http://pubsonline.informs.org/journal/msom

MANUFACTURING & SERVICE OPERATIONS MANAGEMENT
Articles in Advance, pp. 1–22
ISSN 1523-4614 (print), ISSN 1526-5498 (online)

Data-Driven Management of Post-transplant Medications: An Ambiguous Partially Observable Markov Decision Process Approach

Alireza Boloori,[a] Soroush Saghafian,[b] Harini A. Chakkera,[c] Curtiss B. Cook[c]

[a] Department of Industrial Engineering, Arizona State University, Tempe, Arizona 85281; [b] Harvard Kennedy School, Harvard University, Cambridge, Massachusetts 02138; [c] Divisions of Transplantation and Endocrinology, Mayo Clinic Hospital, Phoenix, Arizona 85054

Contact: aboloori@asu.edu (AB); soroush_saghafian@hks.harvard.edu, http://orcid.org/0000-0002-9781-6561 (SS); harini.chakkera@mayo.edu (HAC); cook.curtiss@mayo.edu (CBC)

Received: February 5, 2018
Revised: September 19, 2018; February 7, 2019
Accepted: March 17, 2019
Published Online in Articles in Advance: January 31, 2020

https://doi.org/10.1287/msom.2019.0797

Abstract. *Problem definition*: Organ-transplanted patients typically receive high amounts of immunosuppressive drugs (e.g., tacrolimus) as a mechanism to reduce their risk of organ rejection. However, because of the diabetogenic effect of these drugs, this practice exposes them to a greater risk of new-onset diabetes after transplantation (NODAT), and hence, becoming insulin dependent. We study and develop effective medication management strategies to address the common conundrum of balancing the risk of organ rejection versus that of NODAT. ***Academic/practical relevance*:** Our research contributes to the healthcare operations management literature by developing a robust stochastic decision-making framework that allows for incorporating (1) false-positive and false-negative errors of medical tests, (2) inevitable estimation errors when data sets are used, (3) variability among physician' attitudes toward ambiguous outcomes, and (4) dynamic and patient risk-profile-dependent progression of health conditions. ***Methodology*:** We apply an ambiguous partially observable Markov decision process (APOMDP) approach where dynamic optimization with respect to a "cloud" of possible models allows us to make decisions that are robust to potential misspecifications of risks. ***Results*:** We first provide various structural results that facilitate characterizing the optimal medication policies. Utilizing a clinical data set, we then compare the performance of the optimal medication policies obtained from our APOMDP model with the policies currently used in the medical practice. We observe that, in one year after transplant, our proposed policies can improve the life expectancy of each patient up to 4.58%, while reducing the medical expenditures up to 11.57%. ***Managerial implications*:** Balancing the risks of organ rejection and diabetes complications and considering factors such as physicians' attitudes toward ambiguous outcomes, partial observability of medical tests, and patient-specific risk factors are shown to result in more cost-effective strategies for management of post-transplant medications compared with the current medical practice. Finally, simultaneous management of medications can facilitate the care coordination process between transplantation/nephrology and endocrinology departments of a hospital that are typically in charge of administering such medications.

Funding: This work was partially supported by the National Science Foundation [Award CMMI-1562645].
Supplemental Material: The online appendix is available at https://doi.org/10.1287/msom.2019.0797.

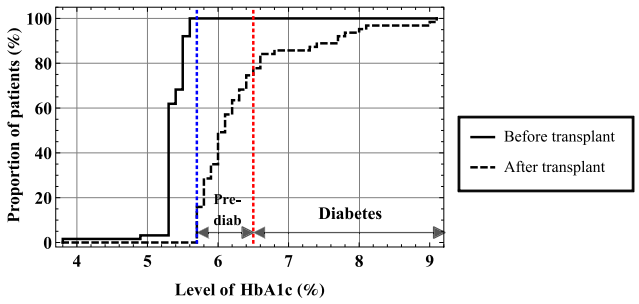
Keywords: ambiguous POMDP • cloud of models • conservatism level • kidney transplant • immunosuppressive drug • diabetes

1. Introduction

As reported by the United Network of Organ Sharing (2018), nearly 20,000 kidney transplantations were conducted in the United States in 2017 (140,992 cases since 2010). According to the Organ Procurement and Transplantation Network (2011), the average cumulative probability of 1-to-10-year organ rejection after kidney transplantation is estimated to be 6.35%–48.7%. To reduce the risk of organ rejection after transplant, physicians typically use an intensive amount of an immunosuppressive (also known as anti-rejection) drug (e.g., tacrolimus). However, because of the well-known *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 illustrate this point, we use a data set of 407 patients who had kidney transplant surgery at our partner hospital between 1999 and 2006. Based on this data set, Figure 1 depicts the empirical cumulative distribution functions (cdfs) of blood glucose levels

Figure 1. (Color online) Empirical Cdfs of Patients' HbA1c Levels in Our Data Set: An Illustration of the Diabetogenic Effect of Immunosuppressive Drugs



Note. The left (right) vertical dotted line shows the threshold for prediabetes (diabetes) as defined by American Diabetes Association (2012).

(measured by the hemoglobin A1c (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 operation (Bentley and Hanson 2011). To control the risk of elevated blood glucose levels, a patient may need diabetes medications (e.g., insulin). However, in the current practice, immunosuppressive drugs and diabetes medications are typically prescribed by different departments (transplantation/nephrology and endocrinology, respectively) of a hospital. This, in turn, results in a sequential management of these medications, which may reduce the efficacy of treatments. In addition, diabetes medications cannot be prescribed arbitrarily, because unnecessary use of such medications is harmful (Kromann et al. 1981). Therefore, the use of a diabetes medication should be coordinated with the intensity of the immunosuppressive drug used. Despite guidelines on how to manage these medications separately, there is currently no clear guideline on how to coordinate these regimens (i.e., how to simultaneously manage these medications). 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 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 the 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.

Ambiguity attitudes. Incomplete/imprecise information (which results in the foregoing estimation errors) typically makes physicians face ambiguity with respect to unknown consequences of treatment choices and their impact on a patient's health outcomes. Furthermore, physicians have a range of ambiguity attitudes in prescribing treatments: whereas some show high conservatism (high ambiguity aversion), others may exhibit low conservatism (low ambiguity aversion; see, e.g., Han et al. 2009, Arad and Gayer 2012, Berger et al. 2013).

Static and dynamic risk factors. Both static/time-invariant (e.g., race and gender) and dynamic/time-variant (e.g., blood pressure (BP) and body mass index (BMI)) 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 abovementioned 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 the ambiguous partially observable Markov decision process (APOMDP), an extension of the traditional POMDP approach recently proposed by Saghafian (2018). 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 toward model misspecifications (caused by physicians' ambiguity attitudes), and (4) several dynamic and/or static risk factors (age, gender, race, diabetes history, BMI, BP, total cholesterol (Chol), high-density lipoprotein (HDL) cholesterol, low-density lipoprotein (LDL) cholesterol, triglyceride (TG), and uric acid (UA)). This approach enables us to provide the first 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. 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 (QOL) and life expectancy). Using these estimations along with the APOMDP approach, we then generate risk-specific medication strategies for use in practice.

2. For non-white patients with age ≥ 50, no diabetes history, and low-risk levels of cholesterol, HDL, LDL, triglyceride, and uric acid, we find that, under the optimal medication policy, a more conservative physician prescribes more intensive regimens of immunosuppressive drugs as well as diabetes medications than a less conservative one. This implies that for patients with these risk factors, a more conservative physician should be more concerned about both risks of organ rejection and NODAT compared with a less conservative physician. However, this does not hold for male patients with age < 50, diabetes history, hypertension, and high-risk levels of cholesterol, HDL, and LDL.

3. Variations in physicians' attitudes toward ambiguity will not have a homogeneous impact on the intensity of drugs prescribed under the optimal policy. Thus, drug intensification (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.

4. Our study sheds light on the predictors of tacrolimus dose variability. Specifically, we find that risk factors such as age, gender, race, BMI, blood pressure, HDL, and LDL make patients more vulnerable to the risk of organ rejection. Furthermore, the diabetogenic effect of tacrolimus is more likely to influence male patients with age ≥ 50, diabetes history, hypertension, high cholesterol, and low HDL. This implies that when using high-dose tacrolimus, such patients become more dependent on diabetes medications than others.

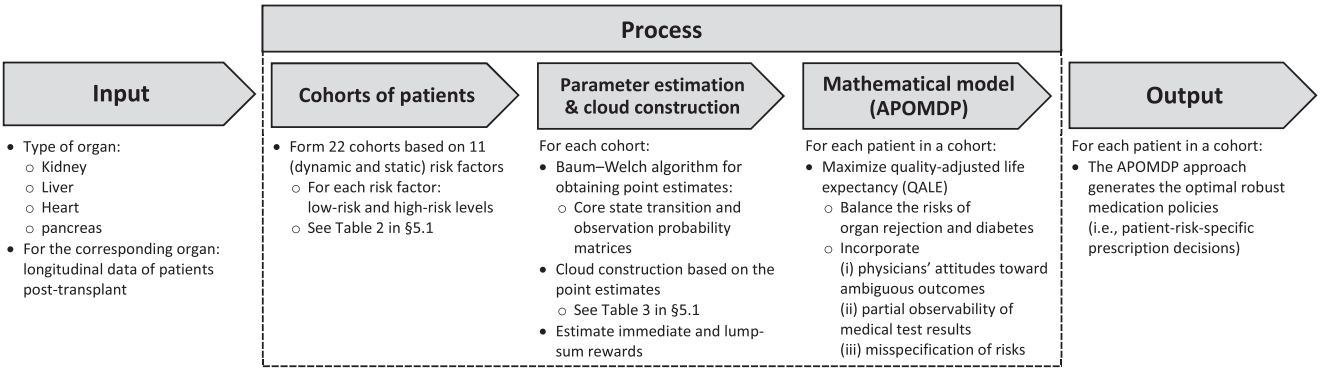
5. We compare the performance of the optimal medication policies that we obtain from the APOMDP approach with (a) benchmarks from the current medical practice and (b) medication policies that arise when one uses a traditional POMDP approach. We consider performance measures such as quality-adjusted life expectancy (QALE), medical expenditure, and the intensity of prescribed medications. Some of the main insights generated from our comparison are as follows:

- Compared with the current medical practice, and depending on different risk factors, our optimal medication policies can improve (per patient per year) the average (a) QALE up to 4.58% and (b) medical expenditures up to 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.
- 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, our contributions are twofold: (1) We demonstrate the use of the APOMDP approach to make *robust* dynamic decisions under both imperfect state information and model misspecifications. Because 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 (DM) 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 the piecewise-linearity and convexity property), which enables us to solve our APOMDP formulation optimally. We also establish (a) an analytical link between the 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.

In closing this section, we provide a road map for the implementation of our APOMDP approach in the management of post-transplant medications. Figure 2 shows a data-driven decision support system (DSS) that not only can assist physicians in their post-transplant medications management decisions, but can also influence medical guidelines. This DSS can achieve these goals by using our proposed approach to better balance risks of organ rejection and diabetes complications (based on each patient's characteristics), while incorporating physicians' attitudes toward ambiguous outcomes along with various other factors

Figure 2. Data-Driven DSS for Post-Transplant Medication Management: An Implementation Road Map



such as false-positive and false-negative error rates of medical tests and lack of data for valid estimation.

The rest of this paper is organized as follows. In Section 2, we provide a brief literature review. In Section 3, we present our APOMDP approach, and in Section 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 Section 5. Finally, we conclude the paper in Section 6 and discuss some avenues for future research.

2. Related Studies

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

2.1. 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 (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.

2.2. Operations Research/Management Science Studies on the Pretransplant Period

The majority of operations research (OR)/management science (MS) studies on transplantation focus on the *pretransplant* 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, 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.

2.3. Studies on POMDP Applications in Healthcare

In the medical decision-making field, POMDP models have been applied 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 static/dynamic risk factors. From the medical perspective, the latter is an improvement, because age and history of screening/treatment are the typical risk factors that have been considered thus far in the extant literature.

2.4. Studies on Robust Dynamic Decision Making

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, 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 (2018) proposes an APOMDP approach, where a controller makes decisions based on a weighted average of both the worst and the best possible outcomes. Moreover, unlike the abovementioned literature on robust MDPs, the APOMDP approach 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 are inevitable (e.g., because

of false positive/negative errors of medical tests). Considering the worst and the best possible outcomes (as opposed to all possible outcomes) is also important for partially observable systems, because it does not add much to the computational complexity.

Applications of robust dynamic decision making in medical problems have been centered around robust MDP formulations. Goh et al. (2018) develops a robust Markov chain framework for analyzing cost-effectiveness of colorectal cancer screening policies. Steimle et al. (2018) proposes a multimodel MDP for managing blood pressure and cholesterol, where model ambiguity is considered by averaging the performance of a given policy across different MDP models. Kaufman et al. (2011) and Zhang et al. (2017) model max-min MDPs for optimizing decisions on liver transplantation and glycemic control in diabetes management, respectively, where transition probabilities can vary within an uncertainty set. Compared to this stream, our work is the first study of a medical decision-making problem that considers both (1) model ambiguity and (2) behavioral attitudes of physicians toward ambiguity.

2.5. Studies on Measuring Ambiguity Attitudes

The ambiguity attitude of a decision maker can be characterized by either parametric or nonparametric methods. In the former, the ambiguity attitude is represented by utility-based models from the economics literature (see, e.g., Arad and Gayer 2012, Peysakhovich and Karmarkar 2015), whereas in the latter, it is measured by using behavioral scales based on sociodemographic characteristics of decision makers (see, e.g., Han et al. 2009). Our APOMDP framework is a parametric approach based on the so-called α-maxmin expected utility (α-MEU) preferences (Ghirardato et al. 2004), which measure a convex combination of the lowest (i.e., maxmin) and the highest (i.e., maxmax) possible outcomes based on the parameter $\alpha \in [0, 1]$. The parameter α captures a range of individuals' attitudes toward ambiguity, such that its high (low) values represent high (low) levels of ambiguity aversion (for empirical investigations of the α-MEU function, see Ahn et al. 2014 and the references therein).

The so-called range of ambiguity attitude has been estimated or set by the extant literature endogenously or exogenously. In the former, this parameter is inferred by conducting hypothesis testing with survey/questionnaire-based experiments (see, e.g., Chen et al. 2007). However, in the latter, this parameter is set without resorting to empirical experiments (see, e.g., Ahn et al. 2014). Compared with this stream, we can employ the DSS (shown in Figure 2) to implement our APOMDP approach and optimize decisions (about medication regimens) for any level of ambiguity attitude in [0, 1]. Therefore, our methodology can also be used to determine the best level of ambiguity attitude (i.e., the one that yields the highest QALE among all possible levels). Based on this premise, our findings in this paper are not predictive of physicians' behavior. Instead, they are prescriptive: they generate insights into what physicians should be targeting in their practice (both given their own level of ambiguity attitude and across all such possible levels).

2.6. Other 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, 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

We use a discrete-time, finite-horizon APOMDP approach (see Saghafian 2018) to determine optimal decisions that maximize QALE of a patient with respect to risks of organ rejection and NODAT complications. At each patient's visit, a decision maker—typically a physician—measures the patient's (1) lowest concentration of tacrolimus (in the body) known as the *trough level*, or C_0, and (2) blood glucose level. Then, after evaluating whether the patient has a low, medium, or high C_0 and whether he or she 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 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 or 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.

In addition to identifying optimal decisions and investigating their cost-effectiveness, we use this setting to study unnecessary intensification of prescribed medications. We do so by comparing the effect of using (a) lower dosages of tacrolimus and (b) using insulin versus not using it. Furthermore, our notion of simultaneous prescriptions facilitates the care coordination between the transplantation/nephrology and

endocrinology departments of a hospital, which are typically in charge of administering tacrolimus and insulin, respectively.

3.1. The Elements of the APOMDP Approach

The elements of our APOMDP approach are as follows: decision epochs, core state space, observation state space, action space, ambiguity set, core state transition probability, observation probability, information space, belief space, immediate reward, lump-sum reward, ambiguity attitude set, and discount factor. 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 $n = 1, 2, \ldots, N$, where n represents the number of months since transplant. We consider one year post-transplant as our planning horizon ($N = 12$), because it represents the time period during which medication management strategies are (a) most important and (b) most variable among physicians particularly for tacrolimus regimens (see, e.g., Staatz and Tett 2004, Schiff et al. 2007).

Core state space. $S = \{\Delta, \nabla\} \cup \mathcal{S}$, where $\mathcal{S} = \{s_i, i = 1, 2, \ldots, 9\}$ and s_i's are as described in Table 1. In addition, Δ and ∇ represent "death" and "organ rejection," respectively. We note that ∇ and Δ 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. $O = \{\Delta, \nabla\} \cup \mathbb{O}$, where $\mathbb{O} = \{o_i, i = 1, 2, \ldots, 9\}$, 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 while having organ survival and diabetic conditions.

Table 1. Description of Parts of Core Health States and Actions

State	Transplant condition[a] (tacrolimus C_0)	Diabetes condition
s_1	Low	Diabetes (type II)
s_2	Medium	
s_3	High	
s_4	Low	Prediabetes
s_5	Medium	
s_6	High	
s_7	Low	Healthy
s_8	Medium	
s_9	High	
Action	Prescription	
	Tacrolimus dose	Insulin use
a_1	High	Yes
a_2	Medium	
a_3	Low	
a_4	High	No
a_5	Medium	
a_6	Low	

[a]With the patient experiencing an organ survival.

Action space. $A = \{a_i, i = 1, 2, \ldots, 6\}$, where a_i's are described in Table 1. Letting $a \preceq \widehat{a}$ represent the fact that a is more intensive than $\widehat{a}$ (or $\widehat{a}$ is less intensive than a), it can be seen from Table 1 that $a_1 \preceq a_2 \preceq a_3$, $a_4 \preceq a_5 \preceq a_6$, $a_1 \preceq a_4$, $a_2 \preceq a_5$, $a_3 \preceq a_6$, and $a_1 \preceq a_6$. Thus, a_1 (a_6) corresponds to administrating the most (least) intensive medication regimen. Similarly, we use the notation $a \npreceq \widehat{a}$ to represent situations where $a \preceq \widehat{a}$ does not hold (i.e., when either $\widehat{a} \preceq a$ or when there is no ordering between the two).

Ambiguity set (cloud of models). $M = \{m_1, m_2, \ldots, m_K\}$, where K is the number of models in the cloud. As mentioned in Section 1, estimating transition and observation probability matrices from a data set is subject to errors. This, in turn, results in model misspecifications which warrants the cloud of models (as opposed to a single model). Each model in M represents a different estimation for the core state transition and observation probability matrices (defined below). In Section 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. $\mathbf{P}_m = \{\mathbf{P}_m^a : a \in A\}$, where for each $a \in A$, $\mathbf{P}_m^a = [p_m^a(j|i)]_{i,j \in S}$, 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 $m \in M$.

Observation probability. $\mathbf{Q}_m = \{\mathbf{Q}_m^a : a \in A\}$, where for each $a \in A$, $\mathbf{Q}_m^a = [q_m^a(o|j)]_{j \in S, o \in O}$, 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. $\Pi = \{\boldsymbol{\pi} = [\pi_i]_{i \in S} \in \mathbb{R}^{|S|} : \sum_{i=1}^{|S|} \pi_i = 1, \pi_1, \pi_2 \in \{0, 1\}, \pi_3, \ldots, \pi_{11} \in [0, 1]\}$, where $\boldsymbol{\pi}$ is an information vector over the state space S. Because 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 $\boldsymbol{\pi} \neq [1, 0, \ldots, 0]'$ or $\boldsymbol{\pi} \neq [0, 1, \ldots, 0]'$, $\mathbf{b} = [0, 0, b_3, \ldots, b_{11}] = \boldsymbol{\pi}$ (i.e., DM's belief 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 stands for partially observable.)

We use 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 $B(\mathbf{b}, a, o, m) : \Pi_{PO} \times A \times O \times M \to \Pi_{PO}$ 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

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

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

Immediate reward. $\mathbf{r}_n(a) = [r_n(s,a) \geq 0]_{s\in S}$ for $a \in A$, where $r_n(s,a)$ is the quality of life 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 $s \in \mathcal{S}$.

Lump-sum reward. $\mathbf{R}_n = [R_n(s) \geq 0]_{s\in S}$, where $R_n(s)$ is a lump-sum reward (in QALE) gained by a patient whenever she or 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 or her remaining lifetime, or (2) during the planning horizon ($n < N$), if she or 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 or her range of attitude toward ambiguity. We note that this is the same as parameter α in the α-MEU function described in Section 2.

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

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

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

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

where

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

The first term on the right-hand side (RHS) of (5) represents the expected current reward (in QALE) in period n when the belief vector is $\mathbf{b}$ and the action is a. The other terms on 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 ambiguity attitude of a more (less) conservative DM (see, e.g., Chen et al. 2007, Ahn et al. 2014). By varying λ, our framework allows us to capture the behavioral attitudes of physicians and evaluate their effects on the intensity of medications administered. We note that $\lambda = 1$ 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.

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

$$\underline{m}_n(\mathbf{b}, a, \lambda) = \arg\min_{m\in M}\{H_n(\mathbf{b}, a, m, \lambda)\} \quad \text{and}$$
$$\overline{m}_n(\mathbf{b}, a, \lambda) = \arg\max_{m\in M}\{H_n(\mathbf{b}, a, m, \lambda)\}. \tag{7}$$

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

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. Compared to the earlier work of Saghafian (2018) that establishes structural results for general APOMDPs, we make use of the specific properties of the medical problem under consideration and provide (1) a closed-form expression for the piecewise-linear and convex (PLC) value function, (2) an analytical link between the DM's conservatism level and his actions (i.e., the intensity of prescribed medications), (3) a lower bound for the optimal value function, and (4) specific monotonicity results for the optimal policy.

4.1. 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 in the belief vector (Saghafian 2018). This may prevent us from using solution algorithms (similar to those used for POMDPs), because many of them rely on the piecewise-linearity and convexity property of the value function. Thus, to guarantee the piecewise-linearity and convexity 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 M.

Definition 1 (Saghafian 2018). Model $\underline{m}_n(\mathbf{b}, a, \lambda) \in M$ defined in (7) is said to be a BIWC member of M if it is constant in the belief vector $\mathbf{b}$.

This implies that, irrespective of the DM's belief about a patient's health state, there exists a set of transition and observation matrices (given the action and conservatism level) that yields the least total reward (in QALE). If such a model exists in M, then the optimal value function is PLC in the belief vector $\mathbf{b}$ (see Saghafian 2018, proposition 2), and hence can be written as

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

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

Although the existence of a BIWC member in the cloud of models 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 $a\in A$, $\mathbf{P}^a_m\mathbf{Q}^a_m=\mathbf{P}^a_{\hat{m}}\mathbf{Q}^a_{\hat{m}}\mathbf{W}$ for some stochastic matrix $\mathbf{W}$, then model m is said to be *less informative* than model $\hat{m}$. (For notational simplicity, we suppress the dependency on a.) It follows that if one model is less informative than the others, then it is a BIWC member in M (see Saghafian 2018, proposition 3). Utilizing our clinical data set in Online Appendix B.3, 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.

Assuming that 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.

Proposition 1 (Representation of $\boldsymbol{\psi}$-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 (7). Then, the set of $\boldsymbol{\psi}$-vectors ($\Psi_{n,\lambda}$) in (8) can recursively be obtained as*

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

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

where

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

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 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:

$$\widetilde{\Psi}_{n,\lambda}=\Bigg\{\widetilde{\boldsymbol{\psi}}\in\mathbb{R}^{|S|}:\widetilde{\boldsymbol{\psi}}=\mathbf{r}_n(a)+\lambda\Big(\beta\sum_{o\in O}\mathbf{P}^a_{\underline{m}}\mathbf{Q}^{a,o}_{\underline{m}}\,\widetilde{\boldsymbol{\psi}}^{(\mathbf{b},a,o)}_{\underline{m}}\Big)+(1-\lambda)\Big(\beta\sum_{o\in O}\mathbf{P}^a_m\mathbf{Q}^{a,o}_m\,\widetilde{\boldsymbol{\psi}}^{(\mathbf{b},a,o)}_m\Big),\ a\in A,m\in M\setminus\{\underline{m}\},\widetilde{\boldsymbol{\psi}}^{(\mathbf{b},a,o)}_{\underline{m}},\widetilde{\boldsymbol{\psi}}^{(\mathbf{b},a,o)}_m\in\widetilde{\Psi}_{n+1,\lambda}\Bigg\} \quad \forall\ \lambda\in\Lambda,\forall\ n<N. \tag{12}$$

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

4.2. 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 medication regimens. To study this phenomenon, we start by defining the following conditions. In Online Appendix B.5, we also numerically test the validity of conditions in this section using our clinical data set, and discuss whether and when such conditions hold.

Condition 1 (Monotonicity of Reward). (i) *Under any action $a\in A$, the immediate reward vector $\mathbf{r}_n(a)$ is nondecreasing in state $s\in S$, and* (ii) *the lump-sum reward vector $\mathbf{R}_n$ is nondecreasing in state $s\in S$.*

Condition 1 implies that better health states have higher immediate and lump-sum rewards (in QALE). For example, compared to a patient with an organ rejection, a patient with an organ survival is expected to have a higher quality of life (all else equal).

Condition 2 (TP_2 Transitions). *For all $m\in M$ and $a\in A$, the kernels $\mathbf{P}^a_m$ and $\mathbf{Q}^a_m$ are TP_2 (i.e., all their second-order minors are nonnegative).*

Condition 2 imposes a specific ordering between each two consecutive rows of $\mathbf{P}$ and $\mathbf{Q}$ matrices. For example,

this condition implies that, upon taking the same medication regimen, a patient with a better health state is more likely to move to a more favorable state than another patient who is in a worse health state (all else equal).

For later use, here we also define the well-known TP_2 stochastic ordering between two belief vectors. Because each belief vector $\mathbf{b}$ yields 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 $\mathbf{b}$ is said to be dominated by another belief vector $\widehat{\mathbf{b}}$ in the MLR-ordering sense (shown as $\mathbf{b} \preceq_r \widehat{\mathbf{b}}$) if the ratio $\widehat{\mathbf{b}}/\mathbf{b}$ is nondecreasing in its elements.

From the medical standpoint, Definition 2 implies that a patient with associated belief vector $\widehat{b}$ is more likely to be in a better health state than another patient with associated belief vector b. We also need to define the following condition, where, for notational simplicity, we let $\underline{m}(a,\lambda) = \underline{m}_n(\mathbf{b},a,\lambda)$ and $\overline{m}(a,\lambda) = \overline{m}_n(\mathbf{b},a,\lambda)$ for any action a and conservatism level λ. In addition, $B(\mathbf{b},a,o,m)$ is the belief-updating operator defined in Equation (1). Similarly, we denote by $[\Pr\{o|\mathbf{b},a,m\}]_{o\in O}$ the vector of observation probabilities, where $\Pr\{o|\mathbf{b},a,m\}$ is the conditional probability that a DM will make observation o given the belief vector $\mathbf{b}$, action a, and model m (see Equation (2) in Section 3.1).

Condition 3. *Fix belief vector $\mathbf{b} \in \Pi_{PO}$ and time period $n<N$. Then, for all $a,\widehat{a} \in A$ with $a \not\preceq \widehat{a}$, there exists a conservatism level $\lambda^* \in \Lambda$ such that, for all $\lambda \geq \lambda^*$, we have*

(i) $\left[\Pr\{o|\mathbf{b},\widehat{a},\underline{m}(\widehat{a},\lambda^*)\}\right]_{o\in O} \preceq_{TP_2} \left[\Pr\{o|\mathbf{b},a,\underline{m}(a,\lambda^*)\}\right]_{o\in O}$, $[\Pr\{o|\mathbf{b},a,\underline{m}(a,\lambda)\}]_{o\in O} \preceq_{TP_2} [\Pr\{o|\mathbf{b},\widehat{a},\underline{m}(\widehat{a},\lambda)\}]_{o\in O}$;

(ii) $B\big(\mathbf{b},\widehat{a},o,\underline{m}(\widehat{a},\lambda^*)\big) \preceq_{TP_2} B\big(\mathbf{b},a,o,\underline{m}(a,\lambda^*)\big)$, $B(\mathbf{b},a,o,\underline{m}(a,\lambda)) \preceq_{TP_2} B(\mathbf{b},\widehat{a},o,\underline{m}(\widehat{a},\lambda))$; *and*

(iii) *parts* (i) *and* (ii) *also hold for the best model* $\overline{m}$.

To better understand Condition 3, let DM_{base} represent a *baseline* DM with the conservatism level λ^* introduced in Condition 3. Also, we denote by DM_{gen} a general DM with a conservatism level λ such that $\lambda \geq \lambda^*$ (i.e., DM_{gen} is more conservative than DM_{base}). Then, part (i) of Condition 3 has the following implication for the medical practice: DM_{base} (DM_{gen}) is more (less) likely to have a better medical observation if prescribing a less intensive medication regimen (compared to a more intensive one). Furthermore, part (ii) of Condition 3 implies that DM_{base} (DM_{gen}) has a better (worse) updated belief about a patient's health state (in the TP_2 sense) when taking less intensive (than more intensive) medications. Parts (i) and (ii) of Condition 3 also require different utilizations of models (from the ambiguity set) under different conservatism levels: for any $a \in A$ and any $\lambda,\widehat{\lambda} \in \Lambda$ such that $\lambda \neq \widehat{\lambda}$, $\underline{m}(a,\lambda) \neq \underline{m}(a,\widehat{\lambda})$ and $\overline{m}(a,\lambda) \neq \overline{m}(a,\widehat{\lambda})$. Otherwise, unlike our results in Theorem 1 or Corollary 1 (discussed below), the level of conservatism would have no impact on the intensity of medication regimens.

Theorem 1 (Effect of λ on Drug Intensification). *Let $a_n^*(\mathbf{b},\lambda)$ be the optimal medication action for any belief vector $\mathbf{b} \in \Pi_{PO}$, conservatism level $\lambda \in \Lambda$, and time period $n<N$. Also, let λ^* represent the baseline conservatism level introduced in Condition* 3. *Then, under Conditions* 1–3, *for any $\lambda \geq \lambda^*$, we have $a_n^*(\mathbf{b},\lambda) \preceq a_n^*(\mathbf{b},\lambda^*)$.*

Theorem 1 provides insights into conditions under which the optimal medication regimen becomes more intensive as the DM's conservatism level increases compared to a baseline level. This result, however, may not hold for all patients, because the sufficient conditions in Theorem 1 may not hold for them. In particular, in Corollary 1, we show that if for some patients Condition 3 is reserved (i.e., the reverse of orderings and inequalities in parts (i) and (ii) of Condition 4 hold), then the optimal medication regimen for them becomes less intensive as the DM's conservatism level increases. Thus, while under the optimal policy for some patients a more conservative physician prescribes more intensive medications than a less conservative one, for some patients this result might be reversed. In Section 5.2.1, we make use of our clinical data set and shed more light on patient characteristics for which either of these two cases holds.

Corollary 1. *Under Conditions* 1 *and* 2 *and the reverse of* 3, *for any $\lambda \geq \lambda^*$, we have $a_n^*(\mathbf{b},\lambda^*) \preceq a_n^*(\mathbf{b},\lambda)$.*

4.3. 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 $\phi_m^{(\mathbf{b},a)} = \sum_{o\in O} \mathbf{P}_m^a \mathbf{Q}_m^{a,o} \boldsymbol{\psi}_m^{(\mathbf{b},a,o)}$ (for all $\mathbf{b} \in \Pi_{PO}$, $a \in A$, and $m \in M$), where $\boldsymbol{\psi}_m^{(\mathbf{b},a,o)}$ is defined as in* (11). *Then, for any $a,\widehat{a} \in A$ such that $a \preceq \widehat{a}$ and $\lambda \in \Lambda$, vectors $\phi_{\underline{m}(\mathbf{b},\widehat{a},\lambda)}^{(\mathbf{b},\widehat{a})} - \phi_{\underline{m}(\mathbf{b},a,\lambda)}^{(\mathbf{b},a)}$ and $\phi_{\overline{m}(\mathbf{b},\widehat{a},\lambda)}^{(\mathbf{b},\widehat{a})} - \phi_{\overline{m}(\mathbf{b},a,\lambda)}^{(\mathbf{b},a)}$ are nondecreasing in their elements.*

Conditions 4 implies that, when taking a less intensive medication regimen compared with 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 $a_n^*(\mathbf{b},\lambda)$ be the optimal medication action for period n. Then, under Condition* 4, *$\mathbf{b} \preceq_{TP_2} \widehat{\mathbf{b}}$ yields $a_n^*(\mathbf{b},\lambda) \preceq a_n^*(\widehat{\mathbf{b}},\lambda)$.*

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.

4.4. 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 λ) is $O(|M||A||S||\Psi_{n+1,\lambda}|^{|O|})$ (for discussions about the time and space complexities of [PO]MDPs, see Papadimitriou and Tsitsiklis 1987, Hauskrecht 2000). Although we alleviate this effect by implementing the Monahan's (1982) algorithm to eliminate dominated ψ-vectors, to further streamline computational burdens, we now develop a bound for the value function in (4). We let $J_n(\mathbf{b},\lambda)$ be the approximate value function, and let $a^{*J}(\mathbf{b},\lambda)$ 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 computes the expected future reward based on his or her updated belief about the 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 or her expected belief (over all updated belief vectors), and then the reward based on the expected belief:

$$\begin{aligned} J_n(\mathbf{b},\lambda) = \mathbf{b}'\mathbf{r}_n(a^J) &+ \lambda \min_{m\in M}\left\{\beta J_{n+1}\left(\mathbf{b}'\mathbf{P}_m^{a^J},\lambda\right)\right\} \\ &+ (1-\lambda)\max_{m\in M}\left\{\beta J_{n+1}\left(\mathbf{b}'\mathbf{P}_m^{a^J},\lambda\right)\right\}, \end{aligned} \tag{13}$$

where we obtain $\mathbf{b}'\mathbf{P}_m^{a^J}$ from $\sum_{o\in O} Pr\{o|\mathbf{b},a^J,m\}\, B(\mathbf{b},a^J,o,m)$ by following the Bayesian update in Equation (1) and the fact that $\sum_{o\in O}\mathbf{Q}_m^{a,o} = \mathbb{I}$, where $\mathbb{I}$ is an identity matrix. Proposition 2 shows that the optimal value function $V_n(\mathbf{b},\lambda)$ is tightly bounded from below by the approximate value function $J_n(\mathbf{b},\lambda)$.

Proposition 2 (Performance Bound). *Suppose* (i) *the ambiguity set M has a BIWC member,* (ii) $|p_m^a(j|i) - p_{\widehat{m}}^a(j|i)| \le \eta$ *for some* $\eta \ge 0$ $(\forall\, a\in A, \forall\, m,\widehat{m}\in M, \forall\, i,j\in S)$, *and* (iii) $\bar{r}$ *is the maximum possible reward in each period. Also, let* $\epsilon_{n+1} = \epsilon_q \sum_{l=0}^{N-n-1}\beta^l + \epsilon_r\beta^{N-n}$, *where* ϵ_q *and* ϵ_r *are upper bounds for the quality of life and lump-sum reward, respectively. Then, we have*

$$V_n(\mathbf{b},\lambda) - J_n(\mathbf{b},\lambda) \le \min\left\{\frac{\beta\,\eta\,\epsilon_{n+1}\,|S|}{1-\beta}, \frac{\bar{r}(1-\beta^N)}{1-\beta}\right\} \quad \forall\, \mathbf{b}\in\Pi_{PO}, \forall\, \lambda\in\Lambda, \forall\, n<N. \tag{14}$$

In Proposition 2, ϵ_q is a bound for the QOL score, which is a score between 0 and 1. Similarly, ϵ_r is a bound on the lump-sum reward, which is a function of residual life expectancy (RLE) and a discount rate, such that as the discount rate approaches 1, the lump-sum reward approaches QOL (see Section 5.1 for more details regarding these reward parameters). We note that the bound provided by ϵ_{n+1} is relatively tight. For example, it goes to 0 as $\beta\to 0$, and to $(N-n-1)\epsilon_q+\epsilon_r$ as $\beta\to 1$. Also, for $\beta\in[0,1)$, this bound asymptotically approaches $\frac{\epsilon_q}{1-\beta}$ as $N\to\infty$. Furthermore, 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, $J_n(\mathbf{b},\lambda)$ converges to $V_n(\mathbf{b},\lambda)$, 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 M become more similar), η approaches 0; (2) when $\beta\in[0,1)$ and the time horizon increases, ϵ_{n+1} asymptotically approaches $\frac{\epsilon_q}{1-\beta}$, which, in turn, approaches 0 as a patient's health status gets aggravated; and (3) when β approaches 0 (i.e., the DM decides on medications regimens in a myopic approach). Furthermore, when β approaches 1, the performance bound in (14) approaches $N\bar{r}$, which is small when N or $\bar{r}$ is small. In general, the bound in (14) is advantageous for the DM, because it enables him or 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 those influencing medical guidelines.

5.1. Data and Parameter Estimation

5.1.1. 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 to each patient's visit at months 1, 4, and 12 after transplant and includes the following attributes: (1) demographic (e.g., age, race, gender, etc.), (2) clinical (e.g., blood pressure, 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).

5.1.2. Interpolation and Imputation. Because our data set includes only information at months 1, 4, and 12 after 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 after transplant. Prior to that, to replace missing values in the data entries, we employ *multiple imputations by chained equations* by the R computing package (for more details, see, e.g., Buuren and Groothuis-Oudshoorn 2011).

5.1.3. 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 (the median age of patients in our data set is 53 years, and 40% of patients are below 50); (2) non-white race includes Hispanic, black, and Native American; (3) diabetes history refers to the existence of diabetes prior to the time of transplant (among 407 patients, there were 115 patients (28%) with a history of diabetes before or at the time of transplant); (4) the thresholds for classifying risk factors (except for age, gender, race, and blood pressure) as low/high are based on MedPlus (2018); and (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 (American Heart Association 2018).

5.1.4. Choice of Medication Regimens and Health States. 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 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 (for a similar assumption, see also Denton et al. 2009, Mason et al. 2014).

Unlike insulin, which is prescribed in a fixed dosage, physicians prescribe tacrolimus based on C_0 (trough level). 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 after transplant), 8–10 mg/dL (month 4 after transplant), and 6–8 mg/dL (month 12 after transplant). Thus, we label any $C_0 \in [4,8)$, $[8,10)$, and $[10,14]$ 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 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

Table 2. Description of Main Risk Factors and Their Levels (See Also Boloori et al. 2015)

Risk factor	Unit	Low level	High level	Static (S)/dynamic (D)
Age	Years	<50	≥50	S
Gender		Female	Male	S
Race		White	nonwhite	S
Diabetes history		No	Yes	S
Body mass index	kg/m^2	<30 (nonobese)	≥30 (obese)	D
Blood pressure		Normal	Hypertension	D
Total cholesterol	mg/dL	<200	≥200	D
High-density lipoprotein	mg/dL	≥40	<40	D
Low-density lipoprotein	mg/dL	<130	≥130	D
Triglyceride	mg/dL	<150	≥150	D
Uric acid	mg/dL	<7.3	≥7.3	D

or HbA1c ≥ 6.5% (5.7 ≤ HbA1c < 6.5%) is labeled diabetic (pre-diabetic), whereas one with FPG < 100 mg/dL or HbA1c < 5.7% is labeled healthy (American Diabetes Association 2012).

5.1.5. Estimation of Probability Matrices and Cloud Construction. For each cohort of patients in Table 2, we construct a cloud of probabilistic models in two phases.

Phase 1: Point Estimates. We employ the Baum–Welch (BW) algorithm (Welch 2003) to obtain point estimations for core state transition and observation probability matrices (lines 6–9 in Table 3). As inputs to this algorithm, we use (1) the sequence of medical observations (tacrolimus C_0 and blood glucose levels) and actions (prescribed medications) from our clinical data set and (2) initial transition and observation probability matrices. We note that the BW algorithm is iterated 1,000 times to account for the inevitable variability caused by considering random initial probability matrices. Thus, we treat the average outputs of the BW algorithm over all iterations as our point estimates. Despite 1,000 iterations, the resulted point estimates may not be reliable. Thus, we address this issue by constructing the cloud of models.

Phase 2: Cloud Construction. We construct an ambiguity set as a cloud of probabilistic models surrounding the point estimates resulted from Phase 1. We first identify the set of all probability matrices that are within an ϵ-distance from the points estimates. To this end, we characterize the distance by the Kullback–Leibler (KL) divergence criterion (also known as *relative entropy*), which is applied on each row of probability matrices (see Table 3 for the notation used):

$$d_{KL}\left(\mathbf{v}, \mathbf{P}^a_{BW}(i)\right) = \sum_{j\in S} \mathbf{v}(j) \log_2\left(\frac{\mathbf{v}(j)}{p^a_{BW}(j|i)}\right) \quad \forall\, \mathbf{v} \in \mathbb{V}, \forall\, a \in A, \forall\, i \in S \setminus \{\Delta, \nabla\}, \tag{15}$$

where $\mathbf{P}^a_{BW} = \left[p^a_{BW}(j|i)\right]_{i,j\in S}$ is the point estimate returned by the BW algorithm, and $\mathbf{P}^a_{BW}(i)$ is the ith row in matrix $\mathbf{P}^a_{BW}$ (the same procedure is used for matrix $\mathbf{Q}^a_{BW} = \left[q^a_{BW}(o|j)\right]_{j\in S, o\in O}$). We note that we do not apply the KL distance in (15) for the absorbing states (i.e., death Δ and organ rejection ∇) in probability matrices. Instead, we simply consider a unit row vector for the first two rows in these matrices.

Because of the KL divergence in (15), the cloud of models is an infinite set (line 12 in Table 3). However, because we require the existence of a BIWC member in the cloud (see Section 4), we randomly select a finite number (i.e., $|M|$) of samples from this set, such that the BIWC member condition is satisfied (lines 13–15 in Table 3). This, in turn, makes the cloud of models a finite set. In Online Appendix B.3, we provide further details on the existence of a BIWC member in our clinical data set, and in Online Appendix B.4, we validate our estimations of the set of transition and observation probability matrices.

5.1.6. Estimation of the Initial Observation Probability Matrix. Our partner hospital conducts two tests to measure blood glucose levels: if HbA1c is ≥6.5% (5.7 ≤ HbA1c < 6.5%) or FPG is ≥126 mg/dL (100 mg/dL ≤ FPG < 126 mg/dL), then the patient is said to have diabetes type II (pre-diabetes). Each of these tests have

Table 3. A Pseudocode for Constructing the Cloud of Models (Transition and Observation Probability Matrices)

Inputs	1	Initial transition (randomly generated) and observation (see below) probability matrices
	2	Sequence of medical observations and actions (for each cohort) from our data set
	3	Kullback-Leibler (KL) distance bound $= \epsilon \geq 0$ (e.g., $\epsilon = 0.05$)
	4	$\mathbb{V} = \left\{\mathbf{v} = [v_i]_{1\leq i\leq \lvert S\rvert} \in \mathbb{R}_+^{\lvert S\rvert} : \sum_{i=1}^{\lvert S\rvert} v_i = 1\right\}$
	5	Number of distinct models in the cloud (the ambiguity set) $= \lvert M\rvert$
Phase 1	6	**for** i = 1 to 1,000 // number of iterations
	7	**do** *Baum-Welch* algorithm // using inputs 1–2
	8	**return** core state transition and observation probability matrices
	9	**return** point estimates $\mathbf{P}^a_{BW}$ and $\mathbf{Q}^a_{BW}$ for each action $a \in A$ // average of outputs over 1,000 iterations
Phase 2	10	**while** the model informativeness condition is not met for $\mathbf{P}_m$ and $\mathbf{Q}_m$ (for all $m \in M$)
	11	**for each** $a \in A$ and $i = 3$ to $\lvert S\rvert$ // i: any core health state except death and organ rejection
	12	$\mathbb{V}_P(i) = \{\mathbf{v} : \mathbf{v} \in \mathbb{V}, d_{KL}\left(\mathbf{v}, \mathbf{P}^a_{BW}(i)\right) \leq \epsilon\}$, $\mathbb{V}_Q(i) = \{\mathbf{v} : \mathbf{v} \in \mathbb{V}, d_{KL}\left(\mathbf{v}, \mathbf{Q}^a_{BW}(i)\right) \leq \epsilon\}$ // using inputs 3–4
	13	**for** $m = 1$ to $\lvert M\rvert$ // using input 5
	14	**do** randomly select vectors $\mathbf{p} \in \mathbb{V}_P(i)$ and $\mathbf{q} \in \mathbb{V}_Q(i)$
	15	$\mathbf{P}^a_m(i) = \mathbf{p}$ and $\mathbf{Q}^a_m(i) = \mathbf{q}$
	16	**return** probability sets $\mathbf{P}_m$, $\mathbf{Q}_m$ (for all $m \in M$)

their own specificity and sensitivity values (see, e.g., Bennett et al. 2007). Using the notations in Table 4, we then have

$$sp^H = 1 - \left(sp^{A1C}_{5.7}(1 - sp^{FPG}_{100}) + sp^{FPG}_{100}(1 - sp^{A1C}_{5.7}) + (1 - \mathrm{sp}^{\mathrm{FPG}}_{100})(1 - \mathrm{sp}^{\mathrm{A1C}}_{5.7})\right), \tag{16a}$$

$$sn^{PD} = sn^{FPG}_{100}(1 - sn^{A1C}_{5.7}) + sn^{A1C}_{5.7}(1 - sn^{FPG}_{100}) + sn^{FPG}_{100} sn^{A1C}_{5.7}. \tag{16b}$$

Note that sp^{PD} is obtained by (16a), and sn^D is obtained by (16b), where, the cutoff values of 5.7 and 100 are replaced by 6.5 and 126, respectively. Letting $\mathbf{Q}^D = [q^D_{ij}]_{i,j\in\{1,2,3\}}$ and $\mathbf{Q}^T = [q^T_{ij}]_{i,j\in\{1,2,3\}}$ be the diabetes, transplant, and overall initial observation probability matrices, respectively, we have

$$\mathbf{Q}^D = \begin{bmatrix} sn^D & sn^{PD}(1-sn^D) & 1 + sn^{PD}sn^D - (sn^{PD} + sn^D) \\ sp^H(1-sp^{PD}) & sp^{PD}sn^{PD} & sp^{PD}(sp^H - sn^{PD}) - sp^H + 1 \\ (1-sp^H)(1-sp^{PD}) & (1-sp^H)sp^{PD} & sp^H \end{bmatrix},$$

$$\mathbf{Q}^T = \begin{bmatrix} sp^T_8 & (1-sp^T_8)sp^T_{10} & (1-sp^T_8)(1-sp^T_{10}) \\ sp^T_{10}(sp^T_8 - sn^T_8) - sp^T_8 + 1 & sp^T_{10}sn^T_8 & sp^T_8(1-sp^T_{10}) \\ 1 + sn^T_8 sn^T_{10} - (sn^T_8 + sn^T_{10}) & sn^T_8(1-sn^T_{10}) & sn^T_{10} \end{bmatrix},$$

$$\mathbf{Q} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & q^T_{11}q^D_{11} & q^T_{11}q^D_{12} & q^T_{11}q^D_{13} & \cdots & q^T_{13}q^D_{11} & q^T_{13}q^D_{12} & q^T_{13}q^D_{13} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & q^T_{31}q^D_{31} & q^T_{31}q^D_{32} & q^T_{31}q^D_{33} & \cdots & q^T_{33}q^D_{31} & q^T_{33}q^D_{32} & q^T_{33}q^D_{33} \end{bmatrix}.$$

There is no consensus in the medical literature about the specificity/sensitivity of the foregoing medical tests. However, the specificity (sensitivity) usually increases (decreases) with increasing cutoff points. This is reflected in the values in Table 4. We note that, using our APOMDP approach, we can account for other reasonable values.

5.1.7. Estimation of Immediate and Lump-Sum Rewards.

As introduced in Section 3, the immediate reward, $r_n(s,a)$, represents the quality of life 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 let $r_n(s^T,a)$ 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 $n \le N-1$, $r_n(s,a) = \langle r_n(s^T,a), r_n(s^D,a)\rangle$, where

$$r_n(s^T,a) = \begin{cases} qol(\text{organ rejection})/12, \\ \quad \text{if } s^T = \text{organ rejection}, \\ qol(\text{organ survival})/12, \\ \quad \text{if } s^T = \text{organ survival (different } C_0\text{'s}); \end{cases} \tag{17a}$$

$$r_n(s^D,a) = \begin{cases} qol(\text{diabetes})/12, & \text{if } s^D = \text{diabetic}, \\ qol(\text{prediabetes})/12, & \text{if } s^D = \text{prediabetic}, \\ qol(\text{healthy})/12, & \text{if } s^D = \text{healthy}. \end{cases} \tag{17b}$$

In (17a) and (17b), 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).

Table 4. Parameters for Calculating Specificity (Spec) and Sensitivity (Sens) of Observing Medical Test Results

Notation	Description	Value
sp^{FPG}_{100}	Spec: healthy (FPG <100 mg/dL)	85%
sp^{FPG}_{126}	Spec: healthy/prediabetes (FPG <126 mg/dL)	90%
sn^{FPG}_{100}	Sens: prediabetes/diabetes (FPG ≥ 100 mg/dL)	90%
sn^{FPG}_{126}	Sens: diabetes (FPG ≥ 126 mg/dL)	85%
$sp^{A1C}_{5.7}$	Spec: healthy (HbA1c $<5.7\%$)	85%
$sp^{A1C}_{6.5}$	Spec: healthy/prediabetes (HbA1c $<6.5\%$)	90%
$sn^{A1C}_{5.7}$	Sens: prediabetes/diabetes (HbA1c $\ge 5.7\%$)	90%
$sn^{A1C}_{6.5}$	Sens: diabetes (HbA1c $\ge 6.5\%$)	85%
sp^H	Spec: healthy (based on FPG and HbA1c)	see (16a)
sp^{PD}	Spec: healthy/prediabetes (based on FPG and HbA1c)	see (16a)
sn^{PD}	Sens: prediabetes/diabetes (based on FPG and HbA1c)	see (16b)
sn^D	Sens: diabetes (based on FPG & HbA1c)	see (16b)
sp^T_8	Spec: low C_0 (Architect threshold <8 mg/dL)	85%
sp^T_{10}	Spec: low/medium C_0 (Architect threshold <10 mg/dL)	90%
sn^T_8	Sens: medium/high C_0 (Architect threshold ≥ 8 mg/dL)	90%
sn^T_{10}	Sens: high C_0 (Architect threshold ≥ 10 mg/dL)	85%

Furthermore, the lump-sum reward denoted by $R_n(s)$ is the QALE that a patient receives based on the core state s whenever she or 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

$$R_n(s) = \frac{qol(s)\left(1 - e^{-r\, RLE(s,n)}\right)}{r} \quad \forall\, s \in S, \forall\, n \leq N, \quad (18)$$

where r is a discount rate that accounts for degradation of the core health state over the remaining lifetime of a patient. In (18), $qol(s) = \langle qol(s^T), qol(s^D)\rangle$, 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 as in (17a) and (17b). Further details about estimating the required parameters (e.g., *qol* and *RLE* scores) can be found in Online Appendix B.1. When comparing our optimal policies with other benchmarks in Section 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 organ rejection outcomes can have a higher impact compared to diabetes 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 (Section 5.2.1) and comparison of our optimal policies with other policies including the current medical practice (Section 5.2.2). As we will discuss, these results have important implications for guideline 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 in Table 2. To illustrate our results for each of these cohorts and for computational tractability, we consider three different values for the DM's conservatism level (i.e., $\lambda \in \{0.0, 0.5, 1.0\}$) and three models for the ambiguity set (i.e., $|M| = 3$). We also set the KL divergence bound ϵ in Table 3 as 0.05. We consider 0.05 instead of lower values such as 0.01 or 0.02 simply to increase the likelihood of satisfying the model informativeness condition. Furthermore, we use a 2-simplex to represent a cut of the belief space under a specific concentration of tacrolimus. 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, although the exact diabetes status is not perfectly known). Although we calculate optimal medications over the entire belief space Π_{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. We aim to provide insights for the medical practice into the following questions:

Question 1. What is the impact of risks of organ rejection and diabetes complications on the optimal medication regimens?

Question 2. What is the impact of various patient risk factors on the optimal medication regimens?

Question 3. What is the impact of DM's conservatism levels on the optimal medication regimens?

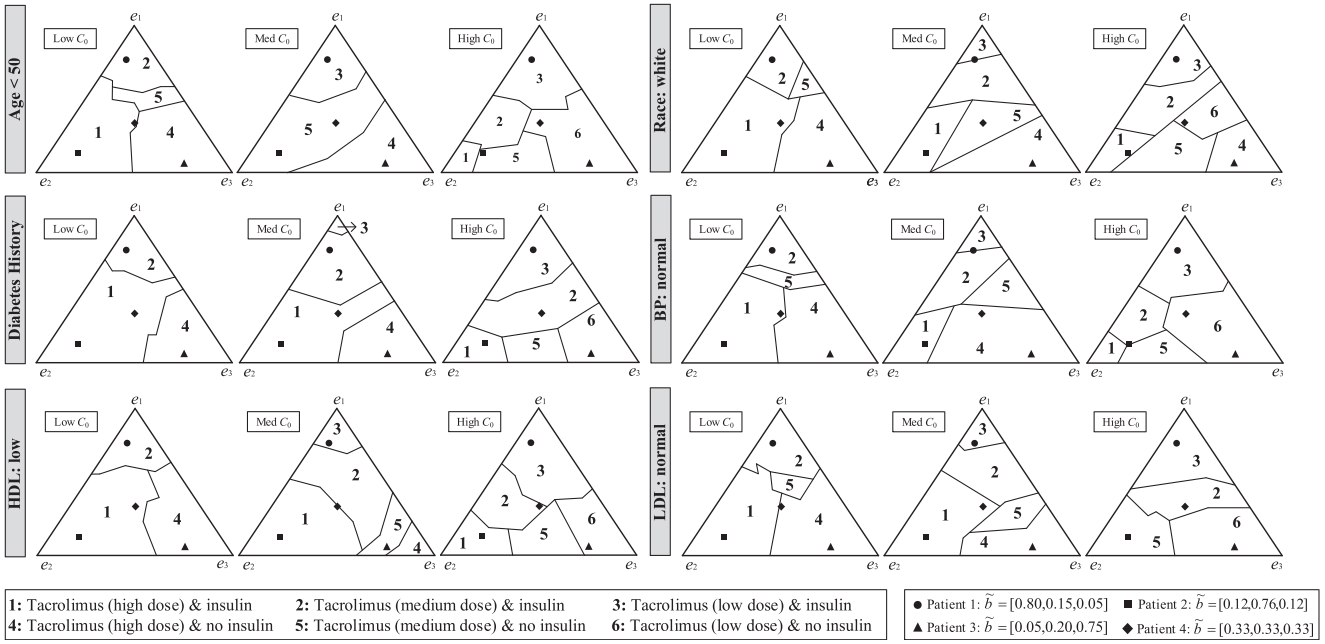
To address these three questions, we summarize our main findings in Observations 1–3 and discuss their implications for the medical practice.

Remark 1. Based on the discussion in Section 2, our observations and implications here are not predictive of what a physician will do under a specific conservatism level. They are rather prescriptive in that they shed light on what a physician should be doing (given his or her conservatism level) based on the optimal policies we find from our APOMDP approach. Because we are able to characterize the optimal policy for any given level of conservatism, we are also able to shed light on the optimal policy that is based on the best conservatism level.

Observation 1 (Optimal Medication Policies). (i) Under low or medium C_0, the optimal tacrolimus regimen is to use the high dose as long as the risk of diabetes is not very high. However, as this risk increases, using less intensive tacrolimus regimens (e.g., medium or low dose) becomes optimal. (ii) Under high C_0, it is optimal to use low-dose tacrolimus regardless of the underlying risk of diabetes. (iii) When tacrolimus is prescribed in medium or high dose, insulin should be used to avoid the potential onset of diabetes, even when the patient has a considerable chance of being diabetes-free.

To better understand Observation 1, let us consider (a) different levels of C_0 (to reflect on different risks of organ rejection), and (b) four patients each corresponding to a specific belief vector (to represent different risks of diabetes complications). These patients are identified in Figure 3 (and Figure EC.3 in Online Appendix C) via vectors $\tilde{b}$. For example, patient 1 has $\tilde{b} = [0.80, 0.15, 0.05]$ (i.e., 80%, 15%, and 5% risks (*perceived* by the DM) of being diabetic, pre-diabetic, and healthy, respectively). Patients 1, 2, and 3 have high risks of being diabetic, pre-diabetic, and

Figure 3. Optimal Medication Policy Regions (Numbers 1–6) for the First Visit Based on Different Risks of Organ Rejection (C_0 Levels) and Diabetes Complications



Note. The terms e_1, e_2, and e_3 represent diabetic, prediabetic, and healthy conditions, respectively; e_j denotes a unit vector with the jth element equal to 1 and other elements equal to 0. Results are for $\lambda = 0.5$.

healthy, respectively, whereas patient 4 has an equal risk among these three conditions. We present the following results from Figure 3 and Figure EC.3:

Low C_0. When the risk of diabetes is not very high (e.g., for Patients 2–4), the optimal tacrolimus regimen is the high dose, which is consistent with the current practice. However, unlike the current practice, we observe that for patients with a high risk of diabetes (e.g., Patient 1) the optimal tacrolimus regimen is the medium dose (for all patient cohorts). In addition, the optimal insulin regimen for Patient 1 (3) is to use (not use) insulin. However, unlike the current practice, insulin is the optimal regimen even when the risk of diabetes is lowered compared with Patient 1: Patient 2 under all cohorts and Patient 4 under all cohorts except being nonwhite and female with no diabetes history and normal levels of Chol, HDL, and LDL.

Medium C_0. When C_0 is medium, using medium-dose tacrolimus is the first choice in the current practice. However, we find that when the risk of diabetes is low (e.g., Patients 2 and 3), the optimal tacrolimus regimen is the high dose (for all patient cohorts). As the diabetes risk increases (e.g., Patients 1 and 4), we find that the optimal tacrolimus regimen becomes the low/medium dose for non-obese, female patients with age < 50, hypertension, normal HDL, and high levels of LDL and TG. In addition, as for patients with low C_0, we observe that for patients with medium C_0, it is optimal to use insulin even when the diabetes risk is relatively low (unlike the current practice). For example, in addition to Patient 1, we find that Patients 2 and 4 (i.e., those with lower risk of diabetes compared with Patient 1) should also be prescribed insulin (for patient cohorts formed by high levels of all risk factors except Chol).

High C_0. When C_0 is high, organ rejection is unlikely, and hence using a low (or medium) dose of tacrolimus is recommended over a high dose in medical practice. Our results confirm the optimality of this recommendation for all patient cohorts. However, as the diabetes risk is lowered (e.g., Patients 3 and 4), using low/medium-dose tacrolimus is optimal only for specific patient cohorts (e.g., nonwhite patients with age < 50 and normal levels of BP and Chol). Also, unlike the current practice, we find that even for patients whose risk of diabetes is not very high (e.g., Patients 2 and 4), it is optimal to use insulin (for obese, female patients with age ≥ 50, diabetes history, and high LDL).

In Observation 1, we addressed Question 1 (i.e., how the optimal medication regimens are affected by different risks of organ rejection and diabetes complications). In the next two observations, we explore the impact of variations in risk factors (Question 2) and the DM's conservatism level λ (Question 3) on medication regimens. Therefore, instead of specific belief vectors (e.g., Patients 1–4 in Observation 1), we consider all belief vectors (i.e., all patients). In particular, we utilize the optimal policy regions depicted in Figure 3 (and Figure EC.3 in Appendix C) and make the following observation.

Observation 2 (Tacrolimus Requirement and the Diabetogenic Effect). Under any conservatism level λ, (i) the optimal policy region for using high-dose tacrolimus is larger for non-white, male, obese patients with age $\geq$ 50, hypertension, low HDL, and high LDL (compared to cohorts formed by the opposing risk levels along each of these risk factors), and (ii) the optimal policy region for using insulin (along with high/medium-dose tacrolimus) is larger for male patients with age $\geq$ 50, diabetes history, hypertension, high Chol, and low HDL (compared with cohorts formed by the opposing risk levels along each of these risk factors).

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 other risk factors such as age, gender, race, BMI, blood pressure, HDL, and LDL. This implies that such risk factors could make patients more vulnerable to the risk of organ rejection, and hence, to offset this effect, the optimal tacrolimus regimens put more emphasis on higher dosages of tacrolimus for such patients. In addition, regarding Observation 2(ii), Figure 3 and Figure EC.3 show (as an example) that the policy regions for actions a_1 and a_2 (i.e., using insulin along with a medium/high dosage of tacrolimus) are larger for patients with age $\geq$ 50 compared with those with age $<$ 50. Observation 2(ii) reveals risk factors under which the diabetogenic effect of tacrolimus is stronger. These findings address Question 2 and are useful for the medical practice, especially because they highlight that the blood glucose level of patients with specific risk factors should be monitored more closely than other patients in the post-transplant period.

Finally, we address Question 3 by making the following observation.

Observation 3 (The Effect of Conservatism Levels). Increasing the conservatism level, λ, results in using (i) more intensive medication regimens (for both tacrolimus and insulin) for non-white patients with age $\geq$ 50, no diabetes history, and low-risk levels of Chol, HDL, LDL, TG, UA, and BMI (both non-obese and obese), and (ii) less intensive tacrolimus regimens for male patients with age $<$ 50, diabetes history, hypertension, and high-risk levels of Chol, HDL, and LDL. However, increasing λ 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.

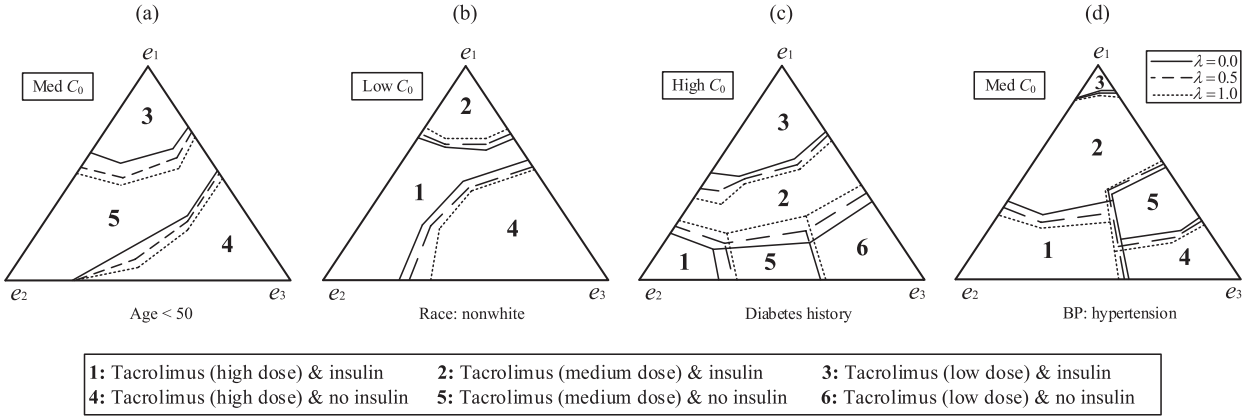
For example, as can be observed from Figure 4(b), for a non-white patient, a higher conservatism level results in larger optimal policy regions for using high-dose tacrolimus (as opposed to medium dose) and insulin (as opposed to not using it). On the other hand, based on Figure 4(a), (c), and (d), we find that for a patient with age $<$ 50, diabetes history, or hypertension, increasing the conservatism level results in smaller optimal policy regions in which higher dose of tacrolimus is prescribed. Regarding this observation, in Section 4, we explored relevant analytical results via Theorem 1 and Corollary 1. In particular, we presented sufficient conditions under which an increase in the conservatism level λ (compared to a baseline level) results in more (or less) intensive medications regimens (equivalently, a larger (or smaller) optimal policy region for such regimens).

Observation 3 has other implications for the medical practice. For non-white patients with age $\geq$ 50, no diabetes history, and normal levels of Chol, HDL, LDL, TG, and UA, Observation 3 implies that a more conservative DM should be more concerned about both risks of organ rejection and NODAT compared to a less conservative DM (which, in turn, results in elevating the intensity of both regimens). However, for male patients with age $<$ 50, diabetes history, hypertension, and high-risk levels of Chol, HDL, and LDL, a more conservative DM should be more concerned about the potential risk of NODAT than that of organ rejection compared to a less conservative DM. This may be due to the diabetogenic effect of tacrolimus, which could make the more conservative DM prescribe less intensive tacrolimus regimens. Also, for white, female patients with 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, there is no significant difference between a more conservative DM and a less conservative one in utilizing medications optimally to balance risks of organ rejection and diabetes complications.

Finally, Observation 3 reveals that variations in physicians' attitudes toward ambiguity will not show a homogeneous pattern with respect to the intensity of the drugs used, if physicians follow the optimal policy. Thus, drug intensification (i.e., use of intensified levels of medication regimens) observed in the current practice should not be attributed merely to physicians' behavior toward ambiguity. Instead, our findings suggest 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 aim to show the potential impact of considering the ambiguity caused by model misspecifications and the partial observability of medical tests. To this end, we have developed a microsimulation model (see Online Appendix D) to simulate costs and patients' life expectancies during the planning horizon under (1) the optimal policies obtained from our

Figure 4. Variations in Optimal Medication Policies (in the Same Period) Based on Different Conservatism Levels



Note. The terms e_1, e_2,and e_3 represent diabetic, prediabetic, and healthy conditions, respectively; e_j denotes a unit vector with the jth element equal to 1 and other elements equal to 0.

APOMDP approach, (2) four benchmark policies that resemble the current medical practice under different scenarios, and (3) a policy that is obtained by a traditional POMDP (i.e., by ignoring the underlying ambiguity; see Online Appendix D for more details).

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 since transplant. However, there is currently no consensus among physicians on how C_0 level and elapsed time should be incorporated in prescribing tacrolimus (see, e.g., Staatz and Tett 2004, Schiff et al. 2007). To address this variation among physicians, we consider four different benchmark policies that are typically used in the current practice (see Table 5). As Table 5 shows, for the first three months after transplant, tacrolimus is prescribed in high dosages 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. 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., observation o_2, o_5, or o_8) during months 4–6 after 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 compare the APOMDP, POMDP, and benchmark approaches based on three performance measures: (1) average QALE achieved, (2) average medical expenditures (see Online Appendix B.2 for related cost estimations), and (3) average number of times that insulin and different dosage of tacrolimus are prescribed (Tables 6 and 7 show the results). The latter allows us to examine whether our methodology yields less intensive medication regimens compared to the current practice. Furthermore, because dynamic risk factors are subject to change throughout the time horizon, in our 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 of low-risk and high-risk levels of

Table 5. Description of Benchmark Policies Based on Medical Observations and Time Elapsed Since Transplant

		Benchmark			
Month	Observation	1	2	3	4
1–3	$\boldsymbol{o}_1$	a_1	a_1	a_1	a_1
	$\boldsymbol{o}_2$	a_1	a_1	a_1	a_1
	$\boldsymbol{o}_3$	a_1	a_1	a_1	a_1
	$\boldsymbol{o}_4, \boldsymbol{o}_7$	a_4	a_4	a_4	a_4
	$\boldsymbol{o}_5, \boldsymbol{o}_8$	a_4	a_4	a_4	a_4
	$\boldsymbol{o}_6, \boldsymbol{o}_9$	a_4	a_4	a_4	a_4
4–6	$\boldsymbol{o}_1$	a_1	a_1	a_1	a_1
	$\boldsymbol{o}_2$	a_1	a_1	a_2	a_2
	$\boldsymbol{o}_3$	a_1	a_2	a_3	a_3
	$\boldsymbol{o}_4, \boldsymbol{o}_7$	a_4	a_4	a_4	a_4
	$\boldsymbol{o}_5, \boldsymbol{o}_8$	a_4	a_4	a_5	a_5
	$\boldsymbol{o}_6, \boldsymbol{o}_9$	a_4	a_5	a_6	a_6
7–12	$\boldsymbol{o}_1$	a_1	a_1	a_2	a_2
	$\boldsymbol{o}_2$	a_2	a_2	a_2	a_3
	$\boldsymbol{o}_3$	a_2	a_3	a_3	a_3
	$\boldsymbol{o}_4, \boldsymbol{o}_7$	a_4	a_4	a_5	a_5
	$\boldsymbol{o}_5, \boldsymbol{o}_8$	a_5	a_5	a_5	a_6
	$\boldsymbol{o}_6, \boldsymbol{o}_9$	a_5	a_6	a_6	a_6

Table 6. Comparison of Medication Policies (Based on Average QALE and Cost)

		Policy						Improvement of optimal policy over:				
Measure	Cohort	Benchmark 1	Benchmark 2	Benchmark 3	Benchmark 4	POMDP	Optimal	Benchmark 1 (%)	Benchmark 2 (%)	Benchmark 3 (%)	Benchmark 4 (%)	POMDP (%)
QALE(yrs)	Age < 50	16.36	16.56	16.98	16.94	17.08	17.14	4.77	3.50	0.94	1.18	0.35
	Age ≥ 50	9.15	9.33	9.80	9.75	9.82	9.85	7.65	5.57	0.51	1.03	0.31
	Gender: female	17.74	17.88	18.07	18.10	18.12	18.17	2.42	1.62	0.55	0.39	0.28
	Gender: male	15.35	15.51	15.88	15.91	15.90	15.93	3.78	2.71	0.31	0.13	0.19
	Race: white	16.16	16.25	16.58	16.60	16.62	16.68	3.22	2.65	0.60	0.48	0.36
	Race: nonwhite	13.25	13.41	14.04	14.02	13.97	14.02	5.81	4.55	**−0.14**	**0.00**	0.36
	Diabetes history: no	14.66	14.76	15.01	15.14	15.15	15.19	3.62	2.91	1.20	0.33	0.26
	Diabetes history: yes	8.32	8.52	8.73	8.75	8.94	8.96	7.69	5.16	2.63	2.40	0.22
	BMI	13.65	13.79	14.10	14.08	14.12	14.15	3.66	2.61	0.35	0.50	0.21
	BP	13.34	13.52	13.82	13.77	13.80	13.89	4.12	2.74	0.51	0.87	0.65
	Chol	13.06	13.18	13.56	13.50	13.60	13.68	4.75	3.79	0.88	1.33	0.59
	HDL	13.08	13.22	13.49	13.55	13.71	13.78	5.35	4.24	2.15	1.70	0.51
	LDL	13.20	13.37	13.86	13.90	13.82	13.90	5.30	3.96	0.29	**0.00**	0.58
	TG	13.05	13.09	13.46	13.40	13.47	13.51	3.52	3.21	0.37	0.82	0.30
	UA	12.86	12.90	13.08	13.15	13.23	13.25	3.03	2.71	1.30	0.76	0.15
	Average[a]							4.58	3.46	0.83	0.79	0.35
Cost($)	Age < 50	5,733	5,507	5,371	5,405	5,163	5,098	12.46	8.02	5.36	6.02	1.28
	Age ≥ 50	5,811	5,674	5,478	5,495	5,352	5,245	10.79	8.18	4.44	4.77	2.04
	Gender: female	5,950	5,879	5,650	5,536	5,521	5,415	9.88	8.57	4.34	2.23	1.96
	Gender: male	5,963	5,884	5,611	5,588	5,640	5,584	6.79	5.37	0.48	0.07	1.00
	Race: white	5,715	5,630	5,253	5,204	5,144	5,113	11.77	10.11	2.74	1.78	0.61
	Race: nonwhite	6,377	6,205	5,417	5,461	5,504	5,452	16.97	13.81	**−0.64**	0.17	0.95
	Diabetes history: no	5,770	5,691	5,347	5,223	5,218	5,180	11.39	9.86	3.22	0.83	0.73
	Diabetes history: yes	6,863	6,814	6,630	6,597	6,212	6,147	11.65	10.85	7.86	7.32	1.06
	BMI	5,895	5,798	5,520	5,584	5,506	5,443	8.30	6.52	1.41	2.59	1.16
	BP	5,881	5,815	5,451	5,566	5,488	5,274	11.51	10.26	3.36	5.54	4.06
	Chol	6,111	6,005	5,603	5,688	5,566	5,494	11.23	9.30	1.98	3.53	1.31
	HDL	5,835	5,805	5,619	5,545	5,329	5,144	13.43	12.85	9.23	7.80	3.60
	LDL	5,925	5,836	5,497	5,383	5,578	5,391	9.91	8.25	1.97	**−0.15**	3.47
	TG	6,252	6,118	5,780	5,916	5,774	5,616	11.32	8.94	2.92	5.34	2.81
	UA	5,944	5,831	5,674	5,513	5,423	5,280	12.58	10.44	7.46	4.41	2.71
	Average[a]							11.57	9.73	4.23	4.01	1.93

Note. Numbers in bold represent cases where the corresponding policy performs as well as (or better than) the optimal policy.
[a]The average is taken over the percentage of improvement across all cohorts under a policy.

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 6 and 7. We make the following observations from the results presented in Tables 6 and 7:

Observation 4 (Impact). During one year after transplant, compared to other polices (i.e., Benchmarks 1–4 and POMDP), our optimal policy, on average, (i) improves the QALE per patient up to 4.58%, (ii) reduces the medical expenditures per patient up to 11.57%, and (iii) prescribes high-dose tacrolimus up to 3.69 fewer times per patient, medium-dose tacrolimus up to 1.48 more times per patient, low-dose tacrolimus up to 2.09 fewer times per patient, and insulin up to 2.12 more times per patient.

Based on Observation 4 and the results provided in Tables 6 and 7, we shed light on the following implications for medical practitioners, as well as those influencing medical guidelines and recommendations: (1) The improvements in QALE and cost made by our optimal policy are not uniform across all cohorts of patients. From Table 6, 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. (2) Gains obtained by following our proposed policies compared to the current practice are 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 2 in one year (compared to our optimal policy), a patient takes high-dose tacrolimus up to 3.69 more times, while taking insulin up to 2.09 fewer times. As a result, the patient becomes more vulnerable against the diabetogenic effect of tacrolimus and NODAT complications. (3) The comparison between our APOMDP approach and the POMDP approach 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 during one year after transplant. This shows the importance of considering model misspecifications that are inevitable when data are used to estimate parameters: one should not rely on a single model to derive effective medication strategies. (4) The abovementioned improvements in performance measures are obtained over our planning horizon (i.e., one year after transplant). Because, compared to other approaches, the APOMDP

Table 7. Comparison of Medication Policies (Based on the Average Number of Medications Prescribed Under Each Policy)

	Benchmark 1				Benchmark 2				Benchmark 3				Benchmark 4				POMDP				Optimal			
Cohort	1‡	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Age: L[a]	7.76	3.91	0.33	3.97	7.05	2.96	1.99	4	3.99	5.12	2.89	4.07	4.05	3.15	3.8	3.98	4.07	4.35	2.58	5.24	3.66	4.87	2.47	5.12
Age: H	8.13	3.87	0	4.21	6.73	3.08	2.19	4.1	4.13	5.39	2.48	4.17	4.13	3.29	3.58	4.15	4.88	4.33	1.79	7.48	4.68	4.81	1.51	7.62
Gender: L	8.1	3.65	0.25	3.72	7.15	3.25	1.6	4.37	4.24	5.32	2.44	4.45	3.55	2.68	4.77	3.95	4.77	4.05	2.18	5.18	4.13	4.44	2.43	4.94
Gender: H	8.46	3.54	0	4.48	7.08	3.35	1.57	4.28	4.41	4.99	2.6	3.96	4.37	2.73	3.9	4.15	5.18	3.89	1.93	5.47	5.05	4.03	1.92	5.70
Race: L	7.29	4.17	0.54	3.85	6.64	3.1	2.26	4.3	4.16	4.9	2.94	3.6	3.77	2.8	4.43	3.81	4.71	3.83	2.46	4.84	4.65	3.97	2.38	4.89
Race: H	7.56	4.35	0.09	4.34	6.54	3.47	1.99	4.3	3.62	5.24	3.14	3.84	3.72	2.79	4.49	4.37	5.05	3.87	2.08	5.77	4.83	4.06	2.11	5.85
Diabetes history: L	8.38	3.46	0.16	3.89	6.97	2.8	2.23	4.34	4.36	5.24	2.4	3.79	4.12	2.58	4.3	3.88	4.65	3.81	2.54	5.31	4.12	4.27	2.61	4.85
Diabetes history: H	7.54	4.14	0.32	4.07	6.64	2.9	2.46	4.55	3.98	5.53	2.49	4.24	3.74	2.65	4.61	4.41	3.95	4.41	2.64	7.50	3.25	5.03	2.72	8.14
BMI	7.82	3.55	0.63	4.3	7.42	2.9	1.68	3.69	4.08	5.5	2.42	4.34	3.83	2.53	4.64	4.39	4.83	4.17	2.00	6.17	4.15	4.79	2.06	6.88
BP	8.3	3.7	0	4.4	7.36	2.54	2.1	4.26	3.94	5.13	2.93	3.96	4.21	2.64	4.15	4.2	5.07	3.85	2.08	5.96	4.25	4.70	2.05	6.38
Chol	7.86	3.94	0.2	4.11	6.79	2.96	2.25	4.15	4.14	4.93	2.93	4.11	3.73	3.46	3.81	4.11	5.02	3.94	2.04	6.13	4.34	4.79	1.87	6.78
HDL	8.35	3.53	0.12	3.92	7.41	3.09	1.5	3.96	4.12	5.77	2.11	3.98	3.96	3.64	3.4	4.28	4.56	3.76	2.68	6.07	3.78	4.69	2.53	6.67
LDL	7.94	4.06	0	4.46	7.32	3.01	1.67	4.38	3.69	4.99	3.32	4.03	3.82	3.29	3.89	4.34	5.14	3.66	2.20	6.13	4.55	4.52	1.93	6.98
TG	7.61	3.62	0.77	3.96	6.83	3.39	1.78	4.34	3.72	5.19	3.09	4.46	3.56	3.28	4.16	4.08	4.66	3.38	2.96	6.13	4.15	4.03	2.82	6.56
UA	7.74	4.26	0	4.26	6.53	3.23	2.24	4.11	4.24	5.77	1.99	4.4	3.87	3.48	3.62	4.11	4.22	4.03	2.75	5.68	3.97	4.17	2.86	5.88
Avg. diff.	3.69	−0.63	−2.06	−2.09	2.73	−1.41	−0.32	−2.01	−0.18	0.79	0.39	−2.12	−0.34	−1.48	1.82	−2.07	0.48	−0.52	0.04	−0.28				

Notes. L, Low level; H, high level; Avg. diff., average of differences with the optimal policy is taken over all cohorts.
[a]Numbers 1, 2, 3, and 4 are the number of times (on average) that high-, medium-, and low-dose tacrolimus and insulin are prescribed during one year, respectively.

approach could (a) result in better outcomes in each time period and (b) move the patient to a better health state over time, the potential improvements could be more significant if these measures are calculated over a longer horizon (e.g., two years since transplant).

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, 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 model that maximizes the 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 under the optimal policy for some patient cohorts (e.g., nonwhite patients with age ≥ 50, no diabetes history, and low cholesterol), a more conservative physician is more concerned about both risks of organ rejection and NODAT than a less conservative physician. Also, for other patient cohorts (e.g., male patients with age < 50, diabetes history, and hypertension), a more conservative DM is more concerned (under the optimal policy) about the risk of NODAT than that of organ rejection compared with a less conservative physician.

We also compare our proposed optimal policies with four benchmark policies that represent the current medical practice (under different scenarios) and a POMDP approach that ignores the underlying model misspecifications. Our results show that, depending on different risk factors considered for each patient, in one year after transplant our optimal policy (compared to other policies) (a) improves the average QALE up to 4.58%, (b) reduces the medical expenditures per patient up to 11.57%, and (c) prescribes high-dose tacrolimus up to 3.69 fewer times per patient. The other important implications of the above-mentioned results for practitioners and guideline makers are as follows: (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 guideline makers should not rely on a single model to derive effective medication strategies: 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 versus 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 United States in 2015; United Network of Organ Sharing 2018). 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 Section 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 with 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

Ahn D, Choi S, Gale D, Kariv S (2014) Estimating ambiguity aversion in a portfolio choice experiment. *Quant. Econom.* 5(2): 195–223.

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.

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

American Heart Association (2018) Understanding blood pressure readings. Accessed May 5, 2017, http://www.heart.org/HEARTORG/Conditions/HighBloodPressure/KnowYourNumbers/Understanding-Blood-Pressure-Readings_UCM_301764_Article.jsp#.WzQibadKg2w.

Arad A, Gayer G (2012) Imprecise data sets as a source of ambiguity: A model and experimental evidence. *Management Sci.* 58(1): 188–202.

Ata B, Skaro A, Tayur S (2016) OrganJet: Overcoming geographical disparities in access to deceased donor kidneys in the United States. *Management Sci.* 63(9):2776–2794.

Ayer T, Alagoz O, Stout NK (2012) OR forum—A POMDP approach to personalize mammography screening decisions. *Oper. Res.* 60(5):1019–1034.

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. *J. Pharmaceutical Biomedical Anal.* 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 U.S. organ and tissue transplant cost estimates and discussion. Report, Milliman, Brookfield, WI.

Berger L, Bleichrodt H, Eeckhoudt L (2013) Treatment decisions under ambiguity. *J. Health Econom.* 32(3):559–569.

Bertsimas D, Farias VF, Trichakis N (2013) Fairness, efficiency, and flexibility in organ allocation for kidney transplantation. *Oper. Res.* 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.

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

Buuren S, Groothuis-Oudshoorn K (2011) MICE: Multivariate imputation by chained equations in R. *J. Statist. Software* 45(3): 1–67.

Chakkera HA, Weil EJ, Castro J, Heilman RL, Reddy KS, Mazur MJ, Hamawi K, et al. (2009) Hyperglycemia during the immediate period after kidney transplantation. *Clinical J. Amer. Soc. Nephrology* 4(4):853–859.

Chen Y, Katuščák P, Ozdenoren E (2007) Sealed bid auctions with ambiguity: Theory and experiments. *J. Econom. Theory* 136(1): 513–535.

Delage E, Mannor S (2010) Percentile optimization for Markov decision processes with parameter uncertainty. *Oper. Res.* 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 Oper. Management* 16(3):381–400.

Ghirardato P, Maccheroni F, Marinacci M (2004) Differentiating ambiguity and ambiguity attitude. *J. Econom. Theory* 118(2): 133–173.

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 (2018) Data uncertainty in Markov chains: Application to cost-effectiveness analyses of medical innovations. *Oper. Res.* 66(3):697–715.

Han PKJ, Reeve BB, Moser RP, Klein WMP (2009) Aversion to ambiguity regarding medical tests and treatments: Measurement, prevalence, and relationship to sociodemographic factors. *J. Health Comm.* 14(6):556–572.

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

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

Kaufman DL, Schaefer AJ, Roberts MS (2011) Living-donor liver transplantation timing under ambiguous health state transition probabilities—Extended abstract. Accessed January 10, 2017, http://www-personal.umich.edu/~davidlk/pubs/robustLivingDonor.pdf.

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

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

MedPlus (2018) Medical encyclopedia. Accessed May 5, 2017, https://www.nlm.nih.gov/medlineplus/encyclopedia.html.

Monahan GE (1982) State of the art—A survey of partially observable Markov decision processes: Theory, models, and algorithms. *Management Sci.* 28(1):1–16.

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

Organ Procurement and Transplantation Network (2011) OPTN/SRTR annual report: Transplant data 1999–2008. Accessed February 23, 2017, https://srtr.transplant.hrsa.gov/archives.aspx.

Papadimitriou CH, Tsitsiklis JN (1987) The complexity of Markov decision processes. *Math. Oper. Res.* 12(3):441–450.

Peysakhovich A, Karmarkar UR (2015) Asymmetric effects of favorable and unfavorable information on decision making under ambiguity. *Management Sci.* 62(8):2163–2178.

Saghafian S (2018) Ambiguous partially observable Markov decision processes: Structural results and applications. *J. Econom. Theory* 178:1–35.

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

Schiff J, Cole E, Cantarovich M (2007) Therapeutic monitoring of calcineurin inhibitors for the nephrologist. *Clinical J. Amer. Soc. Nephrology* 2(2):374–384.

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

Staatz CE, Tett SE (2004) Clinical pharmacokinetics and pharmacodynamics of tacrolimus in solid organ transplantation. *Clinical Pharmacokinetics* 43(10):623–653.

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.

Steimle LN, Kaufman DL, Denton BT (2018) Multi-model Markov decision processes: A new method for mitigating parameter ambiguity. Working paper, University of Michigan, Ann Arbor.

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

United Network of Organ Sharing (2018) Transplant trends. Accessed May 20, 2017, https://unos.org/data/transplant-trends/#transplants_by_organ_type+year+2017.

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

Whitt W (1982) Multivariate monotone likelihood ratio and uniform conditional stochastic order. *J. Appl. Probab.* 19(3):695–701.

Xu H, Mannor S (2012) Distributionally robust Markov decision processes. *Math. Oper. Res.* 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, Raleigh.

Zhang Y, Steimle LN, Denton BT (2017) Robust Markov decision processes for medical treatment decisions. Working paper, University of Michigan, Ann Arbor.