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 use 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 theoretical results that shed light on the structural properties of the optimal value function and policy. Using a clinical data set, we then compare the optimal policy to the current practice as well as some other benchmarks, and discuss relevant implications for both policy makers and physicians. In particular, our results show that potential 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: January 31, 2018

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 an immunosuppressive (a.k.a. anti-rejection) drug (e.g., tacrolimus). However, due to the well-known

¹This work was partially supported by the National Science Foundation through Grant CMMI-1562645 (PI: Saghafian).
diabetogenic effect, excessive exposure to an immunosuppressive drug may induce New Onset 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 make use of a data set of patients who had kidney transplant surgery at our partner hospital between 1999 and 2006. Based on our data set, 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 amounts of an immunosuppressive drug 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 retransplantation, which is a costly medical operation (Bentley and Hanson 2011). 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 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). 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).

**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).

**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 a dynamic decision-making 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 theory and application. From the application perspective, we contribute to the medical literature by presenting new clinically relevant findings: (1) we calibrate our
APOMDP model based on a clinical data set that we have collected from our partner hospital. 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). Using these estimations along with the 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 NO-DAT. 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 have a homogeneous impact on the intensity of drugs prescribed under the optimal policy. Thus, the drug intensification problem (i.e., use of intensified levels of medication regimens) observed in the current practice should not be attributed merely to physicians’ behavior toward ambiguity. Our result suggests that lack of adherence to (or knowledge of) the optimal medications is the main contributor to using intensive regimens 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, BMI, blood pressure, HDL, and LDL. Specifically, we find that these risk factors make patients more vulnerable to the risk of organ rejection. Furthermore, the diabetogenic effect of tacrolimus—a main immunosuppressive drug—is more likely to influence patients with age $\geq 50$, male gender, diabetes history, hypertension, high cholesterol, and low HDL. This implies that, when using high-dose tacrolimus, such patients typically become more dependent on diabetes medications than others. (5) We compare the performance of the optimal medication policies that we obtain from the APOMDP approach with (a) four benchmarks that are based on the current medical practice, and (b) medication policies that arise when one uses a traditional POMDP approach. We consider performance measures such as quality-adjusted life expectancy (QALE), medical expenditure, and the intensity of prescribed medications. Some of the main insights generated from our comparison are as follows:

- Compared to the current medical practice and depending on different risk factors, our optimal medication policies can improve (per patient per year) the average (a) QALE between 0.79% and 4.58%, and (b) medical expenditures between 4.01% and 11.57%. In particular, for cohorts of
patients formed by age, diabetes history, blood pressure, cholesterol, HDL, and triglyceride, our proposed medication strategies yield the highest improvements in QALE while incurring the least amount of medical expenditure, providing more cost-effective ways of managing medications. Our proposed medications strategies also prescribe use of high-dose tacrolimus up to 3.69 fewer times per patient per year compared to the current medical practice.

- We find that deriving optimal strategies via a traditional POMDP instead of using the APOMDP approach (i.e., ignoring inevitable parameter ambiguities) 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.

From the theory perspective, (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 applications including those in the general field of medical decision-making, our work sheds light on the advantages of an 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. Importantly, this removes the need to perform sensitivity analyses on such potential misspecifications. (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 present results on (a) an analytical link between a decision maker’s ambiguity attitude and the intensity of optimal medication regimens, (b) monotonicity results for the optimal medication policy, and (c) a lower bound for the optimal value function.

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.

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)), and optimal interval for other diabetes medications (see, e.g., Mason et al. (2014)). Unlike this stream of research, we
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.

**Operations Research/Management Science Studies on Pre-Transplant Period.** The majority of Operations Research/Management Science 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., Su and Zenios (2005), Bertsimas 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 mainly for cancer screening research. Examples include mammography screening in breast cancer (see, e.g., Ayer et al. (2012)), 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’ behavioral attitudes toward model misspecifications, and (3) is customized with eleven different risk factors (most of which are time-variant). From the medical perspective, the latter is an improvement, since age and history of screening/treatment are the only 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 Zhang (2014) and Goh et al. (2015), and the references therein. Among theoretical studies addressing robustness in dynamic decision-making, we refer to those solving MDPs with respect to a worst-case scenario (i.e., utilizing 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.\(^2\)

**Studies from the Medical Literature.** 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)), (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 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 (a) are prescribed by insulin, and (b) are put on a fixed dosage of it. Therefore, we (a) consider insulin as the main diabetes medication, and (b) assume it is prescribed in a fixed dosage (see also Denton et al. (2009) 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 label any \(C_0 \in [4,8), [8,10), [10,14]\) mg/dL as “low,” “medium,” and “high,” respectively. Similarly, we use labels “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. These discrete settings are consistent with the literature on therapeutic monitoring of immunosuppressive drugs (see, e.g., Schiff et al. (2007)). Also, from the

\(^2\) A particular advantage of considering the worst and the best possible outcomes (as opposed to all possible outcomes in between) is that it does not add much to the computational complexity. This is especially important for partially observable systems which are known to be computationally very complex (for more details see Saghaian (2017)).
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 labeled as diabetic (pre-diabetic), whereas FPG < 100 mg/dL or HbA1c < 5.7\% is labeled as healthy (ADA 2012).

Based on this premise, we use a discrete-time, finite-horizon ambiguous POMDP (APOMDP) approach, in which, at each patient’s visit, a decision maker (DM hereafter)—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. This is mainly due to 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 rejection 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. 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.

**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. Note that, after the first year, an immunosuppressive drug is prescribed at a maintenance level (i.e., fixed low dosage to sustain organ survival), and hence, the chance of NODAT drops drastically (see, e.g., Ghisdal et al. (2012)). Therefore, we consider one year post-transplant as our planning horizon ($N = 12$) which represents

---

3 We will show in §5.2.2 that the APOMDP approach results in more cost-effective policies than that obtained via a POMDP. The reader should note that gains compared to available robust MDP approaches (see, e.g., Iyengar (2005), Nilim and El Ghaoui (2005), and Xu and Mannor (2012)) are more trivial because such approaches do not allow for partial observability of state which is a key factor in our medical problem.
the time period during which medication management strategies are (a) most important, and (b) most variable among physicians.

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

**Observation state space:** $\mathcal{O} = \{\Delta, \nabla\} \cup \mathcal{O}$, 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 $i$th core state. For instance, $o_1$ is the observation that the patient is in $s_1$: medical tests suggest a low $C_0$ level ($C_0 \leq 8 \text{ mg/dL}$) while having organ survival and diabetic conditions (FPG$\geq126 \text{ mg/dL}$ or HbA1c$\geq6.5\%$).

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

**Ambiguity set (“cloud” of models):** $\mathcal{M} = \{m_1, m_2, \ldots, m_K\}$, 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, to construct this cloud of models.

**Core state transition probability:** $\mathbf{P}_m = \{\mathbf{P}_m^a : a \in \mathcal{A}\}$, where for each $a \in \mathcal{A}$, $\mathbf{P}_m^a = [p_m^a(j|i)]_{i,j \in \mathcal{S}}$, 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 $m \in \mathcal{M}$.  

**Observation probability:** $\mathbf{Q}_m = \{\mathbf{Q}_m^a : a \in \mathcal{A}\}$, where for each $a \in \mathcal{A}$, $\mathbf{Q}_m^a = [q_m^a(o|j)]_{j \in \mathcal{S}, o \in \mathcal{O}}$, 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:** $\Pi = \left\{ \mathbf{\pi} = [\pi_i]_{i \in \mathcal{S}} \in \mathbb{R}^{\left|\mathcal{S}\right|} : \sum_{i=1}^{\left|\mathcal{S}\right|} \pi_i = 1, \pi_1, \pi_2 \in \{0, 1\}, \pi_3, \ldots, \pi_{11} \in [0, 1] \right\}$, where $\mathbf{\pi}$ is an information vector over the state space $\mathcal{S}$. Since $\Delta$ (death) and $\nabla$ (organ rejection) are fully observable states, $\mathbf{\pi} = [1, \ldots, 0]'$ and $\mathbf{\pi} = [0, 1, \ldots, 0]'$ 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 $\mathbf{\pi} \neq [1, 0, \ldots, 0]'$ or $\mathbf{\pi} \neq [0, 1, \ldots, 0]'$, $\mathbf{b} = [0, 0, b_3, \ldots, b_{11}] = \mathbf{\pi}$ (i.e.,

---

4 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 dynamic consistency in general, there exist sufficient conditions under which this property is preserved in APOMDPs (see Saghaian (2017)).
<table>
<thead>
<tr>
<th>State</th>
<th>Health Condition (Transplant)</th>
<th>Health Condition (Diabetes)</th>
<th>Action</th>
<th>Prescription (Transplant)</th>
<th>Prescription (Diabetes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>s₁</td>
<td>Organ survival and low $C_0$</td>
<td>Diabetes (type II)</td>
<td>$a_1$</td>
<td>Use tacrolimus (with high dose)</td>
<td>Use insulin</td>
</tr>
<tr>
<td>s₂</td>
<td>Organ survival and medium $C_0$</td>
<td>Diabetes (type II)</td>
<td>$a_2$</td>
<td>Use tacrolimus (with medium dose)</td>
<td>Use insulin</td>
</tr>
<tr>
<td>s₃</td>
<td>Organ survival and high $C_0$</td>
<td>Diabetes (type II)</td>
<td>$a_3$</td>
<td>Use tacrolimus (with low dose)</td>
<td>Use insulin</td>
</tr>
<tr>
<td>s₄</td>
<td>Organ survival and low $C_0$</td>
<td>Pre-diabetes</td>
<td>$a_4$</td>
<td>Use tacrolimus (with high dose)</td>
<td>Do not use insulin</td>
</tr>
<tr>
<td>s₅</td>
<td>Organ survival and medium $C_0$</td>
<td>Pre-diabetes</td>
<td>$a_5$</td>
<td>Use tacrolimus (medium dose)</td>
<td>Do not use insulin</td>
</tr>
<tr>
<td>s₆</td>
<td>Organ survival and high $C_0$</td>
<td>Pre-diabetes</td>
<td>$a_6$</td>
<td>Use tacrolimus (with low dose)</td>
<td>Do not use insulin</td>
</tr>
<tr>
<td>s₇</td>
<td>Organ survival and low $C_0$</td>
<td>Healthy</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>s₈</td>
<td>Organ survival and medium $C_0$</td>
<td>Healthy</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>s₉</td>
<td>Organ survival and high $C_0$</td>
<td>Healthy</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

DM’s belief about $C_0$ and blood glucose levels in an alive patient without an organ rejection. We also let $\Pi_{PO}$ 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 $b$ under a model $m$ when action $a$ is taken and observation $o$ is made:

$$B(b, a, o, m) = \frac{(b'P^a_mQ^{a, o}_m)^t}{Pr\{o|b, a, m\}}.$$ (1)

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

$$Pr\{o|b, a, m\} = \sum_{i \in S} b_i \sum_{j \in S} p^a_m(j|i) q^o_m(o|j)$$ (2)

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

**Immediate reward:** $r_n(a) = [r_n(s, a) \geq 0]_{s \in S}$ for $a \in A$, where $r_n(s, a)$ is the QALE that a patient accrues when in state $s \in S$ 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 $a \in A$ and $s \in S$.

**Lump-sum reward:** $R_n = [R_n(s) \geq 0]_{s \in S}$, 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 $s \in S$.

**Ambiguity attitude set:** $\Lambda = \{\lambda: 0 \leq \lambda \leq 1\}$, where $\lambda$ 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 $\pi$, DM’s conservatism level $\lambda$, and any period $n \leq N$, we have:

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

where

$$V_n(b,\lambda) = \begin{cases} b' R_n, & \text{if } n = N \\ \max_{a \in A} \{ U_n(b,a,\lambda) \}, & \text{if } n < N. \end{cases}$$ (4)

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

$$U_n(b,a,\lambda) = b' r_n(a) + \lambda \min_{m \in M} \left\{ H_n(b,a,m,\lambda) \right\} + (1 - \lambda) \max_{m \in M} \left\{ H_n(b,a,m,\lambda) \right\},$$ (5)

where

$$H_n(b,a,m,\lambda) = \beta \sum_{o \in \Theta} Pr\{o|b,a,m\} V_{n+1}\left(B(b,a,o,m),\lambda\right).$$ (6)

The first term in the RHS of (5) represents the expected current “reward” (in QALE) in period $n$ when the belief vector is $b$, the action is $a$, the model is $m$, and the DM’s conservatism level is $\lambda$. 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 $\lambda$ 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) $\lambda$ represents the attitude of a more (less) pessimistic/conservative DM. By varying $\lambda$, our framework allows us to capture the behavioral attitudes of physicians, and evaluate their effects on the intensity of medications administered.5

Finally, we define the worst model and the best model in period $n$ as the minimizer and maximizer of $H_n(b,a,m,\lambda)$ defined in (6), respectively:

$$m_n(b,a,\lambda) = \arg \min_{m \in M} \left\{ H_n(b,a,m,\lambda) \right\},$$ (7a)

$$\bar{m}_n(b,a,\lambda) = \arg \max_{m \in M} \left\{ H_n(b,a,m,\lambda) \right\}.$$ (7b)

For the ease of notation, we may refer to these worst and best models as $m$ and $\bar{m}$, respectively.

5 We note that setting $\lambda = 1$ represents an extension of existing robust dynamic programming approaches (see, e.g., Iyengar (2005), Nilim and El Ghaoui (2005)) to settings with partially observable states.
4. Structural Results

We now establish some structural properties, which allow us to analyze our APOMDP model, and thereby gain insights into the simultaneous management of post-transplant medications.

**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 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 make use of the definition of a belief-independent worst-case (BIWC) member in the cloud of models $\mathcal{M}$:

**Definition 1** (Saghafian 2017). $m_n(b,a,\lambda) \in \mathcal{M}$ defined in (7a) is said to be a BIWC member of $\mathcal{M}$, if it is constant in the belief vector $b$.

This implies that, irrespective of the DM’s belief about a patient’s health state, there exists a set of transition and observation matrices (under any action and conservatism level) that yields the least total reward (in QALE). If such a model exists in $\mathcal{M}$, then the optimal value function is PLC in the belief vector $b$ (see Proposition 2 in Saghafian (2017)), and hence, can be written as:

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

where $\Psi_{n,\lambda}$ is some finite set. Equation (8) is analogous to the use of POMDPs proposed by Smallwood and Sondik (1973). Based on (8), 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 $a \in A$, $P_m^nQ_m^n = P_{\hat{m}}^nQ_{\hat{m}}^nW$ for some stochastic matrix $W$, then model $m$ is said to be less informative than model $\hat{m}$. 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)). Utilizing our clinical data set in Online Appendix B.4, we discuss scenarios where the model informativeness condition (and thus the existence of a BIWC member) is satisfied in our setting. In other settings where this property does not hold, one can extend the ambiguity set so that it includes a BIWC member. This will substantially reduce the underlying computational complexity by ensuring that (8) holds, and can provide a close approximation.\(^7\)

\(^6\)For notational simplicity, we suppress the dependency on $a$, and write $m(a)$ as $m$ for any $m \in \mathcal{M}$.

\(^7\)Of note, since the state is continuous (a belief/probability vector) in our setting, it is expected that additional conditions such as existence of a BIWC member are needed for reducing the computational complexity compared to
Assuming that $M$ is such that it has a BIWC member, we now establish a closed-form analytical representation for the set of $\psi$-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.

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

$$\Psi_{N,\lambda} = \{R_N\} \forall \lambda \in \Lambda,$$

$$\Psi_{n,\lambda} = \left\{ \psi \in \mathbb{R}^{|\mathcal{A}|} : \psi = r_n(a) + \lambda \left( \beta \sum_{o \in \mathcal{O}} \mathbf{P}^a_m Q_m^{\alpha_o} \psi^{(b,a,o)}_m \right) + (1 - \lambda) \left( \beta \sum_{o \in \mathcal{O}} \mathbf{P}^a_m Q_m^{\alpha_o} \psi^{(b,a,o)}_m \right), \right.$$

$$\left. a \in \mathcal{A}, \psi^{(b,a,o)}_m, \psi^{(b,a,o)}_m \in \Psi_{n+1,\lambda} \right\} \forall \lambda \in \Lambda, \forall n < N,$$

where

$$\psi^{(b,a,o)}_m = \arg \max_{\psi \in \Psi_{n+1,\lambda}} \{ b^T \mathbf{P}^a_m Q_m^{\alpha_o} \psi \} \forall b \in \Pi_{PO}, \forall a \in \mathcal{A}, \forall m \in \mathcal{M}, \forall o \in \mathcal{O}. \quad (11)$$

The characterization of the set of $\psi$-vectors in Proposition 1 depends on identifying both models $\underline{m}$ and $\overline{m}$. Although $\underline{m}$ can be obtained in the ambiguity set $M$ 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 $\psi$-vectors:

$$\widetilde{\Psi}_{n,\lambda} = \left\{ \widetilde{\psi} \in \mathbb{R}^{|\mathcal{A}|} : \widetilde{\psi} = r_n(a) + \lambda \left( \beta \sum_{o \in \mathcal{O}} \mathbf{P}^a_m Q_m^{\alpha_o} \widetilde{\psi}^{(b,a,o)}_m \right) + (1 - \lambda) \left( \beta \sum_{o \in \mathcal{O}} \mathbf{P}^a_m Q_m^{\alpha_o} \widetilde{\psi}^{(b,a,o)}_m \right), \right.$$

$$\left. a \in \mathcal{A}, m \in \mathcal{M} \setminus \{\underline{m}\}, \widetilde{\psi}^{(b,a,o)}_m, \widetilde{\psi}^{(b,a,o)}_m \in \widetilde{\Psi}_{n+1,\lambda} \right\} \forall \lambda \in \Lambda, \forall n < N. \quad (12)$$

Then, $\Psi_{n,\lambda}$ in (10) can be obtained from $\widetilde{\Psi}_{n,\lambda}$ in (12) by applying the Monahan’s algorithm (Monahan 1982). The equation in (12) implies that, even if we consider all models in $\mathcal{M} \setminus \{\underline{m}\}$, using the Monahan’s algorithm, we can shrink the set of the $\psi$-vectors to those attributed only to $\underline{m}$ and $\overline{m}$.

**Effect of DM’s Conservatism Level on Drug Intensification.** As noted earlier, the DM’s conservatism (i.e., ambiguity attitude) may affect the intensification of medications regimens. To study this phenomenon, we start by defining the following conditions. In Online Appendix B.6, we also numerically test the validity of conditions in this section using our data set, and discuss whether and when such conditions hold.

the robust MDP approaches (see, e.g., Iyengar (2005) and Nilim and El Ghaoui (2005)) where partial observability is not allowed.
CONDITION 1 (Monotonicity of Reward). (i) Under any action \( a \in \mathcal{A} \), the immediate reward vector \( r_n(a) \) is nondecreasing in state \( s \in \mathcal{S} \), and (ii) the lump-sum reward vector \( R_n \) is nondecreasing in state \( s \in \mathcal{S} \) (i.e., better health states have higher immediate and lump-sum rewards).

CONDITION 2 (\( TP_2 \) Transitions). For all \( m \in \mathcal{M} \) and \( a \in \mathcal{A} \), the kernels \( P_m^n \) and \( Q_m^n \) are \( TP_2 \) (i.e., all their second-order minors are non-negative).

Here, we define the well-known \( TP_2 \) stochastic ordering between two belief vectors. Since each belief vector \( b \) is a probability mass function, \( TP_2 \)-ordering (shown as \( \preceq_{TP_2} \)) is equivalent to the weak monotone likelihood ratio (MLR) ordering:

**Definition 2 (Whitt 1982).** A belief vector \( b \) is said to be dominated by another belief vector \( \hat{b} \) in the MLR-ordering sense (shown as \( b \preceq_{MLR} \hat{b} \)) if the ratio \( \hat{b}/b \) is nondecreasing in its elements.

From the medical standpoint, Definition 2 implies that a patient with associated belief vector \( \hat{b} \) is more likely to be in a better health state than another patient with associated belief vector \( b \).

Now, based on the \( TP_2 \)-ordering, we define the following condition:

**Condition 3.** Fix \( n < N \) and \( b \in \Pi_{PO} \), and let \( \overline{m}(a,\lambda) = m_n(b,a,\lambda) \) and \( K_m^n(a|i) = \sum_{j \in \mathcal{S}} p_m^n(j|i) q_m^n(o|j) \). For all \( a, \hat{a} \in \mathcal{A} \) and \( \lambda, \hat{\lambda} \in \Lambda \) such that \( a \leq \hat{a} \) and \( \lambda \leq \hat{\lambda} \): (i) \( \hat{K}_{m(a,\lambda)}^a(\sigma|i) \leq K_m^n(a|i) \) and \( \hat{K}_{m(\mathcal{A},\lambda)}^a(\sigma|i) \geq K_m^n(a,\lambda)(\sigma|i), \forall i \in \mathcal{S}, o \in \mathcal{O}, \) and (ii) \( B(b,\hat{a},o,\overline{m}(\hat{a},\hat{\lambda})) \preceq_{TP_2} B(b,a,o,\overline{m}(a,\lambda)), \forall o \in \mathcal{O}.\)

To understand Condition 3 consider two DMs, DM 1 and DM 2, with conservatism levels \( \lambda \) and \( \hat{\lambda} \), respectively, where \( \lambda \leq \hat{\lambda} \). Assume these DMs are prescribing medication regimens under the best model in the cloud of models. Then, Condition 3(i) implies that for DM 1 (2) the probability of having any observation (for any fixed core health state) decreases (increases) with the intensity of medication regimens. Condition 3(ii) implies that DM 1 (2) has a worse (better) updated belief about a patient’s health state (in the \( TP_2 \) sense) when taking more intensive (than less intensive) medications.

**Theorem 1 (Effect of \( \lambda \) on Drug Intensification).** For all \( b \in \Pi_{PO}, \lambda \in \Lambda, \) and \( a, \hat{a} \in \mathcal{A} \) such that \( a \leq \hat{a} \), suppose \( \min_{m \in \mathcal{M}} \{H_n(b,a,m,\lambda)\} = \min_{m \in \mathcal{M}} \{H_n(b,\hat{a},m,\lambda)\}. \) Also, let \( a_n^*(b,\lambda) \) be the optimal medication action for period \( n \) when the belief vector and conservatism levels are \( b \) and \( \lambda \), respectively. Then, (i) under Conditions 1–3, \( \lambda \leq \hat{\lambda} \) yields \( a_n^*(b,\lambda) \geq a_n^*(b,\hat{\lambda}) \), and (ii) under Conditions 1–2 and reverse of Condition 3, \( \lambda \leq \hat{\lambda} \) yields \( a_n^*(b,\lambda) \geq a_n^*(b,\hat{\lambda}) \).

\(^8\)The reverse of this condition is obtained when the foregoing inequalities/orderings are reversed.
The condition in Theorem 1 implies that the model that yields the lowest reward (in QALE) depends only on the belief \( b \) and conservatism level \( \lambda \) (but not on the action \( a \)). Theorem 1(i) provides insights into conditions under which a higher level of conservatism results in a more intensive medication regimen (under the optimal policy). 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) sheds light on conditions under which a higher level of conservatism results in a less intensive medication regimen (under the optimal policy). This can happen because the DM should be more concerned about the risk of NODAT than an organ rejection. Thus, the DM should avoid prescribing high dosages of tacrolimus because of its diabetogenic effect. Similarly, it can happen because the DM should prefer not to use insulin. In our numerical experiments in §5.2.1 we will dig deeper using our data set, and examine how the intensity of optimal medications regimens changes based on the DM’s conservatism level, and how such changes depend on the patient’s risk factors.

**Monotonicity of the Optimal Medication 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 4.** Suppose the value function is PLC and define vectors 
\[
\phi^{(b,a)}_m = \sum_{o \in O} \mathbf{P}_m^{a,o} \mathbf{Q}_m^{a,o} \psi^{(b,a,o)}_m
\]
(for all \( b \in \Pi_{PO}, a \in A, \) and \( m \in M \)), where \( \psi^{(b,a,o)}_m \) is defined in (11). Then, for any \( a, \hat{a} \in A \) such that \( a \preceq \hat{a} \) and \( \lambda \in \Lambda \), vectors 
\[
\phi^{(b,\hat{a})}_m - \phi^{(b,a)}_m \text{ and } \phi^{(b,\hat{a})}_m - \phi^{(b,a)}_m
\]
are nondecreasing in their elements.

Conditions 4 implies that, when taking a less intensive medication regimen compared to a more intensive one, the resulted difference between the reward to-go (in QALE) is nondecreasing along core health states.

**Theorem 2 (Monotone Optimal Medication Policy).** Let \( a^*_n(b,\lambda) \) be the optimal medication action for period \( n \). Then, under Condition 4, \( b \preceq_{TP_2} \hat{b} \) yields \( a^*_n(b,\lambda) \preceq a^*_n(\hat{b},\lambda) \).

Theorem 2 simplifies the search for an optimal medication policy. For instance, consider two patients, patients 1 and 2, where patient 2 is believed to be in a better health condition than patient 1 (in the \( TP_2 \) sense). Then, 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 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.

**Bounds for the Value Function.** For our numerical experiments, we solve our APOMDP model optimally based on Proposition 1. However, the time complexity of finding an optimal policy (at any period \( n \) and for any conservatism level \( \lambda \)) is \( O(|M||A||\Psi_n|^{\theta} \lambda) \) (see, e.g., Hauskrecht (2000)).\(^9\) Although we alleviate this effect by implementing the Monahan’s algorithm (Monahan 1982) to eliminate dominated \( \psi \)-vectors, to further streamline computational burdens, we now develop a bound for the value function in (4). We let \( J_n(b, \lambda) \) be the approximate value function, and \( a^*J_n(b, \lambda) \) be its corresponding action (denoted by \( a^J \) for the ease of notation). In the optimal value function \( V_n(b, \lambda) \), the DM computes the expected future reward based on his/her updated belief about the patient’s health state (i.e., expected reward-to-go). However, in the approximate value function \( J_n(b, \lambda) \), the DM first obtains his/her expected belief (over all updated belief vectors), and then the reward based on the expected belief:

\[
J_n(b, \lambda) = b^r_n(a^J) + \lambda \min_{m \in M} \left\{ \beta J_{n+1} \left( b^\prime P_m^{a^J}, \lambda \right) \right\} + (1 - \lambda) \max_{m \in M} \left\{ \beta J_{n+1} \left( b^\prime P_m^{a^J}, \lambda \right) \right\},
\]

where we obtain \( b^\prime P_m^{a^J} \) from \( \sum_{o \in O} Pr\{o|b, a^J_m\} B(b, a^J, o, m) \) by following the Bayesian update in (1) and the fact that \( \sum_{o \in O} Q_n^{a^J} = \mathbb{I} \), where \( \mathbb{I} \) is an identity matrix. Proposition 2 shows that the optimal value function \( V_n(b, \lambda) \) is tightly bounded from below by the approximate value function \( J_n(b, \lambda) \). Before establishing this result, we need the following lemma.

**Lemma 1 (\( \psi \)-Vectors Bound).** Suppose \( r_n(s) \leq \epsilon_q (\forall s \in \mathcal{S}, \forall n < N) \) and \( R_N(s) \leq \epsilon_r (\forall s \in \mathcal{S}) \) for some \( \epsilon_q, \epsilon_r \geq 0 \). Then, for each \( \psi \)-vector, when there are \( k \) periods remaining until the end of the horizon, we have \( \psi_i \leq \epsilon_{N-k} = \epsilon_q \sum_{l=0}^{k-1} \beta^l + \epsilon_r \beta^k \) for all \( i \in \mathcal{S} \).

In Lemma 1, \( \epsilon_q \) is a bound for the quality of life (QOL) score, which is a score between 0 and 1. Similarly, \( \epsilon_r \) is a bound on the lump-sum reward, which is a function of residual life expectancy and a discount rate, such that as the discount rate approaches 1, the lump-sum reward approaches QOL (see §5.1 for more details regarding these reward parameters). We note that the bound provided by Lemma 1 is relatively tight. For example, it goes to 0 as \( \beta \to 0 \), and to \( (k\epsilon_q + \epsilon_r) \) as \( \beta \to 1 \). Furthermore, for \( \beta \in [0, 1) \), this bound asymptotically approaches \( \frac{\epsilon_q}{1-\beta} \) as \( k \to \infty \).

\(^9\) See also Papadimitriou and Tsitsiklis (1987) for a discussion about the time and space complexities of (PO)MDPs.
**Proposition 2 (Performance Bound).** Suppose (i) the ambiguity set $\mathcal{M}$ has a BIWC member, (ii) $|p^a_m(j|i) - p^a_\widehat{m}(j|i)| \leq \eta$ for some $\eta \geq 0$ \((\forall a \in \mathcal{A}, \forall m, \widehat{m} \in \mathcal{M}, \forall i,j \in \mathcal{S})\), and (iii) $\tau$ is the maximum possible reward in each period. Then, we have:

$$V_n(b, \lambda) - J_n(b, \lambda) \leq \min \left\{ \frac{\beta \eta \epsilon_{n+1} |\mathcal{S}|}{1 - \beta} \left[ \tau \sum_{l=0}^{N-1} \beta^l \right] \right\} \quad \forall b \in \Pi_{PO}, \forall \lambda \in \Lambda, \forall n < N,$$

where $\epsilon_{n+1}$ is obtained from Lemma 1 by setting $k = N - n - 1$.

Proposition 2 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 (14). We note that, under the following conditions, $J_n(b, \lambda)$ converges to $V_n(b, \lambda)$, making the performance bound in (14) 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), $\eta$ approaches 0, (2) when $\beta \in [0, 1)$ and the time horizon increases, $\epsilon_{n+1}$ asymptotically approaches $\frac{\epsilon_0}{1 - \beta}$, which, in turn, approaches 0 as a patient’s health status gets aggravated (see the discussion after Lemma 1), and (3) when $\beta$ approaches 0 (i.e., the DM decides upon medications regimens in a myopic approach). Furthermore, when $\beta$ approaches 1, the performance bound in (14) approaches $N\tau$ which is small when $N$ or $\tau$ is small. In general, the bound in (14) 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 the following elements from 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, the estimation of the reward functions (in QALE), and the mechanism used to validate our estimated parameters. We then describe the results we have obtained from our numerical experiments, and shed light on their 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
Table 2 Description of main risk factors and their levels (see also Boloori et al. (2015))

<table>
<thead>
<tr>
<th>Risk Factor (Abbreviation)</th>
<th>Unit</th>
<th>Low Level</th>
<th>High Level</th>
<th>Static (S)/Dynamic (D)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>Years</td>
<td>&lt;50</td>
<td>≥50</td>
<td>S</td>
</tr>
<tr>
<td>Gender</td>
<td>—</td>
<td>Female</td>
<td>Male</td>
<td>S</td>
</tr>
<tr>
<td>Race</td>
<td>—</td>
<td>White</td>
<td>non-White</td>
<td>S</td>
</tr>
<tr>
<td>Diabetes history (Diab Hist)</td>
<td>—</td>
<td>No</td>
<td>Yes</td>
<td>S</td>
</tr>
<tr>
<td>Body mass index (BMI)</td>
<td>kg/m²</td>
<td>&lt;30 (non-obese)</td>
<td>≥30 (obese)</td>
<td>D</td>
</tr>
<tr>
<td>Blood pressure (BP)</td>
<td>—</td>
<td>Normal</td>
<td>Hypertension</td>
<td>D</td>
</tr>
<tr>
<td>Total cholesterol (Chol)</td>
<td>mg/dL</td>
<td>&lt;200</td>
<td>≥200</td>
<td>D</td>
</tr>
<tr>
<td>High-density lipoporten (HDL)</td>
<td>mg/dL</td>
<td>≥40</td>
<td>&lt;40</td>
<td>D</td>
</tr>
<tr>
<td>Low-density lipoporten (LDL)</td>
<td>mg/dL</td>
<td>&lt;130</td>
<td>≥130</td>
<td>D</td>
</tr>
<tr>
<td>Triglyceride (TG)</td>
<td>mg/dL</td>
<td>&lt;150</td>
<td>≥150</td>
<td>D</td>
</tr>
<tr>
<td>Uric acid (UA)</td>
<td>mg/dL</td>
<td>&lt;7.3</td>
<td>≥7.3</td>
<td>D</td>
</tr>
</tbody>
</table>

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 2 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.\(^\text{10}\) (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.\(^\text{11}\) (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).

**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 models are provided in detail in Table 3. In these steps: (1) we use the Baum-Welch (BW) algorithm

\(^\text{10}\) 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.

\(^\text{11}\) 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.
(Welch 2003) to obtain point estimations for core state and observation probability matrices (steps 5–11 in Table 3). 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 3). We do so by using the Kullback-Leibler (KL) divergence criterion (a.k.a. relative entropy):

\[
d_{KL}(v_1||P_{BW}^a(i)) = \sum_{j \in S} v_1(j) \log_2 \left( \frac{v_1(j)}{p_{BW}^a(j|i)} \right) \text{ for } v_1 \in \Theta_1 \quad \forall i \in S, \forall a \in A, \tag{15a}
\]

\[
d_{KL}(v_2||Q_{BW}^a(i)) = \sum_{o \in O} v_2(o) \log_2 \left( \frac{v_2(o)}{q_{BW}^a(o|i)} \right) \text{ for } v_2 \in \Theta_2 \quad \forall i \in S, \forall a \in A, \tag{15b}
\]

where for each \(a \in A\), \(P_{BW}^a = [p_{BW}^a(j|i)]_{i,j \in S}\) and \(Q_{BW}^a = [q_{BW}^a(o|i)]_{i,j \in S, o \in O}\) are the point estimates returned by the BW algorithm. or 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 \(\epsilon\) so as to ensure that the cloud of models has a BIWC member (see Definition 1).

In Online Appendix B.4, we discuss scenarios where the model informativeness condition (and thus the existence of a BIWC member) is satisfied in our setting. Also, in Online Appendix B.5, we provide details on how we validate our estimations of the set of transition and observation probability matrices.

**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 \(s \in S\), and the action taken \(a \in A\). 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 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 \(a \in A\) and \(n \leq N - 1\):

\[
r_n(s^T,a) = \langle qol(\text{Organ rejection})/12, qol(\text{Organ survival})/12 \rangle \quad \text{if } s^T = \text{Organ rejection, Organ survival (different } C_0 \text{’s)},
\]

\[
r_n(s^D,a) = \langle qol(\text{Diabetic})/12, qol(\text{Pre-diabetic})/12, qol(\text{Healthy})/12 \rangle \quad \text{if } s^D = \text{Diabetic, Pre-diabetic, Healthy}.\tag{16b}
\]

In (16a)-(16b), 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).

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:

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

(17)

where $r$ is a discount rate which accounts for degradation of the core health state over the remaining lifetime of a patient. In (17), $qol(s) = \langle qol(s^T), qol(s^D) \rangle$, 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 (16a)-(16b).

Further details about estimating the required parameters (e.g., $qol$ and $RLE$ scores) can be found in Online Appendix B.2. When comparing our optimal policies with other benchmarks in §5.2.2, we perform sensitivity analyses on the estimated reward parameters by changing the values of $qol$ and $RLE$ (see Online Appendix E). Moreover, although in our base estimates we assign an equal weight to diabetes and organ rejection outcomes (by taking the average of their related rewards), in our sensitivity analyses (Online Appendix E), we consider different values for $qol$ and $RLE$ such that diabetes outcomes can have a higher or lower impact compared to organ rejection outcomes.

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 current medical practice (§5.2.2). As we will discuss, these results have important implications for policy makers as well as individual physicians and patients.
5.2.1. Robust Optimal Medication Policies. We obtain optimal medication policies from our APOMDP approach separately for 22 cohorts of patients based on the risk factors described in Table 2. 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., $|\mathcal{M}| = 3$). To illustrate our results, we use a 2-simplex to represent 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 $b_3, b_6, b_9 \neq 0$ and $b_4, b_5, b_7, b_8, b_{10}, b_{11} = 0$ (i.e., the patient is alive and 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 $\Pi_{PO}$, which is an 8-simplex, we choose these cuts to understand the interaction of two medications under different risks of organ rejection and diabetes complications. Below, we first summarize the following main observation from our results, and then discuss its implications for the medical practice.

Observation 1 (Optimal Medication Policies). (i) Under low or medium $C_0$, when the perceived risk of diabetes is low, the optimal action is to use “high-dose tacrolimus.” However, as the perceived diabetes risk is elevated, using “medium” or “low-dose tacrolimus” becomes optimal. (ii) Under the optimal policy, when $C_0$ is high, not only patients with higher perceived risks of diabetes, but also those with higher perceived chance of being healthy are typically prescribed by “low-dose tacrolimus.” (iii) Under the optimal policy, when the perceived risk of (pre-)diabetes is comparable to the chance of being healthy, cohorts of patients with high-risk levels (see Table 2) are typically prescribed by the same or more intensive tacrolimus/insulin regimens (compared to low-risk levels). Part (iii) consistently holds under different levels of $C_0$.

To better understand Observation 1, let us consider 4 different patients each corresponding to a specific belief point in the above-mentioned 2-simplex. These patients are identified in Figure 2 via vectors $\tilde{b}$ that belong to the interior of the simplex (and hence, are three-dimensional and have positive elements summing to one). For example, patient 1 has $\tilde{b} = [0.80, 0.15, 0.05]$ (i.e., 80% perceived risk of having diabetes, 15% perceived risk of having pre-diabetes, and 5% perceived risk of being healthy). Patients 1, 2, and 3 have a high risk of being diabetic, pre-diabetic, and healthy, respectively. Patient 4 has an equal risk among these three conditions. Regarding Observation 1(i), Figure 2 (see also Figure EC.2 in Appendix C) shows that, when $C_0$ is low, both patients 2 and 3 (who have considerably lower perceived risks of diabetes compared to patient 1) are prescribed by

\[\text{\footnotesize \cite{12}}} Figure 2 depicts the results for a mid level of conservatism ($\lambda = 0.5$). Later, in Observation 3, we will provide a discussion on the effect of variations in conservatism level on the intensity of optimal medication regimens.
“high-dose tacrolimus.” However, patient 1 is prescribed by “medium-dose tacrolimus.” Furthermore, when $C_0$ is medium, patients 2 and 3 are still put on “high-dose tacrolimus,” whereas the typical optimal action for patient 1 is to use “low-dose tacrolimus.” This suggests that, when a patient has a low level of $C_0$ (and thus an organ rejection is likely), s/he should be put on “high-dose tacrolimus” but only as long as his/her perceived risk of diabetes is not high. Otherwise, to prevent from further diabetes complications (due to the diabetogenic effect of tacrolimus), the patient should be put on a low/medium dosage of tacrolimus. Regarding Observation 1(ii), Figure 2 (and Figure EC.2 in Appendix C) reveal that, when $C_0$ is high, patients 1 and 3 are typically put on “low-dose tacrolimus,” while the difference is in whether or not to use “insulin.” This implies that, when an organ rejection is least likely (i.e., high $C_0$ compared to low/med $C_0$), having different perceived risks of diabetes may not result in different tacrolimus regimens (unlike the case in Observation 1(i)). We also observe that patient 2 (who has a high perceived risk of pre-diabetes) is typically prescribed by “high-dose tacrolimus” under the optimal policy when having high $C_0$. Although we cannot directly interpret the latter, one should note that the area for which using “high-dose tacrolimus” is optimal is very small in the vicinity of patient 2. In other words, any increase in risk of diabetes or chance of being healthy can result in a different optimal tacrolimus regimen (e.g., “med-dose tacrolimus”).

Regarding Observation 1(iii), Figure 2 shows that patient 4 (who is equally likely to be diabetic, pre-diabetic, or healthy) typically receives a more intensive medication regimen when his/her risk factors (see Table 2) are high. For example, when $C_0$ is low (i.e., an organ rejection is more likely), if
patient 4 is female or has normal HDL/LDL or no diabetes history, the optimal action is to use “high-dose tacrolimus” and “no insulin.” However, if patient 4 is male, has low HDL, high LDL, or diabetes history, then he should be put on insulin while maintaining a high tacrolimus dose. Furthermore, when $C_0$ is medium (i.e., an organ rejection is less likely), being female, or having no diabetes history, White race, or normal cholesterol/HDL results in “med-dose tacrolimus” and “no insulin” for patient 4, whereas being male, or having diabetes history, non-White race, or high cholesterol yields “high-dose tacrolimus” and “insulin” for the same patient. An implication for the medical practice is that, when the DM is less certain about the patient conditions (e.g., patient 4 which has an equal perceived risk of being diabetic, pre-diabetic, and healthy compared to other patients), the role of risk factors becomes more important in identifying optimal actions. Furthermore, patients with high levels of risk factors are more dependent on intensive medication regimens. Of note, although there exist differences in the optimal medication regimens depending on risk factors, the observations above are common across different cohorts of patients. The only exception is for the case when the risks of organ rejection and (pre-)diabetes are low (i.e., approaching $e_3$ under high $C_0$). In this case, a patient may be prescribed by high (instead of low) dosage of tacrolimus. However, this happens primarily for high-risk cohorts (e.g., patients with obesity, hypertension, and high cholesterol).

Observation 2 (Tacrolimus Requirement and the Diabetogenic Effect). Under any conservatism level, (i) higher dosage of tacrolimus should be prescribed more for patients with $\text{age} \geq 50$, male gender, non-White race, hypertension, high BMI and LDL, and low HDL than patients with the opposing risk levels, and (ii) patients who are prescribed by medium/high dosage of tacrolimus become more dependent on using insulin under the following risk factors: age $\geq 50$, male gender, diabetes history, hypertension, high Chol, and low HDL.

Regarding Observation 2(i), Figure 2 (and Figure EC.2 in Appendix C) show that the area for which actions $a_1$ and $a_4$ (i.e., using high-dose tacrolimus) are optimal is larger for patients with $\text{age} \geq 50$ than for other patients. A similar effect can be observed for the other risk factors in Observation 2(i). 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(i) suggests that the dosage of tacrolimus should be adjusted based on many other static or dynamic risk factors of patients. In particular, Observation 2(i) implies that age, gender, race, BMI, blood pressure, HDL, and LDL 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 total cholesterol, 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. Furthermore, regarding Observation 2(ii), Figure 2 (and Figure EC.2 in Appendix C) show that the policy regions for actions $a_1$ and $a_2$ (i.e., using insulin along with medium/high dosage of tacrolimus) is greater are larger for patients with age $\geq 50$ than for other patients. A similar effect can be observed for the other risk factors in Observation 2(ii). Observation 2(ii) 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.

**Observation 3 (The Effect of Conservatism Levels).** Increasing the conservatism level, $\lambda$, results in using (i) more intensive medication regimens for patients with age $<50$, age $\geq 50$, non-White race, no diabetes history, low-risk levels of Chol, HDL, LDL, TG, and UA, 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 Chol, HDL, and LDL. However, increasing $\lambda$ does not change the intensity of medication regimens for patients with White race, female gender, normal blood pressure, and high-risk levels of TG and UA.

For example, as can be observed from Figure 3 (parts (a) and (b)), for a patient with age $<50$ or non-White race, a higher conservatism level in the same time period typically results in a larger area where the optimal action is to prescribe a more intensive medication regimen (e.g., high-dose tacrolimus as opposed to medium-dose tacrolimus). On the other hand, based on Figure 3 (parts (c) and (d)), for a patient with diabetes history or hypertension, increasing the conservatism level in the same time period can yield a smaller area where the optimal action is to prescribe a more intensive medication regimen.

These results support the theoretical insights we gained on the impact of $\lambda$ on drug intensification (see Theorem 1 in §4). In particular, Observation 3 implies that, for patients with non-White race, no diabetes history, and normal levels of Chol, HDL, LDL, TG, and UA, a more conservative DM should be more concerned about the risk of organ rejection than the potential risk of NODAT compared to a less conservative DM. On the other hand, for patients with male gender, diabetes history, hypertension, and high-risk levels of Chol, HDL, and LDL, this result is reversed. Furthermore, for patients with White race, female gender, normal blood pressure, and high-risk levels of TG and UA, increasing the conservatism level does not drastically affect the intensity of prescribed medications
under the optimal policy. This, in turn, implies that, for these cohorts, the DM should be equally concerned about risks of organ rejection and diabetes complications.

Observation 3 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.

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) four benchmark policies that resemble the current medical practice under different scenarios, and (2) a policy that is obtained by a traditional POMDP (i.e., by ignoring the underlying ambiguities, and considering only one model at a time). Comparing our approach to the policies in (1) reveals the potential 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 Policies. In the current medical practice, the outcomes of medical tests (observations) are treated as the actual health state of the patient (see, e.g., Bennett et al. (2007)), based on which physicians prescribe medication regimens. Furthermore, tacrolimus is typically administered based on a combination of an observation (i.e., $C_0$ level) and time elapsed post-transplant. However, there is currently no consensus among physicians on how $C_0$ level and elapsed time should be incorporated in prescribing tacrolimus. To address this variation among physicians, we consider four different benchmark policies that are typically used in the current practice (see Table 4). As

---

Figure 3 Variations in optimal medication policies (in the same period) based on different conservatism levels ($e_1, e_2, e_3$ represent diabetic, pre-diabetic, and healthy conditions, respectively; $e_j$ denotes a unit vector with $j^{th}$ element equal to 1 and other elements equal to 0)
Table 4 shows, for the first three months post-transplant, tacrolimus is prescribed in high dosage in all of these four benchmark policies. This is consistent with the fact that in the current practice patients are consistently kept on high levels of tacrolimus during the first months post-transplant (see, e.g., Ghisdal et al. (2012)) so as to avoid organ rejection (i.e., by maintaining a high chance of organ survival). However, after the first three months, the four policies differ: benchmark 1 (4) represents the most (least) intensive policy for prescribing tacrolimus. For example, when the patient is observed to have medium $C_0$ (i.e., observations $o_2$, $o_5$, or $o_8$) during months 4-6 post-transplant, the regimen under benchmark 1 is to use high dosage of tacrolimus (i.e., actions $a_1$ or $a_4$), whereas the regimen under benchmark 4 is to use medium dosage of tacrolimus (i.e., actions $a_2$ or $a_5$). Moreover, consistent with the current practice, in all four benchmark policies, insulin is not prescribed for a patient who is observed to be diabetic free (i.e., a patient with FPG $< 126$ mg/dL or HbA1C $< 6.5$%).

We have developed and used a micro-simulation model (see Online Appendix D) to compare the optimal policy from the APOMDP approach with both the four benchmark policies and with the POMDP policy. Tables 5 and 6 show the results of our comparisons based on three performance measures: (1) average QALE achieved, (2) average medical expenditures (see Online Appendix B.3 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 current practice (i.e., whether or not our proposed policies alleviate the drug intensification problem observed in the current practice). Furthermore, since dynamic risk factors are subject to change throughout the time horizon, in this micro-simulation we allow each dynamic risk factor to take either a low or a high level in each period (i.e., unlike static risk factors, we do not run the simulation for each low/high-risk levels of dynamic risk factors, separately). Considering seven dynamic and four static risk factors in our study, we therefore have $7 + 4 \times 2 = 15$ (and not 22) cohorts of patients in Tables 5 and 6. We make the following observations from the results presented in Tables 5 and 6:

**Observation 4 (Impact).** During one year post-transplant, our optimal policy (i) improves the QALE per patient on average by 4.58%, 3.46%, 0.83%, 0.79%, and 0.35% compared to benchmark
Table 5: Comparison of medication policies (based on avg. QALE and cost)

<table>
<thead>
<tr>
<th>Measure</th>
<th>Cohort</th>
<th>Policy</th>
<th>Improvement of Optimal policy over†</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bench 1</td>
<td>Bench 2</td>
<td>Bench 3</td>
</tr>
<tr>
<td>QALE(yrs)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age&lt;50</td>
<td>16.36</td>
<td>16.56</td>
<td>16.98</td>
</tr>
<tr>
<td>Age≥50</td>
<td>9.15</td>
<td>9.33</td>
<td>9.80</td>
</tr>
<tr>
<td>Gender:Female</td>
<td>17.74</td>
<td>17.88</td>
<td>18.07</td>
</tr>
<tr>
<td>Gender:Male</td>
<td>15.35</td>
<td>15.51</td>
<td>15.88</td>
</tr>
<tr>
<td>Race:White</td>
<td>16.16</td>
<td>16.25</td>
<td>16.58</td>
</tr>
<tr>
<td>Race:non-White</td>
<td>13.25</td>
<td>13.41</td>
<td>14.04</td>
</tr>
<tr>
<td>DiabHist:No</td>
<td>14.66</td>
<td>14.76</td>
<td>15.01</td>
</tr>
<tr>
<td>DiabHist:Yes</td>
<td>8.32</td>
<td>8.52</td>
<td>8.73</td>
</tr>
<tr>
<td>BMI</td>
<td>13.65</td>
<td>13.79</td>
<td>14.10</td>
</tr>
<tr>
<td>BP</td>
<td>13.34</td>
<td>13.52</td>
<td>13.82</td>
</tr>
<tr>
<td>Chol</td>
<td>13.06</td>
<td>13.18</td>
<td>13.56</td>
</tr>
<tr>
<td>HDL</td>
<td>13.20</td>
<td>13.37</td>
<td>13.86</td>
</tr>
<tr>
<td>LDL</td>
<td>13.22</td>
<td>13.49</td>
<td>13.55</td>
</tr>
<tr>
<td>TG</td>
<td>13.05</td>
<td>13.46</td>
<td>13.40</td>
</tr>
<tr>
<td>UA</td>
<td>12.86</td>
<td>12.90</td>
<td>13.08</td>
</tr>
<tr>
<td>Average‡</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

† Numbers in bold represent cases where the corresponding policy performs as well as (or better than) the Optimal policy.
‡ Average is taken over the percentage of improvement across all cohorts under a policy.

1, 2, 3, 4, and POMDP policies, respectively, and (ii) reduces the medical expenditures per patient on average by 11.57%, 9.73%, 4.23%, 4.01%, and 1.93% compared to benchmark 1, 2, 3, 4, and POMDP policies, respectively.

Observation 5 (Intensity of Medications). During one year post-transplant, our optimal policy prescribes (i) high-dose tacrolimus 3.69, 2.73, and 0.48 fewer times per patient compared to benchmark 1, 2, and POMDP policies, respectively, and 0.18 and 0.34 more times per patient compared to benchmark 3 and 4 policies, respectively, (ii) medium-dose tacrolimus 0.63, 1.41, 1.48, and 0.52 more times per patient compared to benchmark 1, 2, 4, and POMDP policies, respectively, and 0.79 fewer times per patient compared to benchmark 3 policy, (iii) low-dose tacrolimus 2.09 and 0.32 more times per patient compared to benchmark 1 and 2 policies, respectively, and 0.39, 1.82, and 0.04 fewer times per patient compared to benchmark 3, 4, and POMDP policies, respectively, and (iv) insulin 2.09, 2.01, 2.12, 2.07, and 0.28 more times per patient compared to benchmark 1, 2, 3, 4, and POMDP policies, respectively.

Based on Observations 4 and 5 and the results provided in Tables 5 and 6, we shed light on the following implications for medical practitioners and policy makers:

(1) The improvements in QALE and cost made by our optimal policy is higher versus benchmark policies 1 and 2 than the other benchmark policies. The intensity of medications prescribed under these policies could be a contributing factor. For example, by following benchmark policies 1 and
Table 6 Comparison of medication policies (based on avg. number of medications prescribed under each policy)

<table>
<thead>
<tr>
<th>Cohort</th>
<th>Bench 1</th>
<th>Bench 2</th>
<th>Bench 3</th>
<th>Bench 4</th>
<th>POMDP</th>
<th>Optimal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age:L</td>
<td>3.91</td>
<td>3.33</td>
<td>3.97</td>
<td>7.05</td>
<td>2.96</td>
<td>1.99</td>
</tr>
<tr>
<td>Age:H</td>
<td>3.87</td>
<td>3.7</td>
<td>6.73</td>
<td>3.08</td>
<td>2.19</td>
<td>4.1</td>
</tr>
<tr>
<td>Gender:L</td>
<td>3.65</td>
<td>0.25</td>
<td>3.72</td>
<td>7.15</td>
<td>3.25</td>
<td>1.6</td>
</tr>
<tr>
<td>Gender:H</td>
<td>3.54</td>
<td>0</td>
<td>4.48</td>
<td>7.08</td>
<td>3.35</td>
<td>1.57</td>
</tr>
<tr>
<td>Race:L</td>
<td>4.17</td>
<td>0.54</td>
<td>3.85</td>
<td>6.64</td>
<td>3.1</td>
<td>2.26</td>
</tr>
<tr>
<td>Race:H</td>
<td>4.35</td>
<td>0.09</td>
<td>4.34</td>
<td>6.54</td>
<td>3.87</td>
<td>1.99</td>
</tr>
<tr>
<td>DiabHist:L</td>
<td>3.46</td>
<td>0.16</td>
<td>3.89</td>
<td>6.97</td>
<td>2.8</td>
<td>2.23</td>
</tr>
<tr>
<td>DiabHist:H</td>
<td>4.14</td>
<td>0.32</td>
<td>4.07</td>
<td>6.64</td>
<td>2.9</td>
<td>2.46</td>
</tr>
<tr>
<td>BMI</td>
<td>3.7</td>
<td>0</td>
<td>4.4</td>
<td>7.36</td>
<td>2.54</td>
<td>2.1</td>
</tr>
<tr>
<td>BP</td>
<td>3.94</td>
<td>0.2</td>
<td>4.11</td>
<td>6.79</td>
<td>2.96</td>
<td>2.25</td>
</tr>
<tr>
<td>Chol</td>
<td>5.5</td>
<td>0.12</td>
<td>3.92</td>
<td>7.41</td>
<td>3.09</td>
<td>1.5</td>
</tr>
<tr>
<td>HDL</td>
<td>3.55</td>
<td>0.63</td>
<td>4.3</td>
<td>7.42</td>
<td>2.9</td>
<td>1.68</td>
</tr>
<tr>
<td>BP</td>
<td>3.7</td>
<td>0</td>
<td>4.4</td>
<td>7.36</td>
<td>2.54</td>
<td>2.1</td>
</tr>
<tr>
<td>TG</td>
<td>3.62</td>
<td>0.77</td>
<td>3.96</td>
<td>6.83</td>
<td>3.39</td>
<td>1.78</td>
</tr>
<tr>
<td>UA</td>
<td>4.26</td>
<td>0</td>
<td>4.26</td>
<td>6.53</td>
<td>3.23</td>
<td>2.24</td>
</tr>
<tr>
<td>Avg. difference with Optimal</td>
<td>3.69</td>
<td>-0.63</td>
<td>-2.06</td>
<td>-2.09</td>
<td>2.73</td>
<td>-1.41</td>
</tr>
</tbody>
</table>

1, 2, 3, 4: number of times (on average) that high-/medium-/low-dose tacrolimus and insulin are prescribed during one year, respectively.

L: low level; H: high level

Average is taken over the differences with the Optimal policy over all cohorts.

2 in one year, a patient typically takes high-dose tacrolimus 3.69 and 2.73 more times compared to our optimal policy, while taking insulin 2.09 and 2.01 fewer times, respectively. As a result, the patient becomes more vulnerable against the diabetogenic effect of tacrolimus and NODAT complications. (2) Gains obtained by following our proposed policies compared to the current practice is not uniform across all cohorts of patients. In particular, from Table 5, we observe that for some cohorts of patients our approach yields the most improvement in QALE while incurring the least amount of medical expenditure. These cohorts include patients with (a) age < 50, (b) diabetes history, (c) normal or hypertensive blood pressure, (d) normal or high levels of cholesterol and triglyceride, and (e) normal or low HDL. (3) 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 incurring between $31 and $214 more medical costs (see Figure 4). This shows the importance of considering model misspecifications that are inevitable when data is used to estimate parameters: one should not rely on a single model to derive effective guidelines.

Finally, in Online Appendix E, we conduct sensitivity analyses on the estimated reward values (where both transplant and diabetes-related parameters are varied simultaneously), and find that the results discussed above are robust to the estimated values.

6. Conclusion

Immunosuppressive medications 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 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 (3) optimal policies are customized for different cohorts of patients.

Analyzing the APOMDP model, we first present some structural properties. These include piecewise-linearity and convexity of the value function, a theoretical link between a decision maker’s conservatism level and the intensity of prescribed medications, monotonicity of the optimal medication policy, and a feasible bound on the value function as an approximation. We then perform various numerical experiments using our clinical data set, and discuss their implications. For example, we observe that for some cohorts (e.g., non-White race, no diabetes history, and low cholesterol) a DM should be more concerned about organ rejection than the potential risk of NODAT. On the contrary, for other cohorts (e.g., male gender, diabetes history, and hypertension), the DM should be more concerned about the risk of NODAT than that of organ rejection.

We also compare our proposed optimal policies with four benchmark policies that represent the current medical practice (under different scenarios), and a POMDP policy that ignores the underlying model misspecifications. 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 4.58%, 3.46%, 0.83%, 0.79%, and 0.35% compared to benchmark 1, 2, 3, 4, and POMDP policies, respectively, (b) reduces the medical expenditures per patient 11.57%, 9.73%, 4.23%, 4.01%, and 1.93% compared to benchmark 1, 2, 3, 4, and POMDP policies, respectively, and (c) prescribes high-dose tacrolimus up to 3.69 fewer times per patient compared to other policies. The other important implications of the above-mentioned results for practitioners and policy makers are: (1) Cohorts of patients formed by age, diabetes history, blood pressure, cholesterol, HDL, and triglyceride will
benefit most from our methodology, because for such patients our approach yields the most improvement in QALE while incurring the least medical expenditure. (2) Practitioners or policy makers should not rely on a single model to derive 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 incurring between $31 and $214 more medical costs during the same period.

Our study has some limitations: (1) We consider 11 different risk factors each having two levels (i.e., low vs. high). This creates as many as $2^{11} = 2,048$ risk profiles for patients. However, we consider $2 \times 11 = 22$ cohorts of patients by changing one risk factor at a time. This allows us to focus on the effect of each individual risk factor separately. However, this disallows us to study the potential interactions between the risk factors. To perform such a study, we note that one needs to estimate transition and observation probabilities for each of the $2^{11}$ risk profiles, which, in turn, requires data of about 10,000 patients (i.e., more than half of all kidney transplantations in the U.S. in 2015 (UNOS 2016)). This is much larger than the number of patients seen at our partner hospital. Furthermore, one needs enough data to estimate the reward functions (e.g., QALE values) for all of these $2^{11}$ cohorts of patients. Nevertheless, as noted earlier, we believe that our approach of considering 22 cohorts of patients is strong enough to detect the impact of each risk factor on optimal prescription of medications. (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).

Future research can extend our work in two other directions. First, our approach can be applied to other solid organs (e.g., liver and pancreas) with the goal of creating a multi-organ data-driven decision-support system. Compared to kidney transplantation, where one can use dialysis when facing organ rejection, 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, and this, 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 further reduce expenditures while improving the care delivery process.
References


