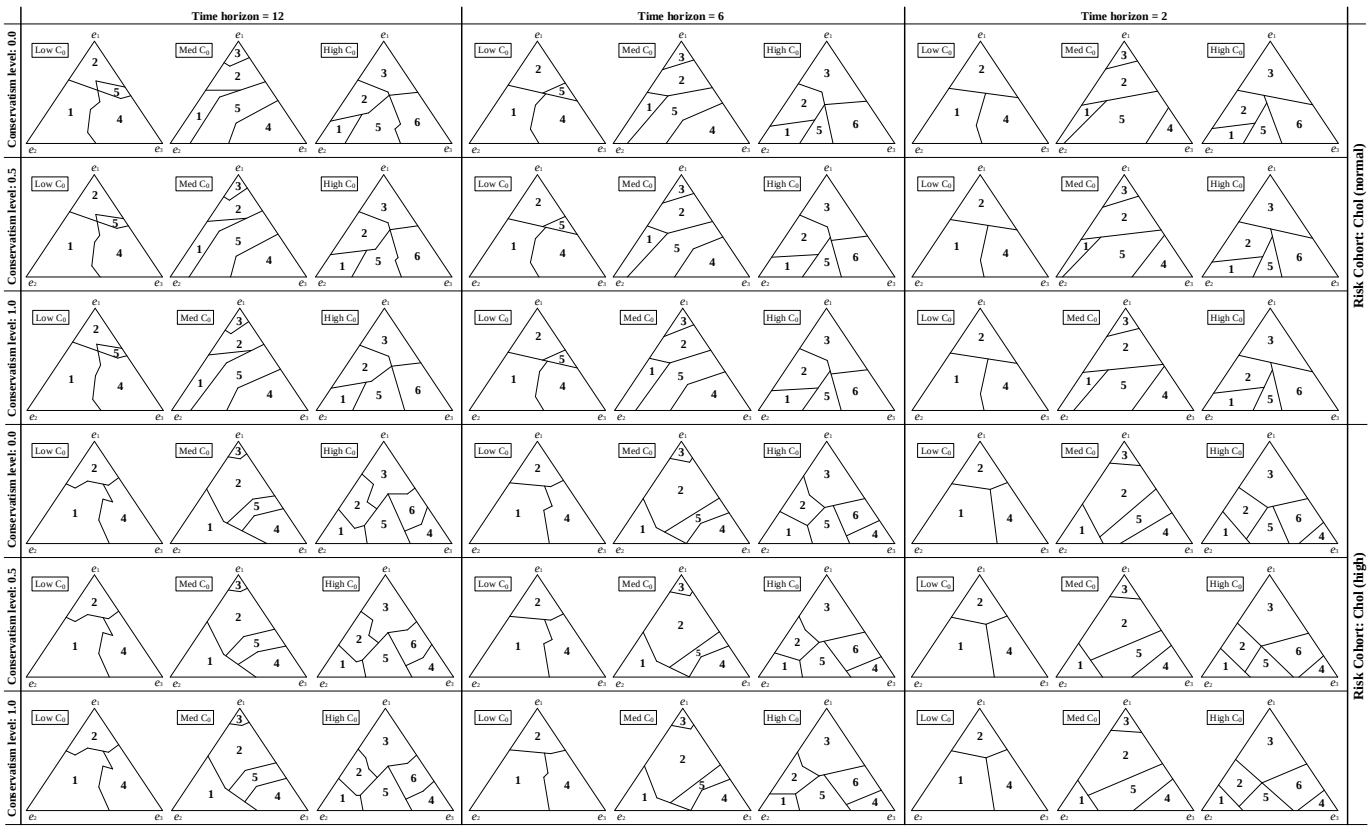

Figure 3 Optimal medication policies for patient cohorts "Chol: normal" and "Chol: high" (Number i inside a simplex corresponds to optimal action a_i^* for the first visit ($n = 1$) (see Table 2); e_j denotes a unit vector with j^{th} element equal to 1 and other elements equal to 0; e_1, e_2, e_3 represent diabetic, pre-diabetic, and healthy conditions, respectively)

For example, as can be observed from Figure 3, when a patient has normal cholesterol levels, increasing the conservatism level from 0 to 1 typically results in larger regions for which a high dosage of tacrolimus is prescribed under the optimal policy. However, this result is reversed when the patient is in danger of elevated cholesterol levels.

Observation 1 matches the insights we gained from Theorem 1: depending on patient characteristics, a higher conservatism level may induce a higher or lower level of drug intensification. Specifically, Observation 1 reveals that, for patients with non-White race, no diabetes history, low-risk levels of Chol, HDL, LDL, TG, and UA, a more conservative DM will prescribe more intensive tacrolimus regimens under the optimal policy. This suggests that, for such cohorts, the DM should be typically more concerned about the risk of organ rejection than the potential risk of NODAT. Moreover, regardless of age (i.e., age<50 vs. age≥50) and BMI (non-obese vs. obese), increasing the conservatism level results in more intensive tacrolimus regimens. On the other hand, for patients with male gender, diabetes history, hypertension, and high-risk levels of BP, Chol, HDL, and LDL, a more conservative DM should prescribe a lower dosage of tacrolimus. The implication for the medical practice is that the DM should typically be more concerned about diabetes complications than

an organ rejection for such patients, and hence, seek to combat the diabetogenic effect of tacrolimus. Furthermore, for patients with White race, female gender, no blood pressure, and high-risk levels of TG and UA, Observation 1 indicates that increasing the conservatism level does not drastically affect the intensity of prescribed medications under the optimal policy. This, in turn, implies that, for such cohorts, the DM should typically be equally concerned about risks of organ rejection and diabetes complications.

Observation 1 also reveals that the variations in physicians attitude toward ambiguity will not show a homogeneous pattern with respect to the intensity of the drugs used, if physicians follow the optimal policy. Thus, the drug intensification problem observed in practice should not be attributed merely to physicians' behavior toward ambiguity. Instead, this result suggests that lack of adherence to (or knowledge of) the optimal medications might be the main cause of using intensive regimens in the current practice.

Observation 2 **(Interactions: Risk Factors and Tacrolimus Requirement).** *Under any conservatism level, higher dosage of tacrolimus should be prescribed more for patients with age $\geq$50, male gender, non-White race, diabetes history, high BMI, hypertension, high Chol, and high LDL than patients with the opposing risk levels.*

It is known in the medical literature that age and race can be predictors of tacrolimus dose variability (see, e.g., Yasuda et al. (2008)). However, Observation 2 suggests that the dosage of tacrolimus should be adjusted based on many other static or dynamic risk factors of patients. In particular, Observation 2 implies that age, gender, race, diabetes history, BMI, blood pressure, total cholesterol, and LDL cholesterol are among the risk factors that make patients more vulnerable to the risk of organ rejection. Therefore, to offset this vulnerability, the optimal tacrolimus regimens put more emphasis on higher dosages of tacrolimus for such patients. In contrast to the above-mentioned risk factors, our results show that HDL, TG, and UA are not good predictors of tacrolimus dose variability: we barely observe any difference between the percentage of patients with opposing levels of these risk factors who are prescribed with high dosage of tacrolimus under the optimal policy.

Observation 3 **(Interactions: Risk Factors and the Diabetogenic Effect).** *Under all conservatism levels, patients who are prescribed high/medium dosage of tacrolimus become more dependent on using insulin when they have the following risk factors: age$\geq$50, male gender, White race, diabetes history, hypertension, high total cholesterol, low HDL, high TG, and high UA.*

This observation reveals patient risk factors under which the diabetogenic effect of tacrolimus is stronger. This finding is important for the medical practice, especially because it highlights that the blood glucose level of patients with such risk factors should be monitored more closely than other patients in the post-transplant period.

Finally, it should be noted that the combination of Observations 2 and 3 can provide further insights for the medical practice. For example, from Observation 2, we know that a patient with age$\geq$50 will be more likely to be prescribed by higher dosage of tacrolimus. From Observation 3, we

find that the same patient would be more dependent on insulin if being prescribed by medium/high dosage of tacrolimus. Combining these two findings, a physician will be able to fine tune the medication regimens for that patient (the same deduction could be obtained for other risk factors).

5.2.2. Comparison of Optimal Policies with the Current Practice. We now compare the optimal policies we obtain from our APOMDP approach with: (1) a *Benchmark* policy that resembles the current medical practice, and (2) a policy that will be obtained if one utilizes a traditional POMDP by ignoring the underlying ambiguities, and considering only one model.[15] Comparing our approach to the policy in (1) reveals the advantage of our policy compared to the current medical practice, and comparing it to the policy in (2) illuminates the need for using an APOMDP rather than a POMDP.

Benchmark Policy. We denote by $a^B_{n,o}$ the joint actions under the Benchmark policy (B: Benchmark) and observation o in period n, and define it as:

$$a^B_{n,o} = \begin{cases} a_1, & \text{if } n \leq 6 \text{ and } o \in \{o_1, o_2, o_3\}, \\ a_4, & \text{if } n \leq 6 \text{ and } o \in \{o_4, o_5, o_6, o_7, o_8, o_9\}, \\ a_2, & \text{if } 6 < n \leq 9 \text{ and } o \in \{o_1, o_2, o_3\}, \\ a_5, & \text{if } 6 < n \leq 9 \text{ and } o \in \{o_4, o_5, o_6, o_7, o_8, o_9\}, \\ a_3, & \text{if } 9 < n \leq 11 \text{ and } o \in \{o_1, o_2, o_3\}, \\ a_6, & \text{if } 9 < n \leq 11 \text{ and } o \in \{o_4, o_5, o_6, o_7, o_8, o_9\}. \end{cases} \tag{21}$$

Note that $a^B_{n,o}$ is independent of belief vector $\mathbf{b}$: to resemble the current practice, the outcomes of medical tests (observations) are treated as the actual health state of the patient (Bennett et al. 2007). In addition, in the current medical practice, insulin is typically prescribed only if a patient is observed to be diabetic. Furthermore, if not prescribed in combination with other immunosuppressive therapies, tacrolimus is typically administered in high dosage in months 0-6 post-transplant and in medium (low) dosage during months 6–9 (9–11) post-transplant. These are all captured in the above definition of $a^B_{n,o}$.

We use a micro-simulation model (see Online Appendix I) to compare the optimal policy from our APOMDP approach with the Benchmark and POMDP policies. Table 6 shows the results of our comparisons based on three performance measures: (1) average QALE achieved, (2) average medical expenditures (see Online Appendix H for related cost estimations), and (3) average number of times that insulin and different dosage of tacrolimus are prescribed. The latter allows us to examine whether or not our methodology yields less intensive medication regimens compared to the Benchmark policy (i.e., whether or not it alleviates the drug intensification problem observed in the current practice). We make the following observations from the results presented in Table 6.

Observation 4 **(Impact).** *Depending on the cohort of patients, during one year post-transplant our optimal policy (i) improves the average QALE per patient between 0.66% and 5.16% compared*

[15] To be fair, we consider the performance of this policy after getting average over all possible models.

Table 6 Comparison of medication policies

Risk Factor	Performance Measure	Policy: Benchmark	Policy: POMDP	Policy: Optimal	Improvement of Optimal policy over Benchmark	Improvement of Optimal policy over POMDP
Age < 50	Avg. QALE (yrs)	16.77	17.08	17.14	**2.21**%	**0.35**%
	Avg. Cost ($)	5,417	5,163	5,098	**5.89**%	**1.26**%
	# Med Used	(6,3,2,5.49)[1]	(4.07,4.35,2.58,5.24)	(3.66,4.87,2.47,5.12)	(**2.34**↓, **1.87**↑, **0.47**↑, **0.37**↓)[2]	(**0.41**↓, **0.52**↑, **0.11**↓, **0.12**↓)
Age ≥ 50	Avg. QALE (yrs)	9.48	9.82	9.85	**3.90**%	**0.31**%
	Avg. Cost ($)	5,705	5,352	5,245	**8.06**%	**2.00**%
	# Med Used	(6,3,2,5.52)	(4.88,4.33,1.79,7.48)	(4.68,4.81,1.51,7.62)	(**1.32**↓, **1.81**↑, **0.49**↓, **2.10**↑)	(**0.20**↓, **0.48**↑, **0.28**↓, **0.14**↑)
Gender: Female	Avg. QALE (yrs)	18.05	18.12	18.17	**0.66**%	**0.28**%
	Avg. Cost ($)	5,752	5,521	5,415	**5.86**%	**1.92**%
	# Med Used	(6,3,2,5.49)	(4.77,4.05,2.18,5.18)	(4.13,4.44,2.43,4.94)	(**1.87**↓, **1.44**↑, **0.43**↑, **0.55**↓)	(**0.64**↓, **0.39**↑, **0.25**↑, **0.24**↓)
Gender: Male	Avg. QALE (yrs)	15.77	15.90	15.93	**1.01**%	**0.19**%
	Avg. Cost ($)	5,877	5,640	5,584	**4.99**%	**0.99**%
	# Med Used	(6,3,2,5.50)	(5.18,3.89,1.93,5.47)	(5.05,4.03,1.92,5.70)	(**0.95**↓, **1.03**↑, **0.08**↓, **0.20**↑)	(**0.13**↓, **0.14**↑, **0.01**↓, **0.23**↑)
Race: White	Avg. QALE (yrs)	16.42	16.62	16.68	**1.58**%	**0.36**%
	Avg. Cost ($)	5,508	5,144	5,113	**7.17**%	**0.60**%
	# Med Used	(6,3,2,5.51)	(4.71,3.83,2.46,4.84)	(4.65,3.97,2.38,4.89)	(**1.35**↓, **0.97**↑, **0.38**↑, **0.62**↓)	(**0.06**↓, **0.14**↑, **0.08**↓, **0.05**↑)
Race: Non-White	Avg. QALE (yrs)	13.61	13.97	14.02	**3.01**%	**0.36**%
	Avg. Cost ($)	6,255	5,804	5,661	**9.50**%	**2.46**%
	# Med Used	(6,3,2,5.49)	(5.05,3.87,2.08,5.77)	(4.83,4.06,2.11,5.85)	(**1.17**↓, **1.06**↑, **0.11**↑, **0.36**↑)	(**0.22**↓, **0.19**↑, **0.03**↑, **0.08**↑)
Diab Hist: No	Avg. QALE (yrs)	15.05	15.15	15.19	**0.93**%	**0.26**%
	Avg. Cost ($)	5,427	5,218	5,180	**4.55**%	**0.72**%
	# Med Used	(6,3,2,5.50)	(4.65,3.81,2.54,5.31)	(4.12,4.27,2.61,4.85)	(**1.88**↓, **1.27**↑, **0.61**↑, **0.65**↓)	(**0.53**↓, **0.46**↑, **0.07**↑, **0.46**↓)
Diab Hist: Yes	Avg. QALE (yrs)	8.52	8.94	8.96	**5.16**%	**0.22**%
	Avg. Cost ($)	6,833	6,212	6,147	**10.04**%	**1.05**%
	# Med Used	(6,3,2,5.50)	(3.95,4.41,2.64,7.50)	(3.25,5.03,2.72,8.14)	(**2.75**↓, **2.03**↑, **0.72**↑, **2.64**↑)	(**0.70**↓, **0.62**↑, **0.08**↑, **0.64**↑)
BMI	Avg. QALE (yrs)	13.83	14.12	14.15	**2.31**%	**0.21**%
	Avg. Cost ($)	5,717	5,506	5,443	**4.79**%	**1.14**%
	# Med Used	(6,3,2,5.52)	(4.83,4.17,2.00,6.17)	(4.15,4.79,2.06,6.88)	(**1.85**↓, **1.79**↑, **0.06**↑, **1.36**↑)	(**0.68**↓, **0.62**↑, **0.06**↑, **0.61**↑)
Blood Pressure	Avg. QALE (yrs)	13.64	13.80	13.89	**1.83**%	**0.65**%
	Avg. Cost ($)	5,659	5,488	5,274	**6.80**%	**3.90**%
	# Med Used	(6,3,2,5.52)	(5.07,3.85,2.08,5.96)	(4.25,4.70,2.05,6.38)	(**1.75**↓, **1.70**↑, **0.05**↑, **0.86**↑)	(**0.82**↓, **0.85**↑, **0.03**↓, **0.42**↑)
Cholesterol	Avg. QALE (yrs)	13.47	13.60	13.68	**1.56**%	**0.59**%
	Avg. Cost ($)	5,870	5,566	5,494	**6.41**%	**1.29**%
	# Med Used	(6,3,2,5.49)	(5.02,3.94,2.04,6.13)	(4.34,4.79,1.87,6.78)	(**1.66**↓, **1.79**↑, **0.13**↓, **1.29**↑)	(**0.68**↓, **0.85**↑, **0.17**↓, **0.65**↑)
HDL	Avg. QALE (yrs)	13.36	13.71	13.78	**3.14**%	**0.51**%
	Avg. Cost ($)	5,814	5,329	5,144	**11.52**%	**3.47**%
	# Med Used	(6,3,2,5.51)	(4.56,3.76,2.68,6.07)	(3.78,4.69,2.53,6.67)	(**2.22**↓, **1.69**↑, **0.53**↑, **1.16**↑)	(**0.78**↓, **0.93**↑, **0.15**↓, **0.60**↑)
LDL	Avg. QALE (yrs)	13.61	13.82	13.90	**2.13**%	**0.58**%
	Avg. Cost ($)	5,745	5,578	5,391	**6.16**%	**3.35**%
	# Med Used	(6,3,2,5.47)	(5.14,3.66,2.20,6.13)	(4.55,4.52,1.93,6.98)	(**1.45**↓, **1.52**↑, **0.07**↓, **1.51**↑)	(**0.59**↓, **0.86**↑, **0.27**↓, **0.85**↑)
Triglyceride	Avg. QALE (yrs)	13.35	13.47	13.51	**1.20**%	**0.30**%
	Avg. Cost ($)	6,094	5,774	5,616	**7.84**%	**2.74**%
	# Med Used	(6,3,2,5.49)	(4.66,3.38,2.96,6.13)	(4.15,4.03,2.82,6.56)	(**1.85**↓, **1.03**↑, **0.82**↑, **1.07**↑)	(**0.51**↓, **0.65**↑, **0.14**↓, **0.43**↑)
Uric Acid	Avg. QALE (yrs)	13.00	13.23	13.25	**1.92**%	**0.15**%
	Avg. Cost ($)	5,815	5,423	5,280	**9.20**%	**2.64**%
	# Med Used	(6,3,2,5.52)	(4.22,4.03,2.75,5.68)	(3.97,4.17,2.86,5.88)	(**2.03**↓, **1.17**↑, **0.86**↑, **0.36**↑)	(**0.25**↓, **0.14**↑, **0.11**↑, **0.20**↓)

[1] (a, b, c, d): high-/medium-/low-dose tacrolimus and insulin are used for a, b, c, and d times during a year, respectively.
[2] ↑ (↓) represents the amount of increase (decrease) in number of times a medication is prescribed.

to the Benchmark policy and between 0.15% and 0.65% compared to the POMDP policy, (ii) reduces the average medical expenditures per patient between 4.55% and 11.52% compared to the Benchmark policy and between 0.60% and 3.90% compared to the POMDP policy, and (iii) prescribes high-dose tacrolimus between 0.95 and 2.75 fewer times per patient compared to the Benchmark policy, and between 0.06 and 0.82 fewer times per patient compared to the POMDP policy. However, our optimal

policy prescribes medium-dose tacrolimus between 0.97 and 2.03 more times per patient compared to the Benchmark policy, and between 0.14 and 0.93 more times per patient compared to the POMDP policy.

Observation 4 and the results provided in Table 6 have several important implications for medical practitioners as well as policy makers:

(1) Our methodology can improve the average QALE while incurring lower medical expenditures compared to the current practice. That is, our proposed methodology yields a more *cost-effective* medication policy than the current medical practice. Furthermore, we note that the improvements reported are garnered over the course of one year post-transplant. During a longer horizon, we expect to observe more improvements.

(2) Considering the traditional cost per QALE measure of cost-effectiveness, Figure 4 illustrates the performance of our optimal policy compared to the Benchmark policy for various cohorts of patients. As can be seen from this figure, cohorts of patients formed by age (both cohorts), non-White race, diabetes history, HDL (both cohorts), and blood pressure (both cohorts) will benefit most from our methodology, because our approach yields the most improvement in QALE while incuring the least amount of medical expenditure. However, for the cohorts formed by gender (both cohorts), no diabetes history, and total cholesterol (both cohorts), such improvements are not significant, and hence, practitioners could stay with the current practice for such patients.

(3) For some cohorts of patients (e.g., formed by non-White race, diabetes history, and HDL (both risk levels)), the improvement in QALE gained from our optimal policy (compared to the current medical practice) is 3 months or more within one year post-transplant. The main reason for this striking outcome is the fact that tacrolimus is heavily prescribed in the current medical practice during this period. However, after one year post-transplant, tacrolimus dosage is gradually reduced on a maintenance level, and hence, the potential risk of NODAT drops after the first year.[16]

(4) The comparison between our APOMDP approach and the POMDP policy reveals that, had we ignored the underlying model misspecifications, each patient would have lost between 0.02 and 0.09 QALE on average (i.e., between 1.04 and 4.68 weeks), while incuring between \$31 and \$214 more medical costs (see Figure 5). This shows the importance of considering model misspecifications that are inevitable when data is used to estimate parameters: practitioners, researchers, and policy makers should not rely on a single model to derive cost-effective guidelines.

(5) The medical expenditure reported in Observation 4 (ii) does not cover the direct cost of medications used, because our purpose is to compare the monetary values of disutilities imposed by risks of organ rejection and diabetes complications resulted from different policies (see Online Appendix H). However, our findings show that our proposed methodology results in less intensive medications strategies compared to the current medical practice. Since less intensive regimens imply less direct

[16] As noted in §3, this is the main reason that we analyze the one-year post-transplant in our APOMDP model.

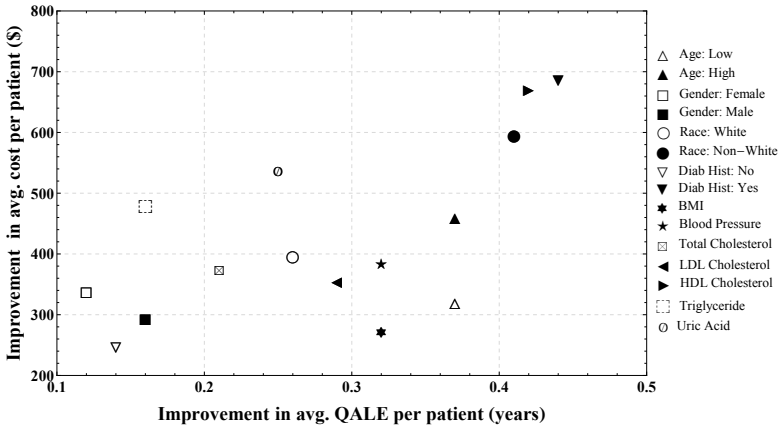

Figure 4 **Improvements in average medical expenditure and QALE gained per patient over one year post-transplant: A comparison between the optimal and Benchmark policies**

medication costs, our results also suggest that the total medical expenditure under our optimal policy is less than that under the current practice.

(6) Comparing the optimal and Benchmark policies, the average number of times that low-dose tacrolimus and insulin are prescribed does not show a homogeneous pattern across different risk factors (unlike cases for high or medium-dose tacrolimus). For example, for the cohorts of patients formed by age$\geq$50, male gender, cholesterol, and LDL, low-dose tacrolimus is prescribed less by the optimal policy than by the Benchmark policy. However, for all other cohorts of patients, the optimal policy prescribes more low-dose tacrolimus compared to the Benchmark policy. This implies that, for the former group of risk factors, a physician should attempt to diminish the intensity of tacrolimus regimens to offset the potential risk of NODAT. Moreover, for the cohorts of patients formed by age$\geq$50, male gender, non-White race, diabetes history and all dynamic risk factors, the optimal policy increases the average number of insulin prescriptions compared to the medical practice, whereas, for cohorts formed by age$<$50, female gender, White race, and no diabetes history, the optimal policy decreases this number compared to the medical practice. This, in turn, supports the implications we discussed in Observation 3 about the customization of insulin requirements with respect to different risk factors.[17]

6. Conclusion

Immunosuppressive medications (a.k.a. anti-rejection drugs) are currently intensively prescribed in the post-transplant period to ensure a low risk of organ rejection. However, this practice has been shown to increase the risk of new-onset diabetes after transplantation (NODAT), which, in turn, necessitates the use of medications such as insulin. To provide guidelines for the simultaneous

[17] In Online Appendix J, we conduct various sensitivity analyses, and show that the findings discussed above are robust to fluctuation in estimated reward parameters.

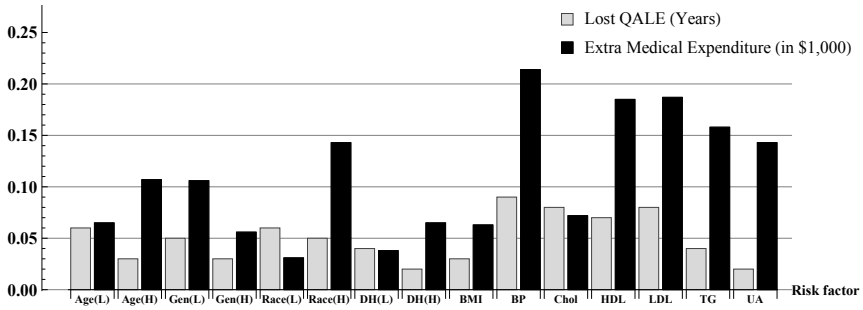

Figure 5 **APOMDP vs. POMDP: Lost QALE and extra medical expenditure imposed by ignoring the underlying model misspecification (L: low level, H: high level, Gen: gender, and DH: diabetes history)**

management of post-transplant medications such as tacrolimus and insulin, we develop an ambiguous POMDP (APOMDP) model that maximizes the quality-adjusted life expectancy (QALE) of patients, while controlling the risk of organ rejection and NODAT. Utilizing our APOMDP approach along with a data set of patients who underwent kidney transplantation at our partner hospital, we establish a data-driven approach in which (1) the physician's ambiguity attitude toward model misspecifications is defined based on a combination of the worst and the best possible outcomes in the "cloud" of models, (2) core state and observation transition probability matrices are patient risk-factor specific but subject to potential estimation errors and other causes of misspecifications, and (3) optimal policies are patient-centric and customized for different cohorts of patients. Moreover, our notion of simultaneous prescriptions facilitates the care coordination between "Transplantation/Nephrology" and "Endocrinology" departments of a hospital that are typically in charge of administering tacrolimus and insulin, respectively.

Analyzing the APOMDP model, we first present several structural properties to facilitate the search for finding optimal medication strategies. These include piecewise-linearity and convexity of the value function, a feasible bound as an approximation, monotonicity of the value function as well as optimal policy, and a theoretical link between a decision maker's conservatism level and the intensity of prescribed medications. We then perform various numerical experiments using our clinical data set, and discuss their implications for policy makers and physicians. These include: (1) for patients with non-White race, no diabetes history, low-risk levels of cholesterol, HDL, LDL, triglyceride, and uric acid, a DM should be more concerned about organ rejection than the potential risk of NODAT. On the contrary, for patients with male gender, diabetes history, hypertension, and high-risk levels of cholesterol, HDL, and LDL, the DM should be more concerned about the risk of NODAT than that of organ rejection, and hence, s/he should seek to combat the diabetogenic effect of tacrolimus. (2) Variations in physicians attitude toward ambiguity will not show a homogeneous pattern with respect to the intensity of the drugs used, if physicians follow the optimal policy. Thus, the intensification of drug regimens observed in practice should not be attributed merely to physicians' behavior toward ambiguity. Instead, our results suggest that lack of adherence to (or

knowledge of) the optimal medications might be the main cause of using intensive regimens of drugs in the current practice. (3) The diabetogenic effect of tacrolimus is more likely to influence patients with age$\geq$50, male gender, White race, diabetes history, hypertension, high total cholesterol, and low HDL, high TG, and high UA. This implies that, when taking high-dose tacrolimus, such patients typically become more insulin-dependent than others.

We also compare our proposed optimal policies with two other medication policies: a Benchmark policy that represents the current medical practice, and a POMDP-based policy that ignores the underlying model misspecifications. We do so by considering three performance measures: average QALE, average medical expenditures, and average number of medications used in one year post-transplant. Our results show that, depending on different risk factors considered for each patient, in one year post-transplant our optimal policy (a) improves the average QALE between 0.66% and 5.16% per patient compared to the Benchmark policy and between 0.15% and 0.65% compared to the POMDP policy, (b) reduces the average medical expenditures between 4.55% and 11.52% per patient compared to the Benchmark policy and between 0.60% and 3.90% compared to the POMDP policy, and (c) prescribes high-dose tacrolimus between 0.95 and 2.75 fewer times per patient compared to the Benchmark policy and between 0.06 and 0.82 fewer times per patient compared to the POMDP policy. However, the optimal policy prescribes medium-dose tacrolimus between 0.97 and 2.03 more times per patient compared to the Benchmark policy, and between 0.14 and 0.93 more times compared to the POMDP policy.

The most important implications of the above-mentioned results for practitioners and policy makers are: (1) our methodology can improve the average QALE while incurring lower medical expenditures compared to the current practice, providing a more cost-effective medication policy for use in practice. (2) Cohorts of patients formed by age (both cohorts), non-White race, diabetes history, HDL (both cohorts), and blood pressure (both cohorts) will benefit most from our methodology, because for such patients our approach yields the most improvement in QALE while incuring the least amount of medical expenditure. (3) Practitioners, researchers, or policy makers should not rely on a single model to derive cost-effective guidelines: had we ignored the underlying model misspecifications, each patient on average would have lost between 1.04 and 4.68 weeks of QALE during one year, while incuring between \$31 and \$214 more medical costs durign the same period.

Our study has some limitations: (1) Although the therapeutic monitoring of tacrolimus requires the surveillance of a patient's weight, we do not consider the effect of weight on optimal medication regimens (beyond customizing it based on BMI). To consider the interactions between a patient's weight and his/her other risk factors, future research would need to analyze a larger sample of patients. (2) We consider tacrolimus as the main immunosuppressive drug in this study, based on the practice at our partner hospital. Some of our results might be specific to tacrolimus, and should not be extended to other immunosuppressive drugs without additional analysis. Furthermore, unlike the case at our partner hospital, multiple immunosuppressive drugs may be used in parallel in

some medical practices. Including all such drugs in our APOMDP approach will increase state and action spaces, aggravating the so-called "curse of dimensionality." This will necessitate using some approximation schemes (e.g., utilizing a lower bound approach similar to the one we discussed in §4, or obtaining policies via approximate dynamic programming). (3) Our structural properties provide theoretical insights to facilitate managing medications post-transplant. However, some of the sufficient conditions required for establishing such properties are slightly violated by the estimated parameters from our data set (see, e.g., Online Appendix C). Although we provide an approximation scheme to alleviate this issue, developing more advanced data-driven estimation methods to account for these sufficient conditions while satisfying the required key properties such as the existence of a BIWC member in the "cloud" of models is a valuable path for future research.

Future research can also extend our work in at least two other directions. First, our approach can be applied to other solid organs (e.g., liver, pancreas, and heart) with the goal of creating a holistic multi-organ data-driven decision-support system (see Figure 2). The most distinguishing feature of kidney transplantation is the ability to use dialysis when facing organ rejection. However, dialysis is not feasible for other organs. As a result, risk of organ rejection is expected to be higher for other organs compared to kidney. This can play a key role in estimating reward parameters using data sets, which, in turn, can affect optimal medication policies. Second, future research may consider a resource allocation problem for hospitals, where the challenge is to effectively allocate limited resources (e.g., insulin and tacrolimus along with nurses and beds) to Endocrinology and Nephrology departments of hospitals for managing NODAT patients. This will create coordinated efforts between different parts of a hospital, and hence, may effectively reduce expenditures while improving the care delivery process.

References

ADA (2012) Standards of medical care in diabetes. *Diabetes Care.* 35:S11–S63.

AHA (2015) Understanding blood pressure readings. Retrieved May 05, 2015, http://www.heart.org/HEARTORG/Conditions/HighBloodPressure/AboutHighBloodPressure/Understanding-Blood-Pressure-Readings_UCM_301764_Article.jsp#.VivsfH6rTcs.

Alagoz O, Maillart LM, Schaefer AJ, Roberts MS (2004) The optimal timing of living-donor liver transplantation. *Management Science.* 50(10):1420–1430.

Alagoz O, Bryce CL, Shechter S, Schaefer A, Chang CCH, Angus DC, Roberts MS (2005) Incorporating biological natural history in simulation models: empirical estimates of the progression of end-stage liver disease. *Medical Decision Making.* 25(6):620–632.

Ata B, Skaro A, Tayur S (2016) OrganJet: Overcoming geographical disparities in access to deceased donor kidneys in the United States. *Management Science.* http://dx.doi.org/10.1287/mnsc.2016.2487.

Ayer T, Alagoz O, Stout NK (2012) OR forum-a POMDP approach to personalize mammography screening decisions. *Operations Research.* 60(5):1019–1034.

Ayer T, Alagoz O, Stout NK, Burnside ES (2015) Heterogeneity in women's adherence and its role in optimal breast cancer screening policies. *Management Science.* 62(5):1339–1362.

Bazin C, Guinedor A, Barau C, Gozalo C, Grimbert P, Duvoux C, Furlan V, Massias L, Hulin A (2010) Evaluation of the Architect tacrolimus assay in kidney, liver, and heart transplant recipients. *Journal of Pharmaceutical and Biomedical Analysis.* 53(4):997–1002.

Bennett CM, Guo M, Dharmage SC (2007) HbA1c as a screening tool for detection of type 2 diabetes: a systematic review. *Diabetic Medicine.* 24(4):333–343.

Bentley TS, Hanson SG (2011) 2011 US organ and tissue transplant cost estimates and discussion. *Milliman Research Report.* 1–20.

Bertsimas D, Farias VF, Trichakis N (2013) Fairness, efficiency, and flexibility in organ allocation for kidney transplantation. *Operations Research.* 61(1):73–87.

Boloori A, Saghafian S, Chakkera HA, Cook CB (2015) Characterization of remitting and relapsing hyperglycemia in post-renal-transplant recipients. *PLoS One.* 10(11):e0142363. doi:10.1371/journal.pone.0142363.

Bowman LJ, Brennan DC (2008) The role of tacrolimus in renal transplantation. *Expert Opinion on Pharmacotherapy.* 9(4):635–643.

Buuren S, Groothuis-Oudshoorn K (2011) MICE: multivariate imputation by chained equations in R. *Journal of Statistical Software.* 45(3):1–67.

Chakkera HA, Weil EJ, Castro J, Heilman RL, Reddy KS, Mazur MJ, Hamawi K, Mulligan DC, Moss AA, Mekeel KL, and others (2009) Hyperglycemia during the immediate period after kidney transplantation. *Clinical Journal of the American Society of Nephrology.* 4(4):853–859.

Delage E, Mannor S (2010) Percentile optimization for Markov decision processes with parameter uncertainty. *Operations Research.* 58(1):203–213.

Denton BT, Kurt M, Shah ND, Bryant SC, Smith SA (2009) Optimizing the start time of statin therapy for patients with diabetes. *Medical Decision Making.* 29:351–367.

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

Ghisdal L, Van Laecke S, Abramowicz MJ, Vanholder R, Abramowicz D (2012) New-onset diabetes after renal transplantation risk assessment and management. *Diabetes Care.* 35(1):181–188.

Goh J, Bayati M, Zenios SA, Singh S, Moore D (2015) Data uncertainty in Markov chains: Application to cost-effectiveness analyses of medical innovations. Working paper. http://web.stanford.edu/ bayati/papers/markov_chain.pdf.

Hauskrecht M (2000) Value-function approximations for partially observable Markov decision processes. *Journal of Artificial Intelligence Research.* 13:33–94.

Hu C, Lovejoy WS, Shafer SL (1996) Comparison of some suboptimal control policies in medical drug therapy. *Operations Research.* 44(5):696–709.

Iyengar GN (2005) Robust dynamic programming. *Mathematics of Operations Research.* 30(2):257–280.

Jusko WJ, Piekoszewski W, Klintmalm GB, Shaefer MS, Hebert MF, Piergies AA, Lee CC, Schechter P, Mekki QA (1996) Pharmacokinetics of tacrolimus in liver transplant patients. *Clinical Pharmacology & Therapeutics.* 57(3):281–290.

Karlin S, Rinott Y (1980) Classes of orderings of measures and related correlation inequalities. I. Multivariate totally positive distributions. *Journal of Multivariate Analysis.* 10(4):467–498.

Kaufman DL, Schaefer AJ, Roberts MS (2011) Living-donor liver transplantation timing under ambiguous health state transition probabilities–Extended Abstract. *Manufacturing & Service Operations Management (MSOM) Conference.* http://www-personal.umich.edu/ davidlk/pubs/robustLiverExtendedAbstract.pdf.

Kromann H, Borch E, Gale EA (1981) Unnecessary insulin treatment for diabetes. *British Medical Journal.* 283(6303):1386–1388.

Kurt M (2012) The structure of optimal statin initiation policies for patients with type 2 diabetes. PhD thesis, University of Pittsburgh.

Lehmann ED, Deutsch T (1992) A physiological model of glucose-insulin interaction in type 1 diabetes mellitus. *Journal of Biomedical Engineering.* 14(3):235–242.

Maillart LM, Ivy JS, Ransom S, Diehl K (2008) Assessing dynamic breast cancer screening policies. *Operations Research.* 56(6):1411–1427.

Mason JE, England DA, Denton BT, Smith SA, Kurt M, Shah ND (2012) Optimizing statin treatment decisions for diabetes patients in the presence of uncertain future adherence. *Medical Decision Making.* 32:154–166.

Mason JE, Denton BT, Shah ND, Smith SA (2014) Optimizing the simultaneous management of blood pressure and cholesterol for type 2 diabetes patients. *European Journal of Operational Research.* 233:727–738.

MedPlus (2015) Retrieved May 05, 2015, https://www.nlm.nih.gov/medlineplus/encyclopedia.html.

Monahan GE (1982) State of the artâĂŤa survey of partially observable Markov decision processes: theory, models, and algorithms. *Management Science.* 28(1):1–16.

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

OPTN (2011) OPTN/SRTR annual report: Transplant data 1999–2008. Retrieved February 23, 2015, http://www.srtr.org/annual_reports/archives/2009/2009_Annual_Report/.

Papadimitriou CH, Tsitsiklis JN (1987) The complexity of Markov decision processes. *Mathematics of Operations Research.* 12(3):441–450.

Saghafian S (2017) Ambiguous partially observable Markov decision processes: structural results and applications. Working Paper. Harvard University. Available at https://scholar.harvard.edu/files/saghafian/files/apomdp-web.pdf.

Sandikci B, Maillart LM, Schaefer AJ, Roberts MS (2013) Alleviating the patient's price of privacy through a partially observable waiting list. *Management Science.* 59(8):1836–1854.

Sassi F (2006) Calculating QALYs, comparing QALY and DALY calculations. *Health Policy and Planning.* 21(5):402–408.

Smallwood RD, Sondik EJ (1973) The optimal control of partially observable Markov processes over a finite horizon. *Operations Research.* 21(5):1071–1088.

Staatz C, Taylor P, Tett S (2001) Low tacrolimus concentrations and increased risk of early acute rejection in adult renal transplantation. *Nephrology Dialysis Transplantation.* 16(9):1905–1909.

Su X, Zenios SA (2005) Patient choice in kidney allocation: a sequential stochastic assignment model. *Operations Research.* 53(3):443–455.

Taguchi K, Ohmura T, Ohya Y, Horio M, Furukawa K, Jono H, Inomata Y, Saito H (2013) False tacrolimus concentrations measured by antibody-conjugated magnetic immunoassay in liver transplant patient: 2 case reports and literature review. *Experimental and Clinical Transplantation.* 12(5):474–478.

UNOS: United Network of Organ Sharing (2016) Transplant trends. Retrieved December 10, 2016, https://www.unos.org/data/transplant-trends/transplants_by_organ_type+year+2015.

Van'T Riet E, Alssema M, Rijkelijkhuizen JM, Kostense PJ, Nijpels G, Dekker JM (2010) Relationship between A1C and glucose levels in the general Dutch population the New Hoorn study. *Diabetes Care.* 33(1):61–66.

Van Zandt T (2002) An introduction to monotone comparative statics. Retrieved April 20, 2016, http://faculty.insead.edu/vanzandt/teaching/CompStatics.pdf.

Welch LR (2003) Hidden Markov models and the Baum-Welch algorithm. *IEEE Information Theory Society Newsletter.* 53(4):10–13.

Wiesemann W, Kuhn D, Rustem B (2013) Robust Markov decision processes. *Mathematics of Operations Research.* 38(1):153–183.

Xu H, Mannor S (2012) Distributionally robust Markov decision processes. *Mathematics of Operations Research.* 37(2):288–300.

Yasuda SU, Zhang L, Huang SM (2008) The role of ethnicity in variability in response to drugs: focus on clinical pharmacology studies. *Clinical Pharmacology & Therapeutics.* 84(3):417–423.

Zhang J (2011) Partially observable Markov decision processes for prostate cancer screening. PhD thesis, North Carolina State University.

Zhang Y (2014) Robust optimal control for medical treatment decisions. PhD thesis, North Carolina State University.