Robust Markov Decision Processes for Medical Treatment Decisions

Yuanhui Zhang
Operations Research Program, North Carolina State University, yuanhui.zhang@gmail.com

Brian T. Denton
Department of Industrial and Operations Engineering, University of Michigan, btdenton@umich.edu

Medical treatment decisions involve complex tradeoffs between the risks and benefits of various treatment options. The diversity of treatment options that patients can choose over time, and uncertainties in future health outcomes result in a difficult sequential decision making problem. Markov decision processes (MDPs) are commonly used to study medical treatment decisions; however, optimal policies obtained by solving MDPs may be affected by the uncertain nature of the model parameter estimates. In this article, we present a robust Markov decision process treatment model (RMDP-TM) with a controllable uncertainty set formulation for the transition probability matrices (TPMs) of the underlying Markov chain. We show that the RMDP-TM can overcome the common problem of over-conservativeness of the worst-case optimal policy obtained from the RMDP model with fixed uncertainty set formulations for TPMs. We present theoretical analysis to establish computationally efficient methods to solve the RMDP-TM and present its application to a medical treatment decision problem of optimizing the sequence and the start time to initiate medications for glycemic control for patients with type 2 diabetes.

Key words: robust optimization; robust Markov decision process; medical decision making; type 2 diabetes; glycemic control
1. Introduction

Medical treatment decisions for chronic diseases involve complex tradeoffs between benefits and harms of treatment. These decisions are often made by physicians based on results from randomized control trials and/or observational studies. However, decisions are rarely quantified in a way that makes clear the short term costs of medications and harms from medication side effects versus the long term benefits of avoiding disease-related complications. For many chronic diseases, there are multiple treatment options that can potentially be selected over the course of a patient’s lifetime. This results in a difficult sequential decision making problem which seeks to optimally trade off short term harms (e.g. side effects, medication costs) with uncertain long term benefits (e.g. delaying or avoiding disease-related complications or death).

Markov decision process (MDP) models have been used to study optimal control of many medical treatment decisions including liver transplants, HIV, diabetes, and others (Alagoz et al. 2004, 2007, Shechter et al. 2008, Mason et al. 2014). A key component of every MDP model is the transition probability matrix (TPM) which describes stochastic changes in the system over time. The optimal policy of an MDP may be highly dependent on TPMs of the underlying Markov chain (Mannor et al. 2012). A strong assumption of MDP models is that TPMs is known with certainty. For medical treatment decisions, maximum likelihood estimates (MLEs) of the TPMs from a population-based observational data are commonly used; however, MLEs cannot capture the natural variation in transition probabilities caused by patient heterogeneity. Therefore, it is potentially valuable to develop optimization models which can also take into account this variation.

Robust MDP (RMDP) models in a max-min framework have been developed to generate the worst-case optimal policy when parameters in MDP models such as TPMs are subject to uncertainty (Satia and Lave Jr. 1973, White III and Eldeib 1994, Kaufman and Schaefer 2012, Iyengar 2005, Nilim and El Ghaoui 2005, Wiesemann et al. 2013). However, to the best of our knowledge, RMDPs have not yet been applied to health care applications, such as medical decision making problems. In this article, we propose an RMDP treatment model (RMDP-TM) with a controllable
uncertainty set formulation for the TPM for optimizing medical treatment decisions. Similar to the RMDP framework presented in (Nilim and El Ghaoui 2005, Iyengar 2005), we assume that the uncertainty sets of the TPMs satisfy the *rectangular uncertainty property*. In previous work, fixed uncertainty sets for TPMs are commonly used in RMDPs; however, a shortcoming of using such uncertainty set formulations is that they often provide overly-conservative optimal policies. To overcome this problem, we include an *uncertainty budget* as part of the uncertainty set formulation, so that the size of the proposed uncertainty set can be easily controlled.

We present theoretical analysis to establish computationally efficient methods to solve the RMDP-TM, and show that the RMDP-TM is computationally tractable based on the theoretical analysis and numerical experiments. We also present results about the structure of the optimal policy. We use the RMDP-TM version of the glycemic control model based on our previously published paper Zhang et al. (2014) to demonstrate the application of the proposed model to the context of optimizing treatment decisions for patients with type 2 diabetes. We use this example to show the computational efficiency of the proposed solution methods, and to draw conclusions about the feasibility and value of using the proposed RMDP-TM for real-world medical treatment decision problems.

The main contributions of this article are twofold: from the methodological point of view: (1) we present a new robust stochastic optimization model, the RMDP-TM, which is suited to medical treatment decision problems that arise in the context of chronic diseases; (2) we propose a new data-driven controllable uncertainty set model for TPMs, which can be used to control the conservativeness of the optimal policy; (3) we present computationally efficient methods for solving this model and theoretical analysis of the optimal policy. From the application point of view, (1) we present the first results from applying the RMDP to a real-world medical treatment decision problem in the context of glycemic control for patients with type 2 diabetes; (2) we utilize a validated glycemic control model with parameters estimated from a large longitudinal data set comprised of millions of patients’ medical records and pharmacy claims; (3) based on numerical results, we
draw conclusions about the effectiveness of various robust optimal treatment policies which can potentially provide guidance to clinicians and policy makers to make treatment decisions.

The remainder of this paper is organized as follows. In Section 2 we present the mathematical formulation of the RMDP-TM for general medical treatment decision problems. In Section 3, we present theoretical analysis and solution methods for solving the RMDP-TM. In Section 4 we demonstrate the application of the RMDP-TM to optimize the treatment decisions with regard to the sequence and time to initiate hyperglycemia lowering medications for patients with type 2 diabetes. Finally, in Section 5 we highlight our main conclusions.

2. Model Formulation

We formulate the medical treatment decision process as a finite-state, finite-horizon, discrete-time MDP where the underlying Markov chain represents the progression of a patient’s health status. The model includes the following five components:

**Time horizon.** We assume treatment decisions are made at a finite and discrete set of time epochs indexed by $t \in T = \{1, 2, \ldots, T\}$ where 1 represents the time of being diagnosed with a certain disease, and $T$ represents a reasonable upper limit on patients’ age (e.g. age 100). The period after the last epoch $T$ is called the post decision horizon.

**States.** The model includes health states, $L$, treatment states, $M$, and an absorbing state, $D$. The health state, $l_t \in L, \forall t \in T$, represents the severity of the disease at time epoch $t$. The treatment state, $m_t = (m_{1,t}, m_{2,t}, \ldots, m_{n,t}) \in M$, is an $n$–tuple binary vector in which $n$ represents the total number of available treatment options (e.g. available medications); $m_{i,t} = 1, \forall i \in \{1, \ldots, n\}$, represents the patient is on treatment option $i$ at the beginning of the time epoch $t$, otherwise, $m_{i,t} = 0$. Health states and treatment states influence the risk of disease-related complications that the treatment aims to prevent. The absorbing state, $D$, is commonly included in medical treatment decision models (Alagoz et al. 2004, 2007, Shechter et al. 2008, Kurt et al. 2011, Mason et al. 2014). It includes all major disease-related complications that the decision maker aims to prevent (e.g. heart attack, stroke, and renal failure) and death from any causes. The complete set of states in
the model is given by $S = \{L \times M\} \cup \{D\}$. Note that although $L$ and $M$ are defined independently, they are inter-dependent due to the effect of treatment actions on the probability of entering the absorbing state.

**Actions.** The action at time epoch $t \in T \setminus \{T\}$ is denoted by $\alpha_t(\ell_t, m_t) \in A_t$, and it represents the selection of treatment option(s) to initiate during the time period $(t, t+1]$ given the patient is in health state $\ell_t \in L$, and treatment state $m_t \in M$. We adopt the convention that, in finite horizon problems, decisions are not made at the last time epoch $T$. The set of all possible actions at time epoch $t$ is denoted by $A_t = \{(A_1, A_2, \ldots, A_n, t)| A_{i,t} = \{I, D\}, \forall i = \{1, \ldots, n\}\}$ where $A_{i,t} = I, \forall i = \{1, \ldots, n\}$, represents to initiate treatment option $i$ at time epoch $t$, otherwise $A_{i,t} = D$. For a patient in the absorbing state, there are no further actions, i.e., $A_t = \emptyset, t \in T \setminus \{T\}$. We assumed that no future treatment decision will be made during the post decision horizon, therefore the treatment state is assumed to be the same as $m_T$ during the post decision horizon. For ease of expression, we simplify the notation $\alpha_t(\ell_t, m_t)$ to $\alpha_t(s_t)$. Given a patient in treatment state, $m_t$, at the beginning of epoch $t$, and take the action, $\alpha_t(s_t)$, we denote $m_{t+1}(\alpha_t(s_t))$ to be the treatment state at the beginning of the next time epoch $t + 1$. A decision rule, $d_t : S \rightarrow A_t, \forall t \in T \setminus \{T\}$, specifies the action when the patient is in state $s_t \in S$ at time epoch $t \in T \setminus \{T\}$.

**Reward.** The immediate reward is denoted by $r_t(s_t, \alpha_t(s_t)), \forall t \in T \setminus \{T\}$, and it represents the reward received during the time period $(t, t+1]$ given being in state $s_t$ and taking action $\alpha_t(s_t)$ at time epoch $t$. The terminal reward, which defines the boundary condition of the model, is denoted by $r_T(s_T), \forall s_T \in S$, and it represents the expected total reward accumulated during the post decision horizon.

**Probabilities.** There are two types of transition probabilities in the model: probabilities of entering the absorbing state, and probabilities of transitioning among health states. We denote the probability of entering the absorbing state as $p_t^E(s_t, \alpha_t(s_t))$. For any $s_t \in L \times M$, $p_t^E(s_t, \alpha_t(s_t))$ represents the probability of having at least one disease-related complication or death occurs during the time period $(t, t+1]$. For $s_t = D$, $p_t^E(s_t, \alpha_t(s_t))$ represents the probability of staying in the absorbing
state which equals 1. We denote the transition probability between health states conditional on not entering the absorbing state during the time period \((t, t + 1]\) by \(q_{t, \ell_t}(\ell_{t+1}), \forall t \in T \setminus \{T\}, \ell_t, \ell_{t+1} \in L\).

The transition probability from state \(s_t \in S\) to state \(s_{t+1} \in S\) given taking action \(\alpha_t(s_t) \in A_t\) at time epoch \(t\) is defined as follows:

\[
p_{t}(s_{t+1|s_t, \alpha_t(s_t)}) = \begin{cases} 
(1 - p^E_t(s_t, \alpha_t(s_t)))q_{t, \ell_t}(\ell_{t+1}), & \text{if } s_t, s_{t+1} \in L \times M, \\
p^E_t(s_t, \alpha_t(s_t)), & \text{if } s_t \in L \times M, s_{t+1} = D, \\
1, & \text{if } s_t = D,
\end{cases}
\]

### 2.1. MDP formulation

A policy, \(\pi = \{d_1, d_2, \ldots, d_{T-1}\}\), is a sequence of decision rules that specifies the action to be used at each state in each time epoch. The policy induces a probability distribution on the set of all realizations of the MDP. Based on the expected total discounted reward criterion, the value function for a given policy \(\pi\) is defined as follows:

\[
v^\pi_t(s_t) \equiv \mathbb{E}^\pi_{s_t} \left[ \sum_{k=t}^{T-1} \lambda^{k-t}r_k(s_k, \alpha_k(s_k)) + \lambda^{T-t}r_T(s_T) \right], \forall s_t \in S, t \in T.
\]

where \(v^\pi_t(s_t)\) represents the expected total discounted reward accumulated from time epoch \(t\) onward. Based on the definition of the value function for a given policy, the optimal value function for the MDP can be written as follows:

\[
v^\text{MDP}_t(s_t) = \begin{cases} 
\max_{\pi \in \Pi} \mathbb{E}^\pi_{s_t} \left[ \sum_{k=t}^{T-1} \lambda^{k-t}r_k(s_k, \alpha_k(s_k)) + \lambda^{T-t}r_T(s_T) \right], & \forall s_t \in L \times M, \\
0, & \text{otherwise,}
\end{cases}
\]

where \(\Pi\) represents the set of all possible policies. It is well-known that the optimal value function of the MDP (2) can be rewritten recursively as follows:

\[
v^\text{MDP}_t(s_t) = \begin{cases} 
\max_{\alpha_t(s_t) \in A_t} \left\{ r_t(s_t, \alpha_t(s_t)) + \lambda(1 - p^E_t(s_t, \alpha_t(s_t))) \sum_{\ell_{t+1} \in L} q_{t, \ell_t}(\ell_{t+1}) \right\}, & \forall s_t = (\ell_t, m_t) \in L \times M, t \in T \setminus \{T\}, \\
r_T(s_T), & s_T = (\ell_T, m_T) \in L \times M, \\
0, & \text{otherwise},
\end{cases}
\]
where \( v^\text{MDP}_t(s_t) \) denotes the optimal value-to-go. The backward induction algorithm presented in Puterman (1994) is commonly used to solve MDP models. The MDP formulation assumes no uncertainty in the health state TPM, \( Q_t \triangleq [q_{t\ell}(\ell_{t+1})]_{|\mathcal{L}| \times |\mathcal{L}|} \), so we refer to it as the nominal TPM, and denoted by \( \hat{Q}_t \triangleq [\hat{q}_{t\ell}(\ell_{t+1})]_{|\mathcal{L}| \times |\mathcal{L}|} \), from hereafter.

2.2. RMDP-TM formulation

In many medical treatment decision problems, the health state TPM, \( Q_t \) is estimated from a longitudinal data set by using the maximum likelihood method (Craig and Sendi 2002); therefore, it is subject to uncertainty caused by sample variation. We assume that the health state TPM lies in a given uncertainty set \( Q_t \) which has the rectangular uncertainty property, i.e. \( Q_t = \prod_{\ell_t \in \mathcal{L}} Q_{t,\ell_t} \) where \( Q_{t,\ell_t} \) denotes the uncertainty set of row \( \ell_t \) of the TPM, \( Q_t \). The rectangular uncertainty property indicates the choice of the transition probabilities when the system is in state \( \ell_t \in \mathcal{L} \) at epoch \( t \), is independent of the choice of the transition probabilities when the system is in state \( \ell'_t \neq \ell_t \in \mathcal{L} \) at epoch \( t \). This property is key to maintaining the tractability of solving the RMDPs.

We assume the goal of the RMDP-TM is to maximize the expected total discounted worst-case reward. The optimal value function of the RMDP-TM can be written as follows:

\[
v^\text{RMDP-TM}_t(s_t) = \begin{cases} 
\max_{\pi \in \Pi_{t_t \in \Theta_t}} \min_{\theta_t \in \Theta_t} \sum_{k=t}^{T-1} \lambda^{k-t} r_k(s_k, \alpha_k(s_k)) + \lambda^{T-t} r_T(s_T), & \forall s_t \in \mathcal{L} \times \mathcal{M}, \forall t \in T, \\
0, & \text{otherwise,}
\end{cases}
\]  

(4)

where the decision of nature at time epoch \( t \) is the health state TPM, \( Q_t \), the policy of nature from epoch \( t \) onward is a vector of health state TPMs, denoted by \( \theta_t = (Q_t, Q_{t+1}, \ldots, Q_{T-1}) \), and the set \( \Theta_t = \{ (Q_t, Q_{t+1}, \ldots, Q_{T-1}) | Q_k \in \mathcal{Q}_k, \forall k \in \{ t, t+1, \ldots, T-1 \} \} \), represents the set of all admissible policies of nature since epoch \( t \). In this robust context, the optimal value function can be interpreted as the maximum expected total discounted worst-case reward from epoch \( t \) until the patient reaches the absorbing state.

Based on Theorem 1 of (Nilim and El Ghaoui 2005) as well as Theorem 2.2 of (Iyengar 2005), when the uncertainty set, \( \mathcal{Q}_t, \forall t \in T \setminus \{ T \} \), has the rectangular uncertainty property, the optimal
solution of the RMDP-TM model can be obtained by a deterministic Markovian policy, and the optimal value function of the RMDP-TM can be written recursively as follows:

\[
v_{t}^{\text{RMDP-TM}}(s_t) = \begin{cases} 
\max_{\alpha_t(s_t) \in A_t} \left\{ r_t(s_t, \alpha_t(s_t)) + (1 - p_t^E(s_t, \alpha_t(s_t))) \lambda \min_{q_{t,t+1} \in Q_t \times Q_{t+1} \in \mathcal{L}} q_{t,t+1}(\ell_{t+1}) \right\} \times v_{t+1}^{\text{RMDP-TM}}(\ell_{t+1}, m_{t+1}(\alpha_t(s_t))) & \forall s_t = (\ell_t, m_t) \in \mathcal{L} \times \mathcal{M}, t \in \mathcal{T} \setminus \{T\}, \\
r_T(s_T), & s_T = (\ell_T, m_T) \in \mathcal{L} \times \mathcal{M}, \\
0, & \text{otherwise,}
\end{cases}
\]

(5)

where the minimization problem presented in Equation (5) is often referred to as the inner problem:

\[
\sigma_t(s_t, \alpha_t(s_t)) = \min_{q_{t,t+1} \in Q_t \times Q_{t+1} \in \mathcal{L}} q_{t,t+1}(\ell_{t+1}) \sum_{\ell_{t+1} \in \mathcal{L}} q_{t,\ell}(\ell_{t+1}) v_{t+1}^{\text{RMDP-TM}}(\ell_{t+1}, m_{t+1}(\alpha_t(s_t))) \forall s_t \in \mathcal{L} \times \mathcal{M}.
\]

(6)

The RMDP-TM is a time-varying RMDP model since the decisions of nature are allowed to vary across time epochs. The time-invariant counterpart of the RMDP-TM, which requires the decision of nature to be the same for every time epoch, is a more difficult RMDP model due to the correlation of the decisions of nature across the entire time horizon. The optimal value function, of the time-invariant counterpart of the RMDP-TM can be written as follows:

\[
v_{t}^{\text{TL-RMDP-TM}}(s_t) = \begin{cases} 
\max_{\pi \in \Pi_t} \min_{\theta_t} \mathbb{E}_{\theta_t}^{\pi} \left\{ \sum_{k=t}^{T-1} \lambda^{k-t} r_k(s_k, \alpha_k(s_k)) + \lambda^{T-t} r_T(s_T) \right\} & \forall s_t \in \mathcal{L} \times \mathcal{M}, \\
0, & \text{otherwise,}
\end{cases}
\]

(7)

\forall t \in \mathcal{T}, where \( \theta_t^i \equiv (Q, Q, \ldots, Q) \) represents a stationary policy of nature from epoch \( t \) onward. In Section 3 we will provide sufficient conditions under which problems (5) and (7) are equivalent, i.e. when solving (5) generates a stationary policy of nature.

In addition to uncertainty in the health state TPMs, the probability of entering the absorbing state may also be subject to uncertainty. Frequently these transition probabilities are estimated using published statistical risk models. For some diseases there are multiple published risk models, possibly resulting in a range of estimates. Taking this uncertainty into account, the optimality equations (4) become:

\[
v_t'(s_t) = \begin{cases} 
\max_{\pi \in \Pi_t} \min_{(p^F_t, \theta_t) \in (\mathcal{P}_t^F, \theta_t)} \mathbb{E}_{\theta_t}^{(p^F_t, \theta_t)} \left[ \sum_{k=t}^{T-1} \lambda^{k-t} r_k(s_k, \alpha_k(s_k)) + \lambda^{T-t} r_T(s_T) \right] & \forall s_t \in \mathcal{L} \times \mathcal{M}, \\
0, & \text{otherwise,}
\end{cases}
\]
\( \forall t \in \mathcal{T}, \) where
\[
\mathbf{p}^E_t = (p^E_t(s_t, \alpha_t(s_t)), p^E_{t+1}(s_{t+1}, \alpha_{t+1}(s_{t+1})), \ldots, p^E_{T-1}(s_{T-1}, \alpha_{T-1}(s_{T-1}))) \in \mathcal{P}^E_t
\]
is a vector of transition probabilities of entering the absorbing state from epoch \( t \) to \( T - 1 \), and \( \mathcal{P}^E_t \) is a set containing all possible vectors of probabilities of entering the absorbing state from epoch \( t \) to \( T - 1 \). Under the assumption that uncertainty about the ideal choice of risk models for estimating the transition probabilities of entering absorbing state is at the discretion of the decision maker, solving for \( v'_t(s_t) \) is equivalent to solving for \( v^{RMDP-TM}_t(s_t) \) when \( p^E_k(s_k, \alpha_t(s_k)) \) takes the maximum value for all \( k = t, t+1, \ldots, T - 1 \). Thus, in the remainder of this article, we focus on the more challenging RMDP-TM problem in (4).

### 2.3. A Controllable Polyhedral Uncertainty Set

The choice of the uncertainty set \( \mathcal{Q}_t, \forall t \in \mathcal{T} \) plays an important role in determining the computational tractability of solving the RMDP-TM (4), and determining the conservativeness of the corresponding optimal solutions. Some authors have proposed elliptical uncertainty set (Nilim and El Ghaoui 2005, Bertsimas et al. 2011). These provide a good representation of the confidence region of the TPM but they also lead to a nonlinear formulation of the inner problem (6) which can cause the RMDP difficult to solve in practical use.

The controllable polyhedral uncertainty set we propose is based on the interval matrix (IM) model presented in (Nilim and El Ghaoui 2005). The IM model is a type of uncertainty set model with the rectangular uncertainty property. In the IM model, \( \mathcal{Q}^\text{IM}_t \), the uncertainty set of the TPM, is a Cartesian product of the uncertainty set for each row \( \ell_t \) of the TPM, which can be written as
\[
\mathcal{Q}^\text{IM}_t = \prod_{\ell_t \in \mathcal{L}} \mathcal{Q}^\text{IM}_{t, \ell_t}, \quad \text{where} \quad \mathcal{Q}^\text{IM}_{t, \ell_t} \text{ is the uncertainty set for row } \ell_t \text{ of matrix } Q_t \text{ with the following form:}
\]

\[
\mathcal{Q}^\text{IM}_{t, \ell_t} = \left\{ q_{t, \ell_t} \in \mathbb{R}^{||\mathcal{L}||}_{+} : \sum_{\ell_{t+1} \in \mathcal{L}} q_{t, \ell_t}(\ell_{t+1}) = 1, \quad q_{t, \ell_t}(\ell_{t+1}) \leq q^{l}_{t, \ell_t}(\ell_{t+1}) \leq q^{u}_{t, \ell_t}(\ell_{t+1}), \forall \ell_{t+1} \in \mathcal{L} \right\},
\]

\( \forall \ell_t \in \mathcal{L}, \forall t \in \mathcal{T} \setminus \{T\} \). The quantities \( q^{l}_{t, \ell_t}(\ell_{t+1}) \) and \( q^{u}_{t, \ell_t}(\ell_{t+1}) \) denote the lower and upper bounds of \( q_{t, \ell_t}(\ell_{t+1}) \), respectively, and are commonly represented by statistical estimates of the confidence interval.
A common approach to generate the statistical estimates of the confidence interval for transition probabilities based on a longitudinal dataset is to calculate the $100 \times (1 - \alpha/|\mathcal{L}|)\%$ simultaneous confidence intervals for each row of matrix $Q_t$; so that based on the Bonferroni inequality, all $|\mathcal{L}|$ confidence intervals form the $100 \times (1 - \alpha)\%$ confidence region for matrix $Q_t$. For any $\ell_t, \ell_{t+1} \in \mathcal{L}$, the following Equation (8), proposed by (Gold 1963), can be used to calculate the confidence interval for each row of matrix $Q_t$:

$$
\left( \hat{q}_{\ell_t}(\ell_{t+1}) = \left[ \chi^2_{|\mathcal{L}|-1, \alpha/(2|\mathcal{L}|)} \frac{\hat{q}_{\ell_t}(\ell_{t+1})(1-\hat{q}_{\ell_t}(\ell_{t+1}))}{N_{\ell_t}} \right]^{\frac{1}{2}}, \tilde{q}_{\ell_t}(\ell_{t+1}) + \left[ \frac{2^{\frac{1}{2}} \hat{q}_{\ell_t}(\ell_{t+1})(1-\hat{q}_{\ell_t}(\ell_{t+1}))}{N_{\ell_t}} \right]^{\frac{1}{2}} \right)
$$

(8)

where $n_{\ell_t, \ell_{t+1}}$ denotes the observed number of patients who transition from state $\ell_t$ to state $\ell_{t+1}$, $N_{\ell_t} = \sum_{\ell_{t+1} \in \mathcal{L}} n_{\ell_t, \ell_{t+1}}$ denotes the total number of patients in state $\ell_t$, and $\hat{q}_{\ell_t}(\ell_{t+1}) = \frac{n_{\ell_t, \ell_{t+1}}}{N_{\ell_t}}$ is the MLE of the health state transition probability, $q_{\ell_t}(\ell_{t+1})$.

To mitigate the conservativeness of the IM model while maintaining the rectangular uncertainty property, we combine the IM model with an additional uncertainty budget constraint to control the size of the uncertainty set, and therefore control the conservativeness of the robust optimal policy. We refer to our model as the interval model with uncertainty budget (IMUB). For any $\ell_t, \ell_{t+1} \in \mathcal{L}$ such that $\hat{q}_{\ell_t, \ell_t}(\ell_{t+1}) \neq 0$, we define the maximal left/right-hand-side variation associated with probability $q_{\ell_t, \ell_t}(\ell_{t+1})$ as the absolute difference between the lower/upper bound of the transition probability, $q_{\ell_t, \ell_t}(\ell_{t+1})$, and its nominal value, $\hat{q}_{\ell_t, \ell_t}(\ell_{t+1})$; and denote them by $\delta_{\ell_t, \ell_t}^l(\ell_{t+1}) \triangleq \hat{q}_{\ell_t, \ell_t}(\ell_{t+1}) - q_{\ell_t, \ell_t}(\ell_{t+1})$ and $\delta_{\ell_t, \ell_t}^u(\ell_{t+1}) \triangleq q_{\ell_t, \ell_t}(\ell_{t+1}) - \hat{q}_{\ell_t, \ell_t}(\ell_{t+1})$, respectively. In addition, we define the degree of left/right-hand-side variation as the proportion of variation from the lower/upper bound of $q_{\ell_t, \ell_t}(\ell_{t+1})$ to its nominal value, and denote them by $z_{\ell_t, \ell_t}^l(\ell_{t+1})$ and $z_{\ell_t, \ell_t}^u(\ell_{t+1})$, respectively. Therefore, $\forall \ell_t \in \mathcal{L}, t \in \mathcal{T} \setminus \{T\}$, if $q_{\ell_t, \ell_t}(\ell_{t+1}) \leq \hat{q}_{\ell_t, \ell_t}(\ell_{t+1})$, then $z_{\ell_t, \ell_t}^u(\ell_{t+1}) = (q_{\ell_t, \ell_t}(\ell_{t+1}) - \hat{q}_{\ell_t, \ell_t}(\ell_{t+1}))/\delta_{\ell_t, \ell_t}^u(\ell_{t+1})$ and $z_{\ell_t, \ell_t}^l(\ell_{t+1}) = 0$; otherwise, $z_{\ell_t, \ell_t}^u(\ell_{t+1}) = (q_{\ell_t, \ell_t}(\ell_{t+1}) - \hat{q}_{\ell_t, \ell_t}(\ell_{t+1}))/\delta_{\ell_t, \ell_t}^u(\ell_{t+1})$ and $z_{\ell_t, \ell_t}^l(\ell_{t+1}) = 0$. We assume there is no variation when $\hat{q}_{\ell_t, \ell_t}(\ell_{t+1}) = 0$, i.e. $z_{\ell_t, \ell_t}^l(\ell_{t+1})$ and $z_{\ell_t, \ell_t}^u(\ell_{t+1})$ equal 0.

The uncertainty budget on row $\ell_t$ of the TPM, $Q_t$, is denoted by $\Gamma_{\ell_t, \ell_t}$. It defines a limit on the total allowable variations of the probabilities from their nominal values, measured by the sum of
$z^l_{t,\ell_t}(\ell_{t+1})$ and $z^u_{t,\ell_t}(\ell_{t+1})$ for all $\ell_{t+1}$. Thus $\Gamma_{t,\ell_t}$ is limited to the range from 0 to $|L|$. The complete IMUB for row $\ell_t$ of the TPM, $Q_{t}$, can be written as follows:

$$Q^\text{IMUB}_{t,\ell_t}(\Gamma_{t,\ell_t}) = \begin{cases} 
q_{t,\ell_t}(\ell_{t+1}) = \hat{q}_{t,\ell_t}(\ell_{t+1}) - \delta^l_{t,\ell_t}(\ell_{t+1})z^l_{t,\ell_t}(\ell_{t+1}) \\
\quad + \delta^u_{t,\ell_t}(\ell_{t+1})z^u_{t,\ell_t}(\ell_{t+1}), & \forall \ell_{t+1} \in L, \\
\sum_{\ell_{t+1} \in L} q_{t,\ell_t}(\ell_{t+1}) = 1, \\
\sum_{\ell_{t+1} \in L} \left( z^l_{t,\ell_t}(\ell_{t+1}) + z^u_{t,\ell_t}(\ell_{t+1}) \right) \leq \Gamma_{t,\ell_t}, \\
z^l_{t,\ell_t}(\ell_{t+1})z^u_{t,\ell_t}(\ell_{t+1}) = 0, & \forall \ell_{t+1} \in L, \\
0 \leq z^l_{t,\ell_t}(\ell_{t+1}), z^u_{t,\ell_t}(\ell_{t+1}) \leq 1, & \forall \ell_{t+1} \in L, \\
0 \leq q_{t,\ell_t}(\ell_{t+1}) \leq 1, & \forall \ell_{t+1} \in L, \end{cases}$$

(9)

The first equation in (9) represents the variable $q_{t,\ell_t}(\ell_{t+1})$ in terms of its nominal value, and the degree of left-hand-side and right-hand-side variation. Since $q_{t,\ell_t}(\ell_{t+1})$ can only be on one side of its nominal value, $z^l_{t,\ell_t}(\ell_{t+1})$ and $z^u_{t,\ell_t}(\ell_{t+1})$ cannot be positive simultaneously. Therefore, the fourth constraint guarantees that either $z^l_{t,\ell_t}(\ell_{t+1})$ or $z^u_{t,\ell_t}(\ell_{t+1})$ or both are zero. The second and sixth constraints guarantee that the variables, $q_{t,\ell_t}(\ell_{t+1})$, $\forall \ell_{t+1} \in L$, are confined to the probability simplex. The fifth constraint provides the lower and upper bounds for the variables $z^l_{t,\ell_t}(\ell_{t+1})$ and $z^u_{t,\ell_t}(\ell_{t+1})$. The third constraint is the uncertainty budget constraint, which requires the total degree of uncertainty on row $\ell_t$ of the TPM, $Q_t$, to be less than or equal to $\Gamma_{t,\ell_t}$. The IMUB model of the entire TPM, $Q_t$, $\forall t \in T$, can be written as $Q^\text{IMUB}_t(\Gamma_t, \ell_t) = \prod_{\ell_t \in L} Q^\text{IMUB}_{t,\ell_t}(\Gamma_{t,\ell_t}), \forall t \in T$.

**Remark 1:** Notice that when $\Gamma_{t,\ell_t} = 0$, $\forall \ell_t \in L, t \in T$, we have $q_{t,\ell_t}(\ell_{t+1}) = \hat{q}_{t,\ell_t}(\ell_{t+1})$, $\forall \ell_t, \ell_{t+1} \in L, t \in T$, and the RMDP-TM is the MDP model. Likewise, when $\Gamma_{t,\ell_t} = |L|$, $\forall \ell_t \in L, t \in T$, the uncertainty set corresponds to the IM model. Therefore, the uncertainty budget controls the conservativeness of the optimal policy of the RMDP-TM relative to the MDP model.

The IMUB model for TPMs has some similarities to previously proposed uncertainty set models in the robust optimization literature. The budget constraint in the cardinality constrained model proposed by Bertsimas and Sim (2004) requires that the number of coefficients, which can actually change in the interval, be less than or equal to the budget. Similarly, the budget constraint in the
“Lighting does not strike twice” model proposed by Mannor (2012) (Mannor et al. 2012) required that in the TPM of an RMDP model the number of states whose parameters can deviate from their nominal values be bounded by the given budget. Although their models have the capability of controlling the conservativeness of the robust optimal solution, it is hard to interpret the meaning of the budget parameter in practice. In our model on the other hand, the budget parameter is the total degree of variation from the nominal value that is limited by the budget; therefore, intuitively, the budget parameter provides a measurement of how far away the uncertain parameter could be from the nominal value. Our model is most closely related to the D-Norm model proposed by Bertsimas et al. (2011); however, our model has an additional probability constraint, namely, the sum of the probabilities, \( \{q_{t, t_i(l_{t+1})}\}_{t+1 \in \mathcal{L}} \), has to be equal to 1. Moreover, our model is in a sequential dynamic programming framework. As we will show these differences significantly increases the difficulty in solving the RMDP-TM.

3. Analysis and Algorithm for RMDP-TM

In this section, we present theoretical analysis of the RMDP-TM that can be used to establish computationally efficient methods for solving the RMDP-TM. We also provide sufficient conditions under which the time-invariant counterpart of the RMDP-TM can be solved. Complete proofs of Propositions 1-3 are available in Appendix.

In light of the recursive structure in Equation (5), the RMDP-TM can be solved using the robust dynamic programming (RDP) algorithm proposed by Nilim and El Ghaoui (2005). In the RDP algorithm, the expected total discounted worst-case value-to-go is calculated by solving the inner problem (6) at each iteration. Since the inner problem needs to be solved \(|\mathcal{S}| \cdot |\mathcal{T}\{|T\}|\) times, the extra computational cost of solving the RMDP-TM, as compared to solve MDP, depends on how efficiently the inner problem can be solved.

The inner problem (6) with IMUB model is the following nonlinear program:
\[ \min_{t} \sigma_{t}^{\text{IMUB-NLP}}(s_t, \alpha_t(s_t), \Gamma_{t, t}) = \sum_{\ell_{t+1} \in \mathcal{L}} q_{t, \ell} (\ell_{t+1}) v_{t+1}^{\text{RMDP-TM}}(\ell_{t+1}, m_{t+1}(\alpha_t(s_t))) \]

s.t. \[ q_{t, \ell} (\ell_{t+1}) = \hat{q}_{t, \ell} (\ell_{t+1}) - \delta_{t, \ell} (\ell_{t+1}) z_{t, \ell} (\ell_{t+1}) + \delta_{t, \ell} (\ell_{t+1}) z_{t, \ell} (\ell_{t+1}), \forall \ell_{t+1} \in \mathcal{L}, \] (10)

\[ \sum_{\ell_{t+1} \in \mathcal{L}} q_{t, \ell} (\ell_{t+1}) = 1, \] (11)

(IMUB-NLP) \[ \sum_{\ell_{t+1} \in \mathcal{L}} (z_{t, \ell} (\ell_{t+1}) + z_{t, \ell} (\ell_{t+1})) \leq \Gamma_{t, \ell}, \] (12)

\[ z_{t, \ell} (\ell_{t+1}) \cdot z_{t, \ell} (\ell_{t+1}) = 0, \forall \ell_{t+1} \in \mathcal{L}, \] (13)

\[ 0 \leq z_{t, \ell} (\ell_{t+1}), z_{t, \ell} (\ell_{t+1}) \leq 1, \forall \ell_{t+1} \in \mathcal{L}. \] (14)

\[ 0 \leq q_{t, \ell} (\ell_{t+1}) \leq 1, \forall \ell_{t+1} \in \mathcal{L}. \] (15)

The following linear reformulation of the inner problem, called IMUB-LP, is equivalent to IMUB-NLP

\[ \min_{t} \sigma_{t}^{\text{IMUB-LP}}(\ell_t, m_t, \alpha_t(s_t), \Gamma_{t, t}) = \sum_{\ell_{t+1} \in \mathcal{L}} q_{t, \ell} (\ell_{t+1}) v_{t+1}^{\text{RMDP-TM}}(\ell_{t+1}, m_{t+1}(\alpha_t(s_t))) \]

s.t. \[ \sum_{\ell_{t+1} \in \mathcal{L}} q_{t, \ell} (\ell_{t+1}) = 1, \] (16)

(IMUB-LP) \[ \sum_{\ell_{t+1} \in \mathcal{L}} \left[ x_{t, \ell}^{l} (\ell_{t+1}) + x_{t, \ell}^{u} (\ell_{t+1}) \right] \leq \Gamma_{t, \ell}, \] (17)

\[ x_{t, \ell}^{u} (\ell_{t+1}) \geq q_{t, \ell} (\ell_{t+1}) - \frac{\hat{q}_{t, \ell} (\ell_{t+1})}{\delta_{t, \ell} (\ell_{t+1})}, \forall \ell_{t+1} \in \mathcal{L}, \] (18)

\[ x_{t, \ell}^{l} (\ell_{t+1}) \geq \frac{\hat{q}_{t, \ell} (\ell_{t+1}) - q_{t, \ell} (\ell_{t+1})}{\delta_{t, \ell} (\ell_{t+1})}, \forall \ell_{t+1} \in \mathcal{L}, \] (19)

\[ 0 \leq x_{t, \ell}^{l} (\ell_{t+1}), x_{t, \ell}^{u} (\ell_{t+1}) \leq 1, \forall \ell_{t+1} \in \mathcal{L}. \] (20)

\[ 0 \leq q_{t, \ell} (\ell_{t+1}) \leq 1, \forall \ell_{t+1} \in \mathcal{L}, \] (21)

**Proposition 1.** The IMUB-NLP has the same optimal objective function value as the IMUB-LP.
The formal proof of Proposition 1 is given in the Appendix. The basic idea of the proof is to show that for a given optimal solution of IMUB-NLP, a feasible solution to IMUB-LP can be constructed. On the other hand, for a given optimal solution of IMUB-LP, a feasible solution to IMUB-NLP can be constructed. Since the two problems have the same optimal objective function value, the two formulations are equivalent. As we will show in Section 4, the IMUB-LP formulation of the inner problem can significantly reduce the computation time needed to solve the RMDP-TM.

Next, we show that when the uncertainty set corresponds to the IM model, the inner problem (6) can be solved by Algorithm 1 which has the worst-case time complexity of $O(|\mathcal{L}|^2)$.

**Algorithm 1** A fast algorithm to solve the inner problem (6) with the IM model

1. Sort the coefficients of the objective function in nonincreasing order, and relabel the health state according to the order using index $i$

2. For all $i = 1, \ldots, |\mathcal{L}|$, set $y^l_i \leftarrow q^l_{t,\ell}(i)$, $y^u_i \leftarrow q^u_{t,\ell}(i)$

3. **for** $\tau = 1 \rightarrow |\mathcal{L}|$ **do**

4. set $\mathcal{I}_1 \leftarrow \{1, \ldots, \tau - 1\}$ and $\mathcal{I}_2 \leftarrow \{\tau + 1, \ldots, |\mathcal{L}|\}$

5. **if** $1 - \left[ \sum_{j \in \mathcal{I}_1} y^l_j + \sum_{k \in \mathcal{I}_2} y^u_k \right] \in [y^l_{\tau}, y^u_{\tau}]$ **then**

6. $y^*_j \leftarrow y^l_j$, $\forall j \in \mathcal{I}_1$, $y^*_\tau \leftarrow 1 - \left[ \sum_{j=1}^{\tau-1} y^l_j + \sum_{k=\tau+1}^{|\mathcal{L}|} y^u_k \right]$, and $y^*_k \leftarrow y^u_k$, $\forall k \in \mathcal{I}_2$

7. **break**

8. **else**

9. **next** $\tau$

10. **end if**

11. **end for**

**Proposition 2.** Algorithm (1) generates an optimal solution to the inner problem (6) with IM model, and has worst-case time complexity, $O(|\mathcal{L}|^2)$.

The formal proof of Proposition 2 is in the Appendix. The basic idea is to prove that there exists a $\tau$ such that the condition $1 - \left[ \sum_{j \in \mathcal{I}_1} y^l_j + \sum_{k \in \mathcal{I}_2} y^u_k \right] \in [y^l_{\tau}, y^u_{\tau}]$ in the algorithm will be satisfied. Next, we
prove that the primal feasible solution generated by Algorithm 1 is optimal to the inner problem when the uncertainty set is the IM model. This is achieved by constructing a dual solution, and proving it is feasible to the dual problem, and the complementary slackness conditions hold.

Proposition 2 establishes an efficient method for solving an important special case of the RMDP-TM. In Section 4, we show that this results in computation times on the order of those for solving the MDP using the backward induction algorithm (Puterman 1994).

The time-invariant counterpart of the RMDP-TM model is not solvable by the RDP algorithm proposed by Nilim and El Ghaoui (2005) due to the dependency of the nature’s decisions across time epochs. This problem could be solved by using a large semidefinite program proposed by Wiesemann et al. (2013), which does not exploit the recursive structure of the optimal value function; however, this is much more computationally intensive than the RDP algorithm and may not be feasible for large practical problems. Nilim and El Ghaoui (2005) find that the gap between the time-invariant RMDP and the time-varying RMDP goes to zero when the time horizon goes to infinity. For the RMDP-TM, we find sufficient conditions under which the RMDP-TM equals its time-invariant counterpart for a finite horizon. We begin with a definition of nonincreasing worst-case (NIWC) TPM that is relevant to the proposition.

**Definition 1.** The TPM, \( Q_{t}^{NIWC,\Gamma} \) is called an NIWC TPM in the uncertainty set \( Q_{t}^{IMUB}(\Gamma) = \prod_{\ell_{t} \in \mathcal{L}} Q_{t}^{IMUB}(\Gamma) \), if \( \forall \ell_{t} \in \mathcal{L} \), row \( \ell_{t} \) of the TPM, \( Q_{t}^{NIWC,\Gamma} \), denoted by \( q_{t,\ell_{t}}^{NIWC,\Gamma} \), is the optimal solution of the following problem:

\[
\beta_{t}(\ell_{t}) = \min_{q_{t,\ell_{t}} \in Q_{t}^{IMUB}(\Gamma)} \sum_{\ell_{t+1} \in \mathcal{L}} q_{t,\ell_{t}}(\ell_{t+1})c(\ell_{t+1}) \quad \forall \ell_{t} \in \mathcal{T}\setminus\{T\}. \tag{22}
\]

where the coefficients of the objective function, \( \{c(\ell_{t+1})\}_{\ell_{t+1} \in \mathcal{L}} \), are nonincreasing in the health state, \( \ell_{t+1} \).

**Remark 2:** Notice that when \( c(\ell_{t+1}) \) equals \( v_{t+1}^{RMDP-TM}(\ell_{t+1}, m_{t+1}(\alpha_{t}(s_{t})), \forall \ell_{t+1} \in \mathcal{L} \), the problem (22) is equivalent to the inner problem (6).

**Proposition 3.** For the RMDP-TM with optimal value function shown in (5), if the following conditions hold:
(I): The uncertainty set of the TPM, $Q_t$, is $Q_t^{IM}$, an IM model, $\forall t \in T \setminus \{T\}, \mathbf{m}_t \in \mathcal{M}$,

(II): $Q_t^{IM} = Q_{t'}^{IM}, \forall t, t' \in T \setminus \{T\}$,

(III): $r_t(\ell_t, \mathbf{m}_t, \alpha_t(s_t))$ is nonincreasing in $\ell_t$, $\forall \mathbf{m}_t \in \mathcal{M}$, $\alpha_t(s_t) \in \mathcal{A}_t$, and $t \in T \setminus \{T\}$, and $r_T(\ell_T, \mathbf{m}_T)$ is nonincreasing in $\ell_T$, $\forall \mathbf{m}_T \in \mathcal{M}$,

(IV): $p_t^E(\ell_t, \mathbf{m}_t, \alpha_t(s_t))$ is nondecreasing in $\ell_t$, $\forall \mathbf{m}_t \in \mathcal{M}$, $\alpha_t(s_t) \in \mathcal{A}_t$, and $t \in T \setminus \{T\}$, and $p_T^E(\ell_T, \mathbf{m}_T)$ is nondecreasing in $\ell_T$, $\forall \mathbf{m}_T \in \mathcal{M}$,

(V): $Q_{T-1}^{NIWC,|\mathcal{L}|}$ has the increasing failure rate (IFR) property (Barlow and Proschan 1965) then

(a): the optimal value function, $v_t^{RMDP-TM}(\ell_t, \mathbf{m}_t)$, of the RMDP-TM is nonincreasing in $\ell_t$, $\forall \mathbf{m}_t \in \mathcal{M}, t \in T \setminus \{T\}$, and

(b): the optimal policy of nature is $(Q_{T-1}^{NIWC,|\mathcal{L}|}, Q_{T-1}^{NIWC,|\mathcal{L}|}, \ldots, Q_{T-1}^{NIWC,|\mathcal{L}|})$, therefore, it is stationary.

The proof of Proposition 3 is in the Appendix, and is done by induction. These sufficient conditions identify a special case when the time-invariant counterpart of the RMDP-TM can be solved. Conditions (I) and (II) state that the uncertainty sets of the health state TPM for all decision epochs are the same IM model. Condition (III) states that the immediate reward and the terminal reward are nonincreasing as the health state becomes worse for any treatment state. Condition (IV) states that the probability of entering the absorbing state is nondecreasing in health states for any treatment states. These assumptions are reasonable in making medical treatment decisions where a poorer health state is usually associated with a lower reward, and a higher probability of entering the absorbing state. The assumption of the IFR property in Condition (V) has been widely used to analyze the structure of the optimal policy for many applications of MDPs and has been observed to hold empirically in many contexts (Mason et al. 2014, Shechter et al. 2008, Kurt et al. 2011, Alagoz et al. 2004, 2007). Proposition 3 establishes conditions under which the RMDP-TM (4) and its time-invariant counterpart (7) are equivalent. In addition, based on Proposition 3, the computational time for solving a special class of the RMDP-TMs, satisfy the conditions in Proposition 3, can be further reduced to the same as solving an MDP with the TPM set to be the NIWC TPM.
4. Case Study: Optimizing Treatment Decisions for Type 2 Diabetes

In this section, we present a case study of applying the RMDP-TM to optimize the treatment decisions for glycemic control for patients with type 2 diabetes. This case study illustrates the application of the proposed model in the context of type 2 diabetes. It also serves to analyze the performance of the proposed solution methods.

4.1. Background and RMDP-TM Formulation of the Glycemic Control Model

A central focus of managing type 2 diabetes is glycemic control which involves the regulation of blood glucose levels over time. Glycemic control aims to avoid acute daily symptoms of hyperglycemia, to avoid instability in blood glucose over time, and to prevent or delay the development of diabetes-related complications associated with the high blood glucose levels. Glycated hemoglobin (HbA1c) is a commonly used measure of average blood glucose concentration over time. A high HbA1c indicates poor glycemic control, and the need to initiate medication(s) such as one of several oral medications or insulin. HbA1c is obtained via a simple blood test at a recommended frequency of every three months (American Diabetes Association 2013).

Zhang et al. (2014) presented and validated a population-based glycemic control model based on a finite horizon Markov chain with parameters estimated from a retrospective administrative claims dataset with linked laboratory data from a large, nationwide US health plan. This model was used to compare the health benefit, i.e. the quality-adjusted life-years (QALYs) gained, and the total medication costs of different treatment regimens for individuals newly diagnosed with type 2 diabetes.

We used this glycemic control model presented in Zhang et al. (2014) to create an RMDP-TM to optimize the treatment decisions for glycemic control for patients with type 2 diabetes in light of uncertainty in HbA1c transition probability estimates caused by statistical variations. Following is a description of the model in this context.

**Time horizon.** The time horizon starts from the age at diagnosis (55 years old for females and 53 years old for males), and discretized into three-month intervals as recommended by American Diabetes Association (2013) until age 100 years.
**States.** The health states in the model include 10 HbA1c states, $\mathcal{L} = \{\ell(1), \ell(2), \ldots, \ell(10)\}$, defined by a discrete set of clinically relevant ranges of HbA1c levels. As a function of age, the mean HbA1c value for each HbA1c state increases linearly, reflecting the expected deterioration of glycemic control as the patient ages (Sinha et al. 2010). The ranges of HbA1c levels used for categorizing HbA1c states and the mean HbA1c levels for HbA1c states are shown in Appendix Tables 2 and 3. We consider three hyperglycemia-lowering medications: metformin, sulfonylurea, and insulin because they were shown to be the most cost-effective in Zhang et al. (2014). The treatment states are represented as 3-tuple binary vectors in which the first, the second, and the third elements represent the usage status for metformin, sulfonylurea, and insulin, respectively. The absorbing state, $\mathcal{D}$, includes major diabetes-related complications: fatal or non fatal macrovascular events (such as ischemic heart disease, congestive heart failure, and stroke); fatal or non fatal microvascular events (such as blindness, renal failure, and limb amputation); and severe hypoglycemia requiring hospitalization; and death from other causes.

**Actions.** The action is the selection of which medication(s) to initiate at each time epoch. Treatment results in a proportional decrease in HbA1c according to the medication effects estimated from a longitudinal administrative claims dataset including HbA1c records and pharmacy claims (shown in Zhang et al. (2014)). There are no further treatment changes once insulin is initiated, as it is assumed to maintain control of the HbA1c level.

**Rewards.** The immediate reward represents a patient’s QALYs between time epochs. QALYs are widely used for evaluating the health outcome for treatment and health policy decisions (Gold et al. 2002). QALYs adjust a year of life proportionally based on utilities that represent the quality of health that the patient experiences. In this case study, each year is assigned a value between 0 (death) and 1 (perfect health) where the exact value depends on the occurrence of side effects of medication and the disutility of hyperglycemia symptom. For $\forall s_t \in \mathcal{L} \times \mathcal{M}$, the immediate reward is calculated as follows:

$$r_t(s_t, \alpha_t(s_t)) = 0.25 \times (1 - D_{\text{hyper}}(\alpha_t(s_t)))(1 - D_{\text{med}}(\alpha_t(s_t))), \forall \ell_t \in \mathcal{L},$$
where \( D^{\text{hyper}}(\alpha_t(s_t)) \) represents the disutility of hyperglycemia symptom when HbA1c level is above 8\%, \( D^{\text{med}}(\alpha_t(s_t)) \) represents the disutility of taking medications (a quantity measures the side effects of medications) during the 3-month period as shown in Table 1 of Zhang et al. (2014). For \( s_t = D \), \( r_t(s_t, \alpha_t(s_t)) \) is set to be zero. The terminal rewards, \( r_T(s_T) \), are set to be 2.24 years for females and 2.05 years for males based on U.S. life tables (Arias 2011).

**Probabilities.** The probability of entering the absorbing state is calculated as follows:

\[
p^D_t(s_t, \alpha_t(s_t)) = p^0_t + p^{\text{macro}}_t(\ell_t, \alpha_t(s_t)) + p^{\text{micro}}_t(\ell_t, \alpha_t(s_t)) + p^{\text{hypo}}_t(\alpha_t(s_t)).
\]

The probabilities of macro- and micro-vascular events, \( p^{\text{macro}}_t(\ell_t, \alpha_t(s_t)) \) and \( p^{\text{micro}}_t(\ell_t, \alpha_t(s_t)) \) were estimated by using the United Kingdom Prospective Diabetes Study (UKPDS) outcome models in the forms of survival functions (Clarke et al. 2004). The probabilities of severe hypoglycemia, \( p^{\text{hypo}}_t(\alpha_t(s_t)) \), were also obtained from UKPDS Group (1995). The probabilities of death from other causes were obtained from mortality tables from the Centers for Disease Control and Prevention (CDC 2012).

The retrospective administrative claims dataset with linked laboratory data from a large, nationwide US health plan (the same dataset used in Zhang et al. (2014)), was used to estimate parameters related to HbA1c transition probabilities. The population meeting criteria for our study (337 males and 272 females) were 1) aged 40 years or older; 2) have been diagnosed with type 2 diabetes between 1995 and 2010; 3) have received the first non-insulin glucose-lowering medication at least 6 months after enrollment; and 4) have accumulated 15 HbA1c records within 5 years of continuous enrollment, along with complete pharmacy claims records. The nominal HbA1c TPMs were set to be the MLEs of the HbA1c TPMs (shown in Tables 2 and 3), and they were obtained by using the method presented in Denton et al. (2009). Equation (8) was used to calculate left-hand-side and right-hand-side variations of HbA1c transition probabilities for males and females (shown in Appendix Tables 4-7).

The goal of the RMDP-TM version of the glycemic control model is to maximize the worst-case expected total QALYs from the time of diagnosis to the development of the first diabetes-related complication or death. Given the initial HbA1c state distribution, \( \Pi = (\Pi_1, \ldots, \Pi_{10}) \) shown in
Appendix Tables 2 and 3, the QALY results presented in Section 4.3 represent the worst-case expected total QALYs from birth to the first diabetes-related complication or death, and are calculated as $\pi^{RMDP-TM^*} = \text{LYs}_0 + \sum_{i=1}^{10} \Pi_1 v^{RMDP-TM}(\ell(i), (0,0,0))$ where LYs$_0$ represents the expected life years from birth to the time of diagnosis, and $m_1 = (0,0,0)$ represents the initial treatment state for patients not on any diabetes medication at the time of diagnosis.

4.2. Computation Time for RMDP-TM

We present computation times for solving RMDP-TMs in terms of central processing unit (CPU) time. When $\Gamma = 0$, the RMDP-TM is equivalent to the MDP; therefore, we used the backward induction algorithm to solve those instances. When $\Gamma = 1, 2, \ldots, 9$, we used the RDP algorithm where inner problems were solved based on the IMUB-NLP formulation, and the IMUB-LP formulation for comparison. We also used the RDP algorithm to solve the instances of RMDP-TMs when $\Gamma = 10$, the inner problems were solved based on the IMUB-NLP formulation, the IMUB-LP formulation, and the fast algorithm 1 for comparison.

Table 1 shows the computation results for $\Gamma = 1, \ldots, 10$. There are large differences in CPU times when applying different formulations to solve the inner problem. The CPU times for solving RMDP-TMs with the IMUB-LP formulation is an order of magnitude lower than those for solving RMDP-TMs with the IMUB-NLP formulation. In addition, when applying Algorithm 1 to solve the instances of RMDP-TMs with $\Gamma = 10$, the CPU times (4 seconds for both genders) are similar to the CPU times for solving the MDP using the backward induction algorithm. These differences in CPU times demonstrate the importance of Propositions 1 and 2.

4.3. Treatment Policy Performance Comparison

To compare the performance of various treatment policies, we define the *nominal performance* of a policy to be the expected total discounted reward under the nominal TPMs; and the *worst-case performance* of a policy to be the expected total discounted reward under the worst-case criterion when the TPM can vary over its uncertainty set.
Table 1  Computation time (in seconds) for solving RMDP-TMs with the uncertainty budget varies from 0 to 10. Dashes indicate cases that are not applicable to the given solution method.

<table>
<thead>
<tr>
<th>Solution methods</th>
<th>0 (= MDP)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Males</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Backward induction</td>
<td>2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>RDP algorithm+IMUB-NLP</td>
<td>-</td>
<td>10417</td>
<td>10639</td>
<td>10848</td>
<td>10444</td>
<td>10104</td>
<td>10959</td>
<td>10915</td>
<td>10859</td>
<td>10209</td>
<td>10891</td>
</tr>
<tr>
<td>RDP algorithm+IMUB-LP</td>
<td>-</td>
<td>1141</td>
<td>1185</td>
<td>1127</td>
<td>1124</td>
<td>1185</td>
<td>1170</td>
<td>1154</td>
<td>1143</td>
<td>1169</td>
<td>1129</td>
</tr>
<tr>
<td>RDP algorithm+Algorithm 1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>4</td>
</tr>
<tr>
<td><strong>Females</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Backward induction</td>
<td>2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>RDP algorithm+IMUB-NLP</td>
<td>-</td>
<td>10613</td>
<td>10970</td>
<td>10196</td>
<td>10668</td>
<td>10643</td>
<td>10184</td>
<td>10857</td>
<td>10089</td>
<td>10342</td>
<td>10068</td>
</tr>
<tr>
<td>RDP algorithm+IMUB-LP</td>
<td>-</td>
<td>1291</td>
<td>1273</td>
<td>1278</td>
<td>1248</td>
<td>1293</td>
<td>1269</td>
<td>1253</td>
<td>1243</td>
<td>1212</td>
<td>1260</td>
</tr>
<tr>
<td>RDP algorithm+Algorithm 1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>4</td>
</tr>
</tbody>
</table>

Figure 1 shows the nominal performance versus the worst-case performance for no treatment, the treatment guideline (Nathan et al. 2009), and model-based optimal policies including the MDP-optimal policy (i.e. the RMDP-optimal policy with $\Gamma = 0$), and the RMDP-TM optimal policies with $\Gamma > 0$. The treatment guideline refers to the process of initiating metformin as the first-line medication when the patient’s HbA1c is above 7%, and adding sulfonylurea when the HbA1c is above 7% again, and finally initiating insulin (in place of sulfonylurea) once the patient fails to maintain the goal of $\text{HbA1c} \leq 7\%$ with metformin and sulfonylurea together. This treatment guideline was shown to be the most cost-effective (Zhang et al. 2014), therefore, we include it for comparison.

When comparing the nominal performance, all model-based polices outperform the treatment guideline (61.93 QALYs for male and 64.89 QALYs for female). When comparing the worst-case performance, RMDP-TM optimal policies with $\Gamma > 0$ outperform the treatment guideline (61.69 QALYs for male and 64.60 QALYs for female).
Figure 1  The nominal performance versus the worst-case performance based on quality-adjusted life-years to
the first major complication for no treatment (hollow circle), the treatment guideline (solid circle),
the MDP-optimal policy (i.e. the RMDP-optimal policy with Γ = 0) (solid triangle), and RMDP-TM
optimal policies with Γ > 0 (solid diamond).

Among all model-based optimal policies, the MDP-optimal policy (i.e. the RMDP-optimal policy
with Γ = 0) results in the highest nominal performance, the expected QALYs is 62.53 for males
and 65.74 for females. However, its corresponding worst-case performance is the lowest: 61.70
QALYs for males (a reduction of 0.82 QALYs from its nominal performance) and 64.47 QALYs
for females (a reduction of 1.27 QALYs from its nominal performance). To put this in perspective,
these reductions are more than an order of magnitude greater than the use of aspirin for secondary prevention of myocardial infarction in 45-year-old men, an important intervention, which has been estimated to provide a QALY gain of 0.04 per patient (Pignone et al. 2006). On the other hand, the RMDP-optimal policy with \( \Gamma = 10 \) (i.e. the RMDP-optimal policy based on the fixed uncertainty set (i.e. the IM model) of the TPM) is very conservative, which results in the lowest nominal performance: 62.44 QALYs for males (a reduction of 0.09 QALYs from the nominal performance of the MDP-optimal policy) and 65.63 QALYs for females (a reduction of 0.11 QALYs from the nominal performance of the MDP-optimal policy).

Figure 1 shows that the worst-case performance of the MDP-optimal policy can be significantly improved by the alternative RMDP-optimal policies. For males, the worst-case performance of the RMDP-optimal policy with \( \Gamma = 2 \) is 61.82 QALYs, higher than that of the MDP-optimal policy by 0.12 QALYs while sacrificing only 0.01 QALYs in the nominal performance. For females, the worst-case performance of the RMDP-optimal policy with \( \Gamma = 1 \) is 64.64 QALYs, higher than that of the MDP-optimal policy by 0.17 QALYs while sacrificing only 0.01 QALYs in the nominal performance. This example illustrates how the RMDP-TM can be used to understand the trade-off between nominal and worst-case performance, thus allowing the decision maker to limit how conservative the policy is by selecting \( \Gamma \) accordingly.

4.4. Analysis of the time-invariant counterpart of the RMDP-TM for the glycemic control problem

We found that the MLEs of the TPMs for both genders did not satisfy the IFR assumption exactly. Based on the following worst-case violation measurement (Alagon et al. 2004):

\[
\epsilon = \max_{\ell \in \mathcal{T}} \max_{j \in \{1, \ldots, |\mathcal{L}| - 1\}} \max_{i \in \{1, \ldots, |\mathcal{L}|\}} \sum_{s=1}^{|\mathcal{L}|} \left[ q_{t, i,j}(\ell_{t+1}(s)) - q_{t, i/(j+1)}(\ell_{t+1}(s)) \right] \tag{23}
\]

The worst-case violation is 0.125 for females and 0.1191 for males. We also calculated the NIWC TPM for males and females, and found that NIWC TPMs for both genders did not satisfy the IFR assumption exactly either. The violation is 0.0898 for female patients and 0.1129 for male patients.
Therefore, the condition (V) in Proposition 3 did not hold. We observed that for $\Gamma \geq 0$, the optimal policy of nature was not stationary, therefore, this glycemic control version of the RMDP-TM cannot provide the optimal solution to its time-invariant counterpart. Nevertheless, the property would hold for other applications that strictly meet the IFR condition, and thus proposition 3 identifies a special class of treatment problems for which the computationally intractable time-invariant form of the RMDP can be solved easily.

5. Conclusions

We presented a robust stochastic dynamic programming model, the RMDP-TM, that can fit a broad range of medical treatment decisions in which there is uncertainty in transition probabilities. The RMDP-TM, makes a novel connection between stochastic dynamic programming and the robust optimization literature by incorporating an uncertainty budget-based formulation in the MDP setting. In addition, we proposed a new uncertainty model for the TPMs called the IMUB model which has the capability of controlling the size of the uncertainty set in order to conservativeness of the optimal solution.

We proved that the proposed RMDP-TM is computationally tractable with the proposed uncertainty set model. We also presented some theoretical properties that can be exploited to achieve computational efficiency. In addition, we presented reasonable conditions under which an optimal solution to the time-invariant counterpart of the RMDP-TM, in which the policy of nature is required to be stationary, can be found easily by solving the RMDP-TM. Our numerical experiments showed that the RMDP-TM can be solved efficiently with the proposed solution methods. In addition, we provided and proved sufficient conditions under which a special type of RMDP model, namely, the time-invariant RMDP, can be solved.

We applied the proposed models to optimize the treatment decisions for a glycemic control problem in the context of type 2 diabetes. Based on the numerical results, we found that the MDP-optimal policy was associated with the lowest worst-case performance among all other RMDP-optimal policies, and the RMDP-optimal policy based on the fixed uncertainty set (i.e. the IM
model) of the TPM resulted in very conservative nominal performance. We showed that the RMDP-TM could provide alternative policies that were less conservative and can significantly improve the worst-case performance with only moderate loss in the nominal performance. The magnitude of the improvement is much higher than some well-known disease prevention programs such as influenza vaccination, and the use of aspirin for secondary prevention of myocardial infarction.

Acknowledgments
This material is based upon work supported by the National Science Foundation under Grant Number CMMI-1462060 (Denton). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. The authors would like to thank Nilay Shah and Steven Smith from Mayo Clinic for valuable feedback from clinical perspective. The authors also thank Murat Kurt for helpful comments on an early version of this manuscript.

References


Appendix

Proof of Proposition 1 First, we prove that if \((z_{t,\ell_t}^l, x_{t,\ell_t}^u)\) is an optimal solution to the (IMUB-NLP), then we can construct a feasible solution, \((x_{t,\ell_t}^l, x_{t,\ell_t}^u, q_{t,\ell_t}^l, q_{t,\ell_t}^u)\), to the (IMUB-LP) by setting \(x_{t,\ell_t}^l = z_{t,\ell_t}^l\) and \(x_{t,\ell_t}^u = z_{t,\ell_t}^u\). Based on constraints (10) and (13) in the (IMUB-NLP), we have the following relationship. If \(q_{t,\ell_t}^l \leq \hat{q}_{t,\ell_t}\) then

\[
x_{t,\ell_t}^l = z_{t,\ell_t}^l = \frac{q_{t,\ell_t}(\ell_{t+1}) - \hat{q}_{t,\ell_t}(\ell_{t+1})}{\delta_{t,\ell_t}(\ell_{t+1})} \geq 0,
\]

and

\[
x_{t,\ell_t}^u = z_{t,\ell_t}^u = 0.
\]

If \(q_{t,\ell_t}^l > \hat{q}_{t,\ell_t}\) then

\[
x_{t,\ell_t}^l = \frac{\hat{q}_{t,\ell_t}(\ell_{t+1}) - q_{t,\ell_t}(\ell_{t+1})}{\delta_{t,\ell_t}(\ell_{t+1})} \geq 0,
\]

and

\[
x_{t,\ell_t}^u = z_{t,\ell_t}^u = 0.
\]

It follows that (24, 25) imply constraints (18) and (19) are satisfied. Satisfaction of constraints (16), (17), and (20) is implied by the constraints (11), (12), and (14) in the (IMUB-NLP), respectively.

Therefore, \((x_{t,\ell_t}^l, x_{t,\ell_t}^u, q_{t,\ell_t}^l, q_{t,\ell_t}^u)\) is a feasible solution to the (IMUB-LP). Since the (IMUB-NLP) and (IMUB-LP) have the same objective function, it follows that:

\[
\sigma_t^{IMUB-NLP,*}(s_t, \alpha_t(s_t), \Gamma_t, \ell_t) \geq \sigma_t^{IMUB-LP,*}(s_t, \alpha_t(s_t), \Gamma_t, \ell_t)
\]

Next we prove that if \((x_{t,\ell_t}^l, x_{t,\ell_t}^u, q_{t,\ell_t}^l, q_{t,\ell_t}^u)\) is the optimal solution to the (IMUB-LP), then we can construct a feasible solution \((z_{t,\ell_t}^l, z_{t,\ell_t}^u, q_{t,\ell_t}^l, q_{t,\ell_t}^u)\) to the (IMUB-NLP). For any \(\ell_{t+1} \in \mathcal{L}\), we define \(z_{t,\ell_t}^l, z_{t,\ell_t}^u, q_{t,\ell_t}^l, q_{t,\ell_t}^u\) as follows:

\[
q_{t,\ell_t}^l(\ell_{t+1}) = q_{t,\ell_t}^l(\ell_{t+1}),
\]

if \(q_{t,\ell_t}^l(\ell_{t+1}) \leq \hat{q}_{t,\ell_t}(\ell_{t+1})\), then

\[
z_{t,\ell_t}^l(\ell_{t+1}) = \frac{q_{t,\ell_t}(\ell_{t+1}) - \hat{q}_{t,\ell_t}(\ell_{t+1})}{\delta_{t,\ell_t}(\ell_{t+1})},
\]

and if \(q_{t,\ell_t}^l(\ell_{t+1}) > \hat{q}_{t,\ell_t}(\ell_{t+1})\),

\[
z_{t,\ell_t}^u(\ell_{t+1}) = \frac{\hat{q}_{t,\ell_t}(\ell_{t+1}) - q_{t,\ell_t}(\ell_{t+1})}{\delta_{t,\ell_t}(\ell_{t+1})}.
\]

Therefore, constraints (10) and (13) are satisfied. Based on the constraint (16) in the (IMUB-LP), we have \(\sum_{\ell_{t+1} \in \mathcal{L}} q_{t,\ell_t}^l(\ell_{t+1}) = 1\). Therefore, \(\sum_{\ell_{t+1} \in \mathcal{L}} q_{t,\ell_t}^u(\ell_{t+1}) = 1\), namely, the constraint (11) is satisfied. Based on constraints (18) and (19) in the (IMUB-LP), we know that

\[
x_{t,\ell_t}^l(\ell_{t+1}) \geq \frac{q_{t,\ell_t}(\ell_{t+1}) - \hat{q}_{t,\ell_t}(\ell_{t+1})}{\delta_{t,\ell_t}(\ell_{t+1})} = z_{t,\ell_t}^l(\ell_{t+1}),
\]

and

\[
x_{t,\ell_t}^u(\ell_{t+1}) \geq \frac{\hat{q}_{t,\ell_t}(\ell_{t+1}) - q_{t,\ell_t}(\ell_{t+1})}{\delta_{t,\ell_t}(\ell_{t+1})} = z_{t,\ell_t}^u(\ell_{t+1}).
\]
Therefore,
\[
\sum_{\ell_{t+1}} [z_{l,t}^l(\ell_{t+1}) + z_{l,t}^u(\ell_{t+1})] \leq \sum_{\ell_{t+1}} [x_{l,t}^l(\ell_{t+1}) + x_{l,t}^u(\ell_{t+1})] \leq \Gamma_{t,t'}, \tag{29}
\]
and
\[
z_{l,t}^l(\ell_{t+1}) \leq x_{l,t}^l(\ell_{t+1}) \leq 1, \tag{30}
\]
and
\[
z_{u,t}^l(\ell_{t+1}) \leq x_{u,t}^l(\ell_{t+1}) \leq 1. \tag{31}
\]
Inequalities (29), (30), and (31) imply constraints (12) and (14) in the (IMUB-NLP) are satisfied. Thus, \((z_{l,t}^l(\ell_{t+1}), z_{u,t}^l(\ell_{t+1}), \Gamma_{t,t'})\) is a feasible solution to the (IMUB-NLP). Since the (IMUB-NLP) and (IMUB-LP) have the same objective function, it follows that:
\[
\sigma_{t}^{IMUB-NLP,*}(s_t, \alpha_t(s_t), \Gamma_{t,t'}) \leq \sigma_{t}^{IMUB-LP,*}(s_t, \alpha_t(s_t), \Gamma_{t,t'}), \tag{32}
\]
Together (26) and (32) imply
\[
\sigma_{t}^{IMUB-NLP,*}(s_t, \alpha_t(s_t), \Gamma_{t,t'}) = \sigma_{t}^{IMUB-LP,*}(s_t, \alpha_t(s_t), \Gamma_{t,t'}),
\]
and the proof is complete. \(\blacksquare\)

**Proof of Proposition 2:** First, we rewrite the inner problem (6) in the following general form:

\[
\min \quad \psi(c, y^l, y^u) = \sum_{i=1}^{n} c_i y_i \tag{33}
\]
\[
\text{s.t.} \quad y_i \geq y^l_i, \quad y_i \leq y^u_i, \quad \sum_{i=1}^{n} y_i = 1, \quad y_i \geq 0, \quad i = 1, 2, \ldots, n. \tag{34, 35, 36, 37}
\]

(LP.1)

where \(y_i, i = 1, 2, \ldots, n\), are the decision variables, \(c = (c_1, c_2, \ldots, c_n)\), where \(c_i \geq 0, i = 1, 2, \ldots, n\), is the vector of objective function coefficients, \(y^l = (y^l_1, y^l_2, \ldots, y^l_n)\) and \(y^u = (y^u_1, y^u_2, \ldots, y^u_n)\) are vectors of lower and upper bounds of the decision variables, \(y_i, i = 1, 2, \ldots, n\), respectively. We assume that \(0 \leq y^l_i < y^u_i\) without loss of generality.

The assumption that (LP.1) has at least one feasible solution implies that the feasible region of (LP.1) is nonempty. It is also bounded since \(0 \leq y_i \leq 1, \forall i = 1, \ldots, n\). Since the coefficients of the objective function are nonnegative, the value of the objective function is also bounded. Therefore, based on Corollary 2.3 in Bertsimas (Bertsimas and Tsitsiklis 1997), we know that there exists at least one basic feasible solution which is an optimal solution to (LP.1).

Now we analyze the structure of the optimal solution to (LP.1). The standard form of (LP.1) can be written as follows:
\(\min \sum_{i=1}^{n} c_i y_i \) \hspace{1cm} (38)

\[ y_i - a_i = y^l_i, \quad i = 1, 2, \ldots, n, \] \hspace{1cm} (39)

\[ y_i + b_i = y^u_i, \quad i = 1, 2, \ldots, n, \] \hspace{1cm} (40)

\[ \sum_{i=1}^{n} y_i = 1, \] \hspace{1cm} (41)

\[ y_i, a_i, b_i \geq 0, \quad i = 1, 2, \ldots, n. \] \hspace{1cm} (42)

This formulation includes \(3n\) decision variables and \(2n + 1\) constraints. Thus, a basic feasible solution has \(n - 1\) non-basic variables. For any \(i \in \{1, \ldots, n\}\), consider the following eight cases for a basic feasible solution \((y'_1, \ldots, y'_n, a'_1, \ldots, a'_n, b'_1, \ldots, b'_n)\):

1: \(y_i, a_i, b_i\) are all basic variables. In this case, \(y'_i\) is in the range \([y^l_i, y^u_i]\).

2: \(a_i\) is a non-basic variable, and \(y_i, b_i\) are basic variables. In this case, \(y'_i = y^l_i\).

3: \(b_i\) is a non-basic variable, and \(y_i, a_i\) are basic variables. In this case, \(y'_i = y^u_i\).

4: \(y_i\) is a non-basic variable, and \(a_i, b_i\) are basic variables. This is possible if and only if \(y^l_i = 0\). In this case, \(y^*_i\) is at its lower bound 0.

5: \(a_i, y_i\) are both non-basic variables, and \(b_i\) is a basic variable. In this case, the basis matrix includes a row of zeros and this case does not correspond to a basic feasible solution.

6: \(a_i, b_i\) are both non-basic variables, and \(y_i\) is a basic variable. This case corresponds to an infeasible solution since \(y_i\) cannot be simultaneously equal to \(y^l_i\) and \(y^u_i\) by assumption.

7: \(y_i, b_i\) are both non-basic variables, \(a_i\) is basic variable. This case corresponds to an infeasible solution since the constraint \(y_i + b_i = y^u_i > 0\) will be violated.

8: \(y_i, a_i, b_i\) are all non-basic variables. This is infeasible since the constraint \(y_i + b_i = y^u_i > 0\) will be violated.

Therefore, cases (1–4) are the only possible cases. For any \(i \in \{1, \ldots, n\}\), among the three decision variables \(y_i, a_i,\) and \(b_i\), at most one can be a non-basic variable. Since there are \(n - 1\) non-basic variables, there is at most one \(k\) that is of case 1, and all others are of case 2, 3, or 4. Therefore, there exists one optimal solution of (LP.1), which is a basic feasible solution, is of the following structure: at most one \(y_i \in (y^l_i, y^u_i)\), and all others are either at the lower bound or upper bound.

For basic feasible solution \((y'_1, y'_2, \ldots, y'_n)\), assume variable \(y'_r\) corresponding to the case 1. If there is a variable \(y'_k\) such that \(y'_k = y^l_k\) and \(c_k < c_r\), then we can pivot to a better basic feasible solution via the following pivot scheme A.

1: If \(y'_r - y'_k \leq y^u_k - y^l_k\), then

\[ y'_r \leftarrow y'_r, \quad y'_k \leftarrow y'_k + (y'_r - y^l_r), \quad \tau \leftarrow k. \]
2: If \( y'_{\tau} - y'_{k} > y_{l}^{u} - y_{l}^{k} \), then
\[
y'_{\tau} \leftarrow y'_{\tau} - (y_{u}^{u} - y_{u}^{k}), \quad y'_{k} \leftarrow y_{u}^{u}.
\]

On the other hand, if there is a \( k \) such that \( y'_{k} = y_{u}^{u} \) and \( c_{k} > c_{\tau} \), then we can pivot to a better basic feasible solution via the following pivot scheme B.

1: If \( y_{u}^{u} - y'_{\tau} \leq y_{u}^{u} - y_{l}^{k} \), then
\[
y'_{\tau} \leftarrow y_{u}^{u}, \quad y'_{k} \leftarrow y_{u}^{u} - (y_{u}^{u} - y'_{\tau}), \quad \tau \leftarrow k.
\]

2: If \( y_{u}^{u} - y'_{\tau} > y_{u}^{u} - y_{l}^{k} \), then
\[
y'_{\tau} \leftarrow y'_{\tau} + (y_{u}^{u} - y_{l}^{k}), \quad y'_{k} \leftarrow y_{l}^{k}.
\]

Therefore, there exists a basic feasible solution with basic variable \( y_{\tau} \) such that
\[
y_{\tau} = 1 - \sum_{j \in \mathcal{I}_1} y_{j}^{u} - \sum_{k \in \mathcal{I}_2} y_{k}^{u} \in [y'_{\tau}, y_{u}^{u}].
\] (43)

where \( \mathcal{I}_1 = \{ i : c_{i} > c_{\tau} \} \), and \( \mathcal{I}_2 = \{ i : c_{i} \leq c_{\tau} \} \). Thus the existence of the index \( \tau \) in the Algorithm 1 is guaranteed.

Next, we prove that the solution, \( (y^{*}_{1}, \cdots, y^{*}_{n}) \), generated by Algorithm 1 is an optimal solution of (LP.1). The feasibility of \( (y^{*}_{1}, \cdots, y^{*}_{n}) \) is guaranteed by the Algorithm 1. To prove that is optimal, we first consider the dual problem of (LP.1):

\[
\begin{align*}
\text{max} & \quad \sum_{i=1}^{n} (\mu_{i} y_{i}^{l} - \rho_{i} y_{i}^{u}) + w \\
\text{s.t.} & \quad \mu_{i} - \rho_{i} + w \leq c_{i}, i = 1, 2, \cdots, n, \quad \text{(45)} \\
& \quad \mu_{i}, \rho_{i} \geq 0 i = 1, 2, \cdots, n, \quad \text{(46)} \\
& \quad w \quad \text{u.r.s.} \quad \text{(47)}
\end{align*}
\]

(Dual of LP.1)

Based on the primal feasible solution, \( (y^{*}_{1}, \cdots, y^{*}_{n}) \), we can construct a dual solution as follows:

\[
\begin{align*}
w^{*} &= c_{\tau}, \\
\mu_{i}^{*} &= c_{i} - c_{\tau}, & \text{if } i \in \mathcal{I}_1, \\
\rho_{i}^{*} &= 0, \\
\mu_{i}^{*} &= 0, \\
\rho_{i}^{*} &= 0, \\
\mu_{\tau}^{*} &= 0, \\
\rho_{\tau}^{*} &= c_{\tau} - c_{i}, & \text{if } i \in \mathcal{I}_2.
\end{align*}
\] (48)

The dual solution satisfies the constraints (45), and the nonnegativity constraints for \( \mu_{i} \) and \( \rho_{i} \) for all \( i \) are guaranteed by the fact that \( c_{i} \) are in nonincreasing order in \( i \). Therefore, \( (\mu^{*}_{1}, \cdots, \mu^{*}_{n}, \rho^{*}_{1}, \cdots, \rho^{*}_{n}) \),
... $\rho_n^*$, $w^*$) is a feasible solution of (Dual of LP.1). For the pair of primal and dual problems shown in (LP.1) and (Dual of LP.1), the complementary slackness condition can be written as follows:

$$ (c_i - \mu_i - \rho_i - w)y_i = 0, \quad \forall i = 1, \ldots, n \quad (49) $$
$$ \mu_i(y_i - y_i^*) = 0, \quad \forall i = 1, \ldots, n \quad (50) $$
$$ \rho_i(y_i^* - y_i) = 0, \quad \forall i = 1, \ldots, n \quad (51) $$

For any $i \in \{1, 2, \ldots, n\}$, plugging (48) into the left-hand-size of equation (49), condition (49) satisfies. For any $i \in \mathcal{I}_1$, $y_i^* = y_j^*$ and therefore $\mu_i^*(y_i^* - y_j^*) = 0$. For any $i \in \mathcal{I}_2 \cap \{\tau\}$, $\mu_i^* = 0$ and therefore $\mu_i^*(y_i - y_i^*) = 0$. Thus, the condition (50) holds. For any $i \in \mathcal{I}_1 \cap \{\tau\}$, $\rho_i^* = 0$ and therefore $\rho_i^*(y_i^* - y_i^*) = 0$. For any $i \in \mathcal{I}_2$, $y_i^* = y_i^*$ and therefore $\rho_i^*(y_i^* - y_i^*) = 0$. Thus, the condition (51) holds. Since $(\mu_1^*, \ldots, \mu_n^*, \rho_1^*, \ldots, \rho_n^*, w^*)$ and $(y_1^*, \ldots, y_n^*)$ satisfy the complementary slackness conditions (49–51), $(y_1^*, \ldots, y_n^*)$ is an optimal solution to (LP.1).

The complexity of the Algorithm 1 depends on the step 1 (sorting), therefore, it has the worst-case time complexity of $O(|\mathcal{L}|^2)$. □

The following Lemma 1 will be used to prove the Proposition 3. Lemma 1 is similar to the Lemma 4.7.2 in Puterman (1994), however, Lemma 4.7.2 in Puterman (1994) is for the infinite-horizon case, and for nondecreasing sequence $\{v_j\}_{j=0,1,\ldots}$.

**Lemma 1.** Let $\{x_j\}, \{x_j'\}$ be real-value non-negative sequences satisfying

$$ \sum_{j=0}^{N} x_j \geq \sum_{j=0}^{N} x_j' $$

for all $n$, with equality holding for $n = 0$. Suppose $v_{j+1} \leq v_j$ for $j = 0, 1, \ldots, N$, then

$$ \sum_{j=0}^{N} v_j x_j \leq \sum_{j=0}^{N} v_j x_j' $$

**Proof of Lemma 1** Let $n$ be arbitrary and $v_{-1} = 0$. Then

$$ \sum_{j=0}^{N} v_j x_j \leq \sum_{j=0}^{N} x_j \sum_{j=0}^{N} (v_j - v_{j-1}) = \sum_{j=0}^{N} (v_j - v_{j-1}) \sum_{j=0}^{N} x_j + v_0 \sum_{j=0}^{N} x_j $$

$$ \leq \sum_{j=0}^{N} (v_j - v_{j-1}) \sum_{i=j}^{N} x_i' + v_0 \sum_{i=0}^{N} x_i \quad (\because v_j - v_{j-1} \leq 0, \forall j = 1, \ldots, N) $$

$$ = \sum_{j=0}^{N} (v_j - v_{j-1}) \sum_{i=j}^{N} x_i' + v_0 \sum_{i=0}^{N} x_i' \quad (\because \sum_{j=0}^{N} x_j = \sum_{j=0}^{N} x_j') $$

$$ = \sum_{j=0}^{N} (v_j - v_{j-1}) \sum_{i=j}^{N} x_i' = \sum_{j=0}^{N} x_j' \sum_{i=0}^{j} (v_i - v_{i-1}) = \sum_{j=0}^{N} v_j x_j' $$

Therefore, $\sum_{j=0}^{N} v_j x_j \leq \sum_{j=0}^{N} v_j x_j'$ holds.
**Proof of Proposition 3:** First, we show that given Conditions (I) and (II), the NIWC TPMs, \( \{ Q_{t}^{\text{NIWC},|\mathcal{L}|} \}_{t \in \mathcal{T} \setminus \{T\}} \), defined by Definition 1, are the same for all time epochs. By Condition (I) and Proposition 2, each row of the NIWC TPM, \( Q_{t}^{\text{NIWC},|\mathcal{L}|} \), can be generated by using Algorithm 1. The order of the coefficients in the objective function of the problem (22) is fixed (i.e. nonincreasing in health states), therefore, as shown in Algorithm 1, the optimal solution, \( q_{t,\ell}^{\text{NIWC},|\mathcal{L}|}, \forall \ell \in \mathcal{L} \), is only determined by the feasible region of the problem (22) (i.e, the uncertainty set, \( Q_{t}^{\text{IM}} \)). By Condition (II) we know \( Q_{t}^{\text{IM}} = Q_{t'}^{\text{IM}}, \forall \ell \in \mathcal{L}, t, t' \in \mathcal{T} \setminus \{T\} \). Therefore, \( q_{t,\ell}^{\text{NIWC},|\mathcal{L}|} = q_{t',\ell}^{\text{NIWC},|\mathcal{L}|}, \forall t, t' \in \mathcal{T} \setminus \{T\}, \ell \in \mathcal{L} \).

It follows that

\[
Q_{1}^{\text{NIWC},|\mathcal{L}|} = Q_{2}^{\text{NIWC},|\mathcal{L}|} = \cdots = Q_{T-1}^{\text{NIWC},|\mathcal{L}|}
\]  

(52)

Next, we prove results (a) and (b) hold by induction. For the base case \( t = T \), \( v_{T}^{\text{RMDP-TM}}(\ell_{T}, m_{T}) = r_{T}(\ell_{T}, m_{T}) \) is nonincreasing in \( \ell_{T}, \forall m_{T} \in \mathcal{M} \) by Condition (III). Now assume that \( v_{k}^{\text{RMDP-TM}}(\ell_{k}, m_{k}) \) is nonincreasing in \( \ell_{k}, \forall m_{k} \in \mathcal{M} \) and \( k \in \{t + 1, t + 2, \ldots, T - 1\} \), then we need to prove \( v_{t}^{\text{RMDP-TM}}(\ell_{t}, m_{t}) \) is nonincreasing in \( \ell_{t}, \forall m_{t} \in \mathcal{M} \). Given the base case and induction hypothesis, the optimal decision of nature for \( t \in \{t, t+1, \ldots, T-1\} \) is the NIWC TPM, \( Q_{t}^{\text{NIWC},|\mathcal{L}|} \), of the uncertainty set \( Q_{t} \) based on Definition 1. Combined Condition (V) and Equation (52) we know that \( Q_{t}^{\text{NIWC},|\mathcal{L}|} \) has the IFR property. By Lemma 1, we have

\[
\sum_{\ell_{t+1} \in \mathcal{L}} q_{t,\ell_{t}}^{\text{NIWC}}(\ell_{t+1})v_{t+1}^{\text{RMDP-TM}}(\ell_{t+1}, m_{t+1}) \geq \sum_{\ell_{t+1} \in \mathcal{L}} q_{t,\ell_{t}}^{\text{NIWC}}(\ell_{t+1})v_{t+1}^{\text{RMDP-TM}}(\ell_{t+1}, m_{t+1}),
\]

(53)

if the health state \( \ell_{t} \in \mathcal{L} \) is order such that a patient in health state \( \ell_{t} \) is healthier than a patient in health state \( \ell'_{t} \). Combined with the conditions that the one-period reward at decision epoch \( t \) is nonincreasing in \( \ell_{t} \) for any \( m_{t} \in \mathcal{M} \) (Condition (III)), and the probability of entering absorbing state is nondecreasing in \( \ell_{t} \) for any \( m_{t} \in \mathcal{M}, \alpha_{t} \in A_{t}(\ell_{t}, m_{t}) \) (Condition (IV)), we have \( v_{t}^{\text{RMDP-TM}}(\ell_{t}, m_{t}) \) is nonincreasing in \( \ell_{t} \) for any \( m_{t} \in \mathcal{M} \). Therefore, the induction hypothesis is satisfied, and the result (a) follows.

Based on result (a) that the optimal value function is nonincreasing with respect to health state for any decision epochs, the optimal decision of nature for any decision epoch \( t \in \mathcal{T} \setminus \{T\} \) is the NIWC TPM, \( Q_{t}^{\text{NIWC},|\mathcal{L}|} \), of the uncertainty set \( Q_{t} \) based on Definition 1. Combined with Equation (52), we have the optimal policy of nature is \( (Q_{T-1}^{\text{NIWC},|\mathcal{L}|}, \ldots, Q_{T-1}^{\text{NIWC},|\mathcal{L}|}) \) which is stationary. \( \Box \)
Table 2  Glycosylated hemoglobin (HbA1c) used in the RMDP-TM for women. HbA1c range definition at diagnosis, the mean natural HbA1c values for each HbA1c state at diagnosis (prior to initiating medication), the initial HbA1c distributions at diagnosis, and 3-month HbA1c transition probability matrices for women.

<table>
<thead>
<tr>
<th>HbA1c State</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>HbA1c Range</td>
<td>&lt;6</td>
<td>[6.6,5)</td>
<td>[6.5,7)</td>
<td>[7.5,8)</td>
<td>[8.5,9)</td>
<td>[9.5,10)</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean HbA1c value (%)</td>
<td>5.7</td>
<td>6.25</td>
<td>6.74</td>
<td>7.24</td>
<td>7.73</td>
<td>8.23</td>
<td>8.73</td>
<td>9.22</td>
<td>9.72</td>
<td>11.73</td>
</tr>
<tr>
<td>Initial HbA1c Distribution</td>
<td>0.0771</td>
<td>0.1543</td>
<td>0.2125</td>
<td>0.18</td>
<td>0.1105</td>
<td>0.0848</td>
<td>0.0502</td>
<td>0.035</td>
<td>0.0273</td>
<td>0.0683</td>
</tr>
<tr>
<td>HbA1c state 1</td>
<td>0.6471</td>
<td>0.3529</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>HbA1c state 2</td>
<td>0.1800</td>
<td>0.5200</td>
<td>0.2200</td>
<td>0.0600</td>
<td>0.0200</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>HbA1c state 3</td>
<td>0.0435</td>
<td>0.1957</td>
<td>0.4783</td>
<td>0.2174</td>
<td>0.0652</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>HbA1c state 4</td>
<td>0.0192</td>
<td>0.0577</td>
<td>0.2500</td>
<td>0.3846</td>
<td>0.1923</td>
<td>0.0769</td>
<td>0.0192</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>HbA1c state 5</td>
<td>0.0323</td>
<td>0</td>
<td>0.1935</td>
<td>0.2903</td>
<td>0.2258</td>
<td>0.1935</td>
<td>0.0323</td>
<td>0</td>
<td>0.0323</td>
<td>0</td>
</tr>
<tr>
<td>HbA1c state 6</td>
<td>0</td>
<td>0</td>
<td>0.0370</td>
<td>0.1852</td>
<td>0.1852</td>
<td>0.2963</td>
<td>0.2222</td>
<td>0.0370</td>
<td>0.0370</td>
<td>0</td>
</tr>
<tr>
<td>HbA1c state 7</td>
<td>0</td>
<td>0.0588</td>
<td>0</td>
<td>0.0588</td>
<td>0.2353</td>
<td>0.2353</td>
<td>0.1765</td>
<td>0.1176</td>
<td>0.1176</td>
<td>0</td>
</tr>
<tr>
<td>HbA1c state 8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.2500</td>
<td>0.2500</td>
<td>0</td>
<td>0.5000</td>
<td>0</td>
</tr>
<tr>
<td>HbA1c state 9</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.1250</td>
<td>0.1250</td>
<td>0.2500</td>
<td>0.3750</td>
<td>0.1250</td>
</tr>
<tr>
<td>HbA1c state 10</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.0500</td>
<td>0.1500</td>
<td>0.0500</td>
<td>0.1000</td>
<td>0</td>
<td>0.2000</td>
<td>0.4500</td>
</tr>
</tbody>
</table>
Table 3  Glycosylated hemoglobin (HbA1c) used in the RMDP-TM for men. HbA1c range definition at diagnosis, the mean natural HbA1c values for each HbA1c state at diagnosis (prior to initiating medication), the initial HbA1c distributions at diagnosis, and 3-month HbA1c transition probability matrices for men.

<table>
<thead>
<tr>
<th>HbA1c State</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>HbA1c Range</td>
<td>&lt;6</td>
<td>[6,6.5)</td>
<td>[6.5,7)</td>
<td>[7,7.5)</td>
<td>[7.5,8)</td>
<td>[8,8.5)</td>
<td>[8.5,9)</td>
<td>[9,9.5)</td>
<td>[9.5,10)</td>
<td>10</td>
</tr>
<tr>
<td>Mean HbA1c value (%)</td>
<td>5.69</td>
<td>6.25</td>
<td>6.73</td>
<td>7.24</td>
<td>7.74</td>
<td>8.24</td>
<td>8.74</td>
<td>9.21</td>
<td>9.73</td>
<td>11.59</td>
</tr>
<tr>
<td>Initial HbA1c Distribution</td>
<td>0.0694</td>
<td>0.1388</td>
<td>0.1968</td>
<td>0.1626</td>
<td>0.1138</td>
<td>0.0919</td>
<td>0.0619</td>
<td>0.0424</td>
<td>0.0328</td>
<td>0.0896</td>
</tr>
<tr>
<td>HbA1c state 1</td>
<td>0.6667</td>
<td>0.2444</td>
<td>0.0889</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>HbA1c state 2</td>
<td>0.1346</td>
<td>0.5000</td>
<td>0.3654</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>HbA1c state 3</td>
<td>0.0794</td>
<td>0.2222</td>
<td>0.3810</td>
<td>0.2857</td>
<td>0</td>
<td>0.0317</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>HbA1c state 4</td>
<td>0.0175</td>
<td>0.0877</td>
<td>0.2807</td>
<td>0.3860</td>
<td>0.1579</td>
<td>0.0702</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>HbA1c state 5</td>
<td>0</td>
<td>0.0465</td>
<td>0.2093</td>
<td>0.2326</td>
<td>0.2326</td>
<td>0.1628</td>
<td>0.0930</td>
<td>0.0233</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>HbA1c state 6</td>
<td>0</td>
<td>0</td>
<td>0.0800</td>
<td>0.0800</td>
<td>0.2000</td>
<td>0.3600</td>
<td>0.2000</td>
<td>0.0400</td>
<td>0</td>
<td>0.0400</td>
</tr>
<tr>
<td>HbA1c state 7</td>
<td>0.1071</td>
<td>0</td>
<td>0.0357</td>
<td>0.1071</td>
<td>0.0714</td>
<td>0.1429</td>
<td>0.2143</td>
<td>0.2500</td>
<td>0.0357</td>
<td>0.0357</td>
</tr>
<tr>
<td>HbA1c state 8</td>
<td>0</td>
<td>0.0833</td>
<td>0</td>
<td>0.0833</td>
<td>0.2500</td>
<td>0.1667</td>
<td>0</td>
<td>0.2500</td>
<td>0.0833</td>
<td>0.0833</td>
</tr>
<tr>
<td>HbA1c state 9</td>
<td>0.0556</td>
<td>0.0556</td>
<td>0</td>
<td>0.0556</td>
<td>0.1667</td>
<td>0.1111</td>
<td>0.1111</td>
<td>0.2222</td>
<td>0.0556</td>
<td>0.1667</td>
</tr>
<tr>
<td>HbA1c state 10</td>
<td>0</td>
<td>0</td>
<td>0.0588</td>
<td>0.1176</td>
<td>0.0588</td>
<td>0.1765</td>
<td>0.1176</td>
<td>0.0588</td>
<td>0.0882</td>
<td>0.3236</td>
</tr>
</tbody>
</table>
Table 4  Left-hand-side maximum deviation of the TPM in RMDP-TM for women.

<table>
<thead>
<tr>
<th>HbA1c State</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>HbA1c state 1</td>
<td>0.1143</td>
<td>0.1143</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>HbA1c state 2</td>
<td>0.0536</td>
<td>0.0696</td>
<td>0.0577</td>
<td>0.0331</td>
<td>0.0195</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>HbA1c state 3</td>
<td>0.0296</td>
<td>0.0577</td>
<td>0.0726</td>
<td>0.0599</td>
<td>0.0359</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>HbA1c state 4</td>
<td>0.0188</td>
<td>0.0319</td>
<td>0.0592</td>
<td>0.0665</td>
<td>0.0539</td>
<td>0.0364</td>
<td>0.0188</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>HbA1c state 5</td>
<td>0.0313</td>
<td>0</td>
<td>0.0699</td>
<td>0.0804</td>
<td>0.0740</td>
<td>0.0699</td>
<td>0.0313</td>
<td>0</td>
<td>0.0313</td>
<td>0</td>
</tr>
<tr>
<td>HbA1c state 6</td>
<td>0</td>
<td>0</td>
<td>0.0358</td>
<td>0.0737</td>
<td>0.0737</td>
<td>0.0866</td>
<td>0.0789</td>
<td>0.0358</td>
<td>0.0358</td>
<td>0</td>
</tr>
<tr>
<td>HbA1c state 7</td>
<td>0</td>
<td>0.0563</td>
<td>0</td>
<td>0.0563</td>
<td>0.1014</td>
<td>0.1014</td>
<td>0.0911</td>
<td>0.0770</td>
<td>0.0770</td>
<td>0</td>
</tr>
<tr>
<td>HbA1c state 8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.2134</td>
<td>0.2134</td>
<td>0</td>
<td>0.2464</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>HbA1c state 9</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.1153</td>
<td>0.1153</td>
<td>0.1509</td>
<td>0.1687</td>
<td>0.1153</td>
<td>0</td>
</tr>
<tr>
<td>HbA1c state 10</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.0480</td>
<td>0.0787</td>
<td>0.0480</td>
<td>0.0661</td>
<td>0</td>
<td>0.0882</td>
<td>0.1097</td>
</tr>
</tbody>
</table>

Table 5  Right-hand-side maximum deviation of the TPM in RMDP-TM for women.

<table>
<thead>
<tr>
<th>HbA1c State</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>HbA1c state 1</td>
<td>0.3529</td>
<td>0.6313</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>HbA1c state 2</td>
<td>0.2959</td>
<td>0.3848</td>
<td>0.3191</td>
<td>0.1829</td>
<td>0.1078</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>HbA1c state 3</td>
<td>0.1638</td>
<td>0.3186</td>
<td>0.4012</td>
<td>0.3312</td>
<td>0.1983</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>HbA1c state 4</td>
<td>0.1037</td>
<td>0.1761</td>
<td>0.3271</td>
<td>0.3675</td>
<td>0.2977</td>
<td>0.2013</td>
<td>0.1037</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>HbA1c state 5</td>
<td>0.1728</td>
<td>0</td>
<td>0.3865</td>
<td>0.4440</td>
<td>0.4090</td>
<td>0.3865</td>
<td>0.1728</td>
<td>0</td>
<td>0.1728</td>
<td>0</td>
</tr>
<tr>
<td>HbA1c state 6</td>
<td>0</td>
<td>0</td>
<td>0.1980</td>
<td>0.4072</td>
<td>0.4072</td>
<td>0.4786</td>
<td>0.4358</td>
<td>0.1980</td>
<td>0.1980</td>
<td>0</td>
</tr>
<tr>
<td>HbA1c state 7</td>
<td>0</td>
<td>0.3108</td>
<td>0</td>
<td>0.3108</td>
<td>0.5603</td>
<td>0.5603</td>
<td>0.5036</td>
<td>0.4256</td>
<td>0.4256</td>
<td>0</td>
</tr>
<tr>
<td>HbA1c state 8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.7500</td>
<td>0.7500</td>
<td>0</td>
<td>0.5000</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>HbA1c state 9</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.6369</td>
<td>0.6369</td>
<td>0.7500</td>
<td>0.6250</td>
<td>0.6369</td>
<td>0</td>
</tr>
<tr>
<td>HbA1c state 10</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.2654</td>
<td>0.4349</td>
<td>0.2654</td>
<td>0.3654</td>
<td>0</td>
<td>0.4872</td>
<td>0.5500</td>
</tr>
</tbody>
</table>
### Table 6  
Left-hand-side maximum deviation of the TPM in RMDP-TM for men.

<table>
<thead>
<tr>
<th>HbA1c State</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>HbA1c state 1</td>
<td>0.0693</td>
<td>0.0632</td>
<td>0.0418</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>HbA1c state 2</td>
<td>0.0467</td>
<td>0.0683</td>
<td>0.0658</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>HbA1c state 3</td>
<td>0.0336</td>
<td>0.0516</td>
<td>0.0603</td>
<td>0.0561</td>
<td>0</td>
<td>0.0218</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>HbA1c state 4</td>
<td>0.0171</td>
<td>0.0369</td>
<td>0.0587</td>
<td>0.0636</td>
<td>0.0476</td>
<td>0.0334</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>HbA1c state 5</td>
<td>0</td>
<td>0.0317</td>
<td>0.0612</td>
<td>0.0635</td>
<td>0.0635</td>
<td>0.0555</td>
<td>0.0437</td>
<td>0.0227</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>HbA1c state 6</td>
<td>0</td>
<td>0</td>
<td>0.0535</td>
<td>0.0535</td>
<td>0.0789</td>
<td>0.0946</td>
<td>0.0789</td>
<td>0.0386</td>
<td>0</td>
<td>0.0386</td>
</tr>
<tr>
<td>HbA1c state 7</td>
<td>0.0576</td>
<td>0</td>
<td>0.0346</td>
<td>0.0576</td>
<td>0.0680</td>
<td>0.0764</td>
<td>0.0807</td>
<td>0.0346</td>
<td>0.0346</td>
<td></td>
</tr>
<tr>
<td>HbA1c state 8</td>
<td>0</td>
<td>0.0786</td>
<td>0</td>
<td>0.0786</td>
<td>0.1232</td>
<td>0.1060</td>
<td>0</td>
<td>0.1232</td>
<td>0.0786</td>
<td></td>
</tr>
<tr>
<td>HbA1c state 9</td>
<td>0.0532</td>
<td>0.0532</td>
<td>0</td>
<td>0.0532</td>
<td>0.0866</td>
<td>0.0730</td>
<td>0.0730</td>
<td>0.0966</td>
<td>0.0532</td>
<td>0.0866</td>
</tr>
<tr>
<td>HbA1c state 10</td>
<td>0</td>
<td>0</td>
<td>0.0398</td>
<td>0.0545</td>
<td>0.0398</td>
<td>0.0644</td>
<td>0.0545</td>
<td>0.0398</td>
<td>0.0480</td>
<td>0.0791</td>
</tr>
</tbody>
</table>

### Table 7  
Right-hand-side maximum deviation of the TPM in RMDP-TM for men.

<table>
<thead>
<tr>
<th>HbA1c State</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>HbA1c state 1</td>
<td>0.3333</td>
<td>0.3489</td>
<td>0.2311</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>HbA1c state 2</td>
<td>0.2578</td>
<td>0.3777</td>
<td>0.3637</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>HbA1c state 3</td>
<td>0.1855</td>
<td>0.2853</td>
<td>0.3332</td>
<td>0.3100</td>
<td>0</td>
<td>0.1203</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>HbA1c state 4</td>
<td>0.0947</td>
<td>0.2041</td>
<td>0.3242</td>
<td>0.3512</td>
<td>0.2631</td>
<td>0.1843</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>HbA1c state 5</td>
<td>0</td>
<td>0.1749</td>
<td>0.3379</td>
<td>0.3509</td>
<td>0.3509</td>
<td>0.3066</td>
<td>0.2413</td>
<td>0.1252</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>HbA1c state 6</td>
<td>0</td>
<td>0</td>
<td>0.2955</td>
<td>0.2955</td>
<td>0.4357</td>
<td>0.5229</td>
<td>0.4357</td>
<td>0.2135</td>
<td>0</td>
<td>0.2135</td>
</tr>
<tr>
<td>HbA1c state 7</td>
<td>0.3184</td>
<td>0</td>
<td>0.1910</td>
<td>0.3184</td>
<td>0.2651</td>
<td>0.3602</td>
<td>0.4224</td>
<td>0.4457</td>
<td>0.1910</td>
<td>0.1910</td>
</tr>
<tr>
<td>HbA1c state 8</td>
<td>0</td>
<td>0.4346</td>
<td>0</td>
<td>0.4346</td>
<td>0.6808</td>
<td>0.5860</td>
<td>0</td>
<td>0.6808</td>
<td>0.4346</td>
<td>0.4346</td>
</tr>
<tr>
<td>HbA1c state 9</td>
<td>0.2941</td>
<td>0.2941</td>
<td>0</td>
<td>0.2941</td>
<td>0.4784</td>
<td>0.4035</td>
<td>0.4035</td>
<td>0.5337</td>
<td>0.2941</td>
<td>0.4784</td>
</tr>
<tr>
<td>HbA1c state 10</td>
<td>0</td>
<td>0</td>
<td>0.2198</td>
<td>0.3010</td>
<td>0.2198</td>
<td>0.3561</td>
<td>0.3010</td>
<td>0.2198</td>
<td>0.2649</td>
<td>0.4370</td>
</tr>
</tbody>
</table>