Probabilistic Sensitivity Analysis on Markov Decision Processes with Uncertain Transition Probabilities: An Application in Evaluating Treatment Decisions for Type 2 Diabetes

YUANHUI ZHANG[1]
Operations Research Graduate Program,
North Carolina State University,
Raleigh, NC 27695-7913, USA

HAIPENG WU[2]
Optym, LLC,
2153 SE Hawthorne Road,
Gainesville, FL 32641, USA

BRIAN T. DENTON[3]
Department of Industrial and Operations Engineering,
University of Michigan,
1205 Beal Avenue,
Ann Arbor, MI 48109, USA

JAMES R. WILSON[4]
Edward P. Fitts Department of Industrial and Systems Engineering,
North Carolina State University,
Raleigh, NC 27695, USA

JENNIFER M. LOBO[5]
Department of Public Health Sciences,
University of Virginia,
1300 Jefferson Park Avenue,
Charlottesville, VA 22908, USA

[1]E-mail: yuanhui.zhang@gmail.com. Telephone: (919) 271-6331.

[2]E-mail: haipengwu1989@gmail.com. Telephone: (734) 834-3016

[3]Member, Institute of Industrial Engineers. E-mail: btdenton@umich.edu. Telephone: (734) 763-2060. Fax: (734) 764-3451.

[4]Fellow, Institute of Industrial Engineers, and corresponding author. E-mail: jwilson@ncsu.edu. Telephone: (919) 515-6415. Fax: (919) 515-5281.

[5]Member, Institute of Industrial Engineers. E-mail: jem4yb@virginia.edu. Telephone: (434) 924-2813. Fax: (434) 243-3418.

Probabilistic Sensitivity Analysis on Markov Decision Processes with Uncertain Transition Probabilities: An Application in Evaluating Treatment Decisions for Type 2 Diabetes

Markov decision processes (MDPs) are commonly used for decision-making studies in many application domains; however, there are no widely adopted methods for performing sensitivity analysis on the associated transition probability matrices (TPMs). This article describes two simulation-based approaches for conducting probabilistic sensitivity analysis of a given finite-state, finite-horizon, discrete-time MDP with TPMs which may vary within specified uncertainty sets or according to appropriate or specified probability distributions. The first approach assumes no prior knowledge of the TPM's distribution, and each row is sampled uniformly over the corresponding uncertainty set. The second approach involves random sampling from the (truncated) multivariate normal distribution of the TPM's maximum likelihood estimator subject to the condition that each row has nonnegative elements, and sums to one. the TPM independently. The proposed methods are easy to implement and have reasonable computation times. As an illustrative example, these two methods are applied to a medical decision-making problem involving the evaluation of treatment guidelines for glycemic control of patients with type 2 diabetes in which the natural variation in glycated hemoglobin (HbA1c) is modeled as a Markov chain, and the HbA1c transition probabilities are subject to uncertainty.

Keywords: Robustness and sensitivity analysis, Markov decision process, Transition probability matrices, Medical decision-making, Monte Carlo simulation

---

1. Introduction

Markov decision processes (MDPs) are commonly used in decision-making studies for which there is uncertainty in the way that the system evolves over time. Such models represent a dynamic system in terms of a set of mutually exclusive and collectively exhaustive states of the system and the transitions among those states over a given time horizon. The decision maker takes actions at different time epochs to partly control model outcomes.

For a finite-state, finite-horizon, discrete-time MDP, the associated state transition probability matrix (TPM) governs the evolution of the process over the entire time horizon. Typically the TPM is estimated from observed sample paths of the target process and is therefore subject to statistical uncertainty. In some cases, owing to limited historical data, confidence interval (CI) estimators of the individual transition probabilities of the TPM may have half-lengths that are large relative to the corresponding point estimators of those quantities. More generally, a confidence region estimator of the entire TPM may have a large "size" (e.g., the maximum length of a diameter of the confidence region that passes through the point estimator of the TPM) relative to an appropriate matrix norm of the TPM's point estimator (e.g., the l_2-norm; see p. 341

of Horn and Johnson (2013)). In such situations, using only point estimators for the transition probabilities fails to account for this uncertainty.

The Joint Modeling Good Research Practices Task Force of the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) and The Society for Medical Decision Making (SMDM) recommends two types of sensitivity analysis for estimating the sensitivity of model outcomes when there is uncertainty in model parameters: deterministic sensitivity analysis (DSA) and probabilistic sensitivity analysis (PSA) (Briggs et al. 2012). DSA is done by systematically varying the values of one or more parameters of a decision model while keeping the other parameters unchanged, and then comparing the associated outputs to see whether the model's key performance measures are sensitive to changes in the selected parameters (Saltelli et al. 2000). PSA is done by simultaneously varying one or more parameters, with multiple sets of parameter values being sampled from selected probability distributions (Doubilet et al. 1985). Unfortunately, neither method in its conventional form can be used to perform sensitivity analysis on a TPM because in each row of a TPM, the entries must be nonnegative and sum to one.

Bootstrapping has been proposed as a means to conduct sensitivity analysis when the original data set for the observed sample paths is available (Craig and Sendi 2002); however, in practice it is often the case that the original data set is not available, or the region within which the TPM varies, referred to as the associated uncertainty set, is defined subjectively by decision makers, perhaps based on multiple sources of information. Mannor et al. (2007) provided the closed-form approximation of the variance of the value function estimates for a stationary infinite-horizon MDP in light of the closed-form expression of the value function. This approach provides information on the variance of the value function but not the distribution. Although numerous algorithms have been developed for generating random vectors that have a given distribution on a given convex region (Hörmann et al. 2004), most of these algorithms are designed for uniform sampling of convex polytopes or simple polytopes. In general these methods are much more difficult to implement than the methods proposed in this article for random sampling of TPMs; moreover, depending on the structure of the uncertainty set, the former methods may not be directly applicable to random sampling of TPMs.

In this article we present two simulation-based approaches for conducting PSA on the TPM of a finite-state, finite-horizon, discrete-time MDP. In the first method, each row of the TPM is randomly sampled from the uniform distribution on its uncertainty set, which is assumed to be convex and bounded. It can be applied to cases in which (i) there is relatively little information for specifying the uncertainty set; or (ii) the decision maker wants to define the uncertainty set subjectively based on expert opinion. In the second method, each row of the TPM is randomly sampled from the (truncated) multivariate normal distribution of the maximum likelihood estimator (MLE) for that row, subject to the condition that the row has nonnegative elements and sums to one. This approach can be applied to cases in which interest is focused on the uncertainty arising from statistical estimation of the TPM. For a specific one-period reward function defined on each state of the MDP, the main goal is to compute the expected value of the total reward accumulated by the MDP over the given finite time horizon. We perform PSA on the conditional expected value of the total accumulated reward given the TPM, where the TPM is randomly sampled from its corresponding uncertainty set or appropriate probability distribution. We demonstrate the application of those two approaches in a case study concerning an MDP that was designed to evaluate glycemic control treatment guidelines for patients with

type 2 diabetes, and we estimate the impact of the uncertainty in TPMs on a patient's expected quality-adjusted life-years (QALYs) and the expected total medication costs.

The main contributions of this paper are twofold: from the methodological point of view, we present two efficient sampling methods specifically for TPMs which are easy implemented and are applicable to two important forms of uncertainty in TPMs. From the application point of view, to our knowledge, we present the first results of PSA on TPMs of an MDP for a medical treatment problem, and demonstrate the use of these methods to draw conclusions about the sensitivity of the model outcomes.

2. Methods

In this section, we provide the general problem statement and notation in Section 2.1, present the procedure of conducing PSA on TPMs using Monte Carlo simulation in Section 2.2, and provide the two algorithms for sampling TPMs in Section 2.3.

2.1. Problem Statement and Notation

Following standard notation from Puterman (2005), we define a finite-state finite-horizon, discrete-time MDP as a 5-tuple $< \mathscr{T}, \mathscr{S}, \mathscr{A}, p_t(\cdot|s,a), r_t(s,a) >$, where $\mathscr{T} = \{1,2,\ldots,T\}, T < \infty$, represents the set of time epochs, $\mathscr{S} = \{s(1),\ldots,s(n)\}$ represents the set of states in the model where n represents the number of states, and $\mathscr{A}$ represents the set of all possible actions. As a result of choosing action $a \in \mathscr{A}$, in state $s \in \mathscr{S}$ at time epoch t, the decision maker receives a one-period reward, $r_t(s,a)$, and the state at the next time epoch is determined by the probability distribution $p_t(\cdot|s,a)$. For any time epoch $t \in \mathscr{T}\backslash\{T\}$ and action $a \in \mathscr{A}$, the TPM, $P_t^a = [p_t(\cdot|s(1),a)^\intercal, \ldots, p_t(\cdot|s(n),a)^\intercal]^\intercal$, is the matrix consisting of all the one-step transition probabilities, $p_t(s(i)|s(j),a), \forall s(i), s(j) \in \mathscr{S}$.

A *decision rule*, $d_t : \mathscr{S} \to \mathscr{A}, \forall t \in \mathscr{T}\backslash\{T\}$, specifies the action choice when the system occupies state $s \in \mathscr{S}$ at time epoch $t \in \mathscr{T}\backslash\{T\}$, and a *policy*, $\boldsymbol{\pi} = \{d_1, d_2, \ldots, d_{T-1}\}$, is a sequence of decision rules, which specifies the decision to be used at all time epochs. We adopt the convention that, in finite horizon problems, decisions are not made at the last time epoch T, and the reward $r_T(s_T)$ represents the terminal reward when the system occupies state $s_T \in \mathscr{S}$ at time epoch T.

A policy $\boldsymbol{\pi}$ induces a probability distribution on the set of all realizations of the MDPs under the policy $\boldsymbol{\pi}$. We denote $v^{\boldsymbol{\pi}}(s_1|s_1 = s), \forall s \in \mathscr{S}$, to be the conditional expected total reward accumulated over the entire time horizon given the policy $\boldsymbol{\pi}$ is used and the system is in state s at the first time epoch, and we calculate $v^{\boldsymbol{\pi}}(s)$ as follows:

$$v^{\boldsymbol{\pi}}(s_1|s_1 = s) \equiv \mathrm{E}_s^{\boldsymbol{\pi}} \left[\sum_{t=1}^{T-1} r_t(s_t, a_t) + r_T(s_T) \right], \forall s \in \mathscr{S}. \tag{1}$$

We denote $\boldsymbol{\Pi}$ to be the probability distribution of the states at the first time epoch, where

$$\boldsymbol{\Pi} = (\Pi(s(1)), \Pi(s(2)), \ldots, \Pi(s(n))), \sum_{i=1}^{n} \Pi(s(i)) = 1,$$

and $v^{\boldsymbol{\pi}}$ to be the expected total reward over all initial states, and calculate it as follows:

$$v^{\boldsymbol{\pi}} \equiv \sum_{i=1}^{n} \Pi(s(i)) v^{\boldsymbol{\pi}}(s_1 | s_1 = s(i)). \tag{2}$$

Our goal is to examine the sensitivity of the value function with respect to variation in TPMs, i.e., we seek to find the distribution, the maximum, and the minimum values of the value function over all possible TPMs.

2.2. PSA on TPM based on Monte Carlo simulation

When considering uncertain state transition probabilities, the value function (2) becomes the conditional expected total reward accumulated over the entire time horizon given the TPMs governing the state transitions. By assuming that each row of the relevant TPM can be sampled independently from its uncertainty set or its probability distribution, the following Monte Carlo simulation-based PSA procedure can be used to conduct PSA on TPMs for the MDP.

Algorithm 1 Procedure to conduct PSA on TPMs for Markov decision process

```
Step 1: Create M (number of samples) TPMs
for all t ∈ 𝒯\{T} do
    for all a ∈ 𝒜 do
        for all i ∈ {1,...,n} do
            Sample M vectors for the i^th row of the TPM P_t^a
        end for
    end for
end for
Step 2: Generate estimates of the value function
while j ≤ M do
    Calculate the value function based on the j^th sample of the TPM
end while
Step 3: Examine the sensitivity of the value function on TPMs based on the value function estimates
```

Under certain regularity conditions (the smoothness of the value function with respect to the transition probabilities), Algorithm 1, based on random search of a feasible region, will converge in probability to the optimal value (minimum or maximum) of the simulation's expected response over the feasible region–provided that an appropriate scheme is used for increasing each of the following as the total simulation budget increases: (i) the number of points that are randomly sampled from a positive, continuous probability density concentrated on the region; and (ii) the number of independent replications of the stochastic simulation response at each sampled point (Cheng 2013, Chia and Glynn 2013). In the context of an MDP in which the transition probabilities represent the only source of uncertainty, item (ii) above is not relevant; and therefore we can estimate the minimum and maximum values of the conditional expected value of the total accumulated reward given the TPM by randomly sampling a sufficiently large number of TPMs from any positive, continuous density concentrated on the uncertainty set (Devroye (1978), Proposition 4.1).

2.3. Random sampling of the transition probability matrix

We detail two methods for sampling rows of a TPM in Sections 2.3.1 and 2.3.2. All the algorithms introduced in this section have been implemented using MATLAB R2013a (The MathWorks Inc. 2013). The source codes are available from the authors.

From hereafter, we denote $\boldsymbol{Q}$ to be the generic form of an $n \times n$ TPM, and denote $\boldsymbol{q}_i = \left[q_{i,1}, \ldots, q_{i,n}\right]$ to be the i^{th} row where $q_{i,j}$ is the conditional probability that in the next time epoch (period), the underlying Markov chain moves to state s_j given that it is currently in state s_i for $i, j = 1, \ldots, n$. The transition probabilities are subject to the following well-known requirements:

$$\left.\begin{array}{ll} \sum_{j=1}^{n} q_{i,j} = 1, & \text{for } i = 1, \ldots, n, \\ 0 \leq q_{i,j} \leq 1, & \text{for } i, j = 1, \ldots, n \end{array}\right\}. \tag{3}$$

As a consequence of Equation (3), the TPM sampling method used for implementing Line 5 of Algorithm 1 must sample row vectors of the TPM that satisfy Equation (3).

2.3.1. Method 1: Uniform sampling of each row's uncertainty set

A common approach to define the uncertainty set of a single parameter or a set of unrelated parameters is to select a bounded region in which the relevant quantities may vary. If the quantities in question compose a row of a TPM, then the uncertainty set A_i is a subset of Δ^{n-1}, the standard (or unit) simplex of dimension $(n-1)$ in n-dimensional Euclidean space $\mathbb{R}^n$ so that:

$$\boldsymbol{q}_i \in A_i \subseteq \Delta^{(n-1)} \equiv \left\{\boldsymbol{x} = \left[x_1, \ldots, x_n\right] \in \mathbb{R}^n : x_j \geq 0 \text{ for } j = 1, \ldots, n \text{ and } \sum_{j=1}^{n} x_j = 1\right\} \text{ for } i = 1, \ldots, n. \tag{4}$$

The uncertainty set might be specified subjectively, or it might be specified by a separate CI for each individual parameter. The following method can be used to construct a $100 \times (1-\beta)\%$ confidence region for the relevant TPM to be the uncertainty set for that TPM. For $i = 1, \ldots, n$, we seek to construct an uncertainty set A_i for the i^{th} row $\boldsymbol{q}_i$ of the TPM $\boldsymbol{Q} = [q_1^{\intercal}, \ldots, q_n^{\intercal}]^{\intercal}$ such that the uncertainty hyper-rectangle $A = A_1 \times \cdots \times A_n$ is a confidence region estimator of $\boldsymbol{Q}$ with coverage probability

$$\Pr\{\boldsymbol{Q} \in A\} \geq 1 - \beta \text{ for } \beta \in (0,1). \tag{5}$$

To construct A_i for a given value of the confidence coefficient $1-\beta$, we define the following CI estimator of $q_{i,j}$ for $j = 1, \ldots, n$:

$$\text{CI}_{i,j} \equiv \left(\hat{q}_{i,j} - \left[\chi^2_{n-1,\beta/(2n)} \frac{\hat{q}_{i,j}(1-\hat{q}_{i,j})}{N_i}\right]^{1/2}, \hat{q}_{i,j} + \left[\chi^2_{n-1,\beta/(2n)} \frac{\hat{q}_{i,j}(1-\hat{q}_{i,j})}{N_i}\right]^{1/2}\right) \tag{6}$$

where $\hat{q}_{i,j}$ denotes the MLE of $q_{i,j}$ based on N_i observed transitions out of state i, and in general $\chi^2_{k,\omega}$ denotes the ω quantile of the chi-squared distribution with k degrees of freedom for $k = 1, 2, \ldots$ and $\omega \in (0,1)$. Let $A_i \equiv \text{CI}_{i,1} \times \cdots \times \text{CI}_{i,n}$. It follows from Theorem 3 of Gold (1963) that

$$\Pr\{\boldsymbol{q_i} \in A_i\} \geq 1 - \beta/n \text{ for } i = 1, \ldots, n; \tag{7}$$

and therefore the Bonferroni inequality ensures that Equation (5) holds.

Perhaps the simplest approach for conducting PSA on a TPM over a given uncertainty set is to perform the following acceptance-rejection procedure: (a) generate the required random vectors $\{\boldsymbol{q}_i : i = 1, \ldots, n\}$ from the uniform distribution on the standard simplex $\Delta^{(n-1)}$ using the method specified in Theorem 2.1 on p. 207 of Devroye (1986); (b) if the condition $\boldsymbol{q}_i \in A_i$ is satisfied for $i = 1, \ldots, n$ then accept the associated TPM; (c) if the condition of step (b) is not satisfied, then reject the sampled TPM and repeat steps (a) and (b) until a sampled TPM is finally accepted; and (d) repeat steps (a)-(c) until the prespecified number of TPMs have been accepted for use in the PSA. However, a major limitation of such a scheme is its inefficiency, especially when the dimension n of each row grows large because as the dimension increases, the probability of rejection in step (c) approaches one rapidly.

Smith (1984) proposed a random-direction algorithm for generating a series of random points $\{\boldsymbol{X}_j, j = 1, 2, \ldots\}$ within a bounded, convex region of $\mathbb{R}^n$ having a nonempty interior so that asymptotically $\boldsymbol{X}_j$ is uniformly distributed over the region as $j \to \infty$. Since the standard simplex $\Delta^{(n-1)}$ is bounded and convex with a nonempty interior, if the uncertainty set is also convex and has a nonempty interior (such as the intersection of $\Delta^{(n-1)}$ with a hyperrectangle that has nonzero volume), then this procedure can be used to perform PSA on TPMs. For the random direction sampled on each iteration of Smith's scheme, the user must determine the "antipodal" boundary points of the uncertainty set defined by the latest sampled point in the uncertainty set and the latest sampled direction for moving away from that point; and this boundary-point determination generally requires a line search or some other subscheme that has been specially tailored to the problem at hand. In Algorithm 2 given below, Smith's random-direction scheme is modified for general applications of our PSA procedure so that there is no need for the user to provide a subscheme for boundary-point determination. In our experience such user-provided subschemes can be relatively time consuming not only to implement but also to execute in practice. For example, prior to the development of Algorithm 2, we used a variant of Smith's original procedure in which each iteration required solving two linear-programming problems to determine the relevant boundary points of the uncertainty set. By using Algorithm 2 instead, not only did we avoid implementing and solving these linear-programming problems but also we achieved a 60% reduction in the time required to sample 1,000 TPMs of dimension 10×10. Moreover, we have found in practice that Algorithm 2 retains its computational advantage in sampling higher-dimensional TPMs from the uniform distribution on the uncertainty set.

Algorithm 2 Uniform sampling $\boldsymbol{q}_i$ in the uncertainty set A_i for the i^{th} row of the TPM, $\boldsymbol{Q}$

1: Choose an initial point $\boldsymbol{X}_0 \in \text{int}(A_i)$ and set $j \leftarrow 1$, $k \leftarrow 1$.
2: **while** $j \leq W + M$, (W represents the warm-up period, and M represents the number of samples) **do**
3: Starting from $\boldsymbol{X}_{j-1}$, randomly sample a direction $\boldsymbol{d}$ so that for $\lambda \in \mathbb{R}$, the line $\boldsymbol{X}_{j-1} + \lambda \boldsymbol{d}$ lies in the $(n-1)$-dimensional hyperplane containing A_i.
4: Set $\bar{\lambda} \leftarrow \max\{\alpha^+ : \alpha^+ > 0 \text{ and } \boldsymbol{X}_{j-1} + \alpha^+ \boldsymbol{d} \in \Delta^{n-1}\}$ as follows:

4a. Set $\lambda_l^+ \leftarrow \begin{cases} 0, & \text{if } d_l = 0, \\ (1 - x_{j-1,l})/d_l, & \text{if } d_l > 0, \\ -x_{j-1,l}/d_l, & \text{if } d_l < 0, \end{cases}$ for $1 \leq l \leq n$,

4b. Set $\bar{\lambda} \leftarrow \min\{\lambda_l^+ : \lambda_l^+ > 0 \text{ and } 1 \leq l \leq n\}$.
5: Set $\underline{\lambda} \leftarrow \min\{\alpha^- : \alpha^- < 0 \text{ and } \boldsymbol{X}_{j-1} + \alpha^- \boldsymbol{d} \in \Delta^{n-1}\}$ as follows:

5a. Set $\lambda_l^- \leftarrow \begin{cases} 0, & \text{if } d_l = 0, \\ -x_{j-1,l}/d_l, & \text{if } d_l > 0, \\ (1 - x_{j-1,l})/d_l, & \text{if } d_l < 0, \end{cases}$ for $1 \leq l \leq n$,

5b. Set $\underline{\lambda} \leftarrow \max\{\lambda_l^- : \lambda_l^- < 0 \text{ and } 1 \leq l \leq n\}$.
6: Set $\lambda \leftarrow$ a random sample from the Uniform $[\underline{\lambda}, \bar{\lambda}]$ distribution.
7: **while** $\boldsymbol{X}_{j-1} + \lambda \boldsymbol{d} \notin A_i$ **do**
8: if $\lambda \geq 0$, then set $\bar{\lambda} \leftarrow \lambda$
9: else set $\underline{\lambda} \leftarrow \lambda$
10: Set $\lambda \leftarrow$ a random sample from the Uniform $[\underline{\lambda}, \bar{\lambda}]$ distribution
11: **end while**
12: Set $\boldsymbol{X}_j \leftarrow \boldsymbol{X}_{j-1} + \lambda \boldsymbol{d}$
13: Set $j \leftarrow j + 1$
14: **if** $j > W$ **then**
15: Set $\boldsymbol{q}_i^{(k)} \leftarrow \boldsymbol{X}_j$ (k is the index of the sample)
16: Set $k \leftarrow k + 1$
17: **end if**
18: **end while**

In Algorithm 2, Step 1 requires the selection of an initial point $\boldsymbol{X}_0 \in \text{int}(A_i)$, the interior of A_i. Starting from the latest point $\boldsymbol{X}_{j-1}$ visited by the random walk within $\text{int}(A_i)$, in Step 3 we must randomly sample a direction $\boldsymbol{d} = [d_1, \ldots, d_n]$ such that for $\lambda \in \mathbb{R}$, the straight line $\boldsymbol{X}_{j-1} + \lambda \boldsymbol{d}$ lies in the $(n-1)$-dimensional hyperplane containing A_i. To sample $\boldsymbol{d}$, we generate $(n-1)$ independent standard normal random variables $\{Z_l : l = 1, \ldots, n-1\}$, set $\boldsymbol{Z}^* \leftarrow [Z_1, \ldots, Z_{n-1}]$, and take

$$\boldsymbol{d}^* \equiv [d_1, \ldots, d_{n-1}] \leftarrow \boldsymbol{Z}^* / \|\boldsymbol{Z}^*\|, \tag{8}$$

where $\|\boldsymbol{Z}^*\| \equiv \left(\sum_{i=1}^{n-1} Z_i^2\right)^{1/2}$ is the length (norm) of $\boldsymbol{Z}^*$. From Equation (8) we see that $\boldsymbol{d}^*$ is a random direction in $\mathbb{R}^{n-1}$ that is easily mapped to the desired direction $\boldsymbol{d}$ in $\mathbb{R}$ according to

$$\boldsymbol{d} \equiv [\boldsymbol{d}^*, -\boldsymbol{d}^* \mathbf{1}_{n-1}] = [d_1, \ldots, d_{n-1}, -\sum_{j=1}^{n-1} d_j], \tag{9}$$

where $\mathbf{1}_{n-1}$ denotes the $(n-1) \times 1$ vector of ones. It follows immediately from Equation (9) that $\boldsymbol{X}_{j-1} + \lambda \boldsymbol{d}$ lies in the $(n-1)$-dimensional hyperplane containing A_i for every $\lambda \in \mathbb{R}$, as required.

Steps 4 to 12 of Algorithm 2 deliver the next point $\boldsymbol{X}_j$ that is uniformly distributed on the line segment defined by the intersection of the uncertainty set A_i and the straight line passing through the latest sampled point $\boldsymbol{X}_{j-1}$ with the latest sampled direction $\boldsymbol{d}$. In Steps 4 and 5 of Algorithm 2, the quantities $\bar{\lambda}$ and $\underline{\lambda}$ are, respectively, the initial upper and lower bounds on acceptable values for the parameter λ such that the trial point $\boldsymbol{X}_{j-1} + \lambda \boldsymbol{d}$ falls in the standard simplex. If the trial point $\boldsymbol{X}_{j-1} + \lambda \boldsymbol{d}$ sampled in Step 6 does not fall in the uncertainty set A_i then in Steps 7 to 10 the line segment $\{\boldsymbol{X}_{j-1} + \lambda \boldsymbol{d} : \lambda \in [\underline{\lambda}, \bar{\lambda}]\}$ is truncated at the latest trial point so that in Step 11, a new trial point is sampled uniformly within the reduced line segment. This iterative process continues until a trial point is sampled within the uncertainty set A_i and is therefore accepted as the next point $\boldsymbol{X}_j$ in the random walk. Steps 14 to 17 of Algorithm 2 are used to return the points sampled after the warm-up period (W steps).

Remark 1. On each iteration of Algorithm 2, with probability one, steps 7 through 11 will be performed a finite number of times before delivering a point $\boldsymbol{X}_j \in A_i$ that is uniformly distributed on the diameter of A_i passing through $\boldsymbol{X}_{j-1}$ in the latest sampled direction; and this is the same outcome achieved on each iteration of Smith's original random-walk procedure. Therefore as $j \to \infty$, the point $\boldsymbol{X}_j$ delivered by Algorithm 2 is uniformly distributed on A_i.

Smith (1984) has shown that the original random-direction algorithm has an exponential convergence rate to the uniform distribution on the uncertainty set. In view of Remark 1, we see that the gap between the uniform distribution and the distribution of the j^{th} random point $\boldsymbol{X}_j$ generated by Algorithm 2 decreases exponentially in j as $j \to \infty$. Specifically, for any measurable subset

$$\left| \Pr(\boldsymbol{X}_j \in G) - \frac{\text{vol}(G)}{\text{vol}(A_i)} \right| < \left(1 - \frac{\gamma}{n 2^{n-1}}\right)^{j-1} \tag{10}$$

where $\text{vol}(G)$ and $\text{vol}(A_i)$ denote the volumes of G and A_i, respectively; and γ is the maximum value of $\text{vol}(A_i)/\text{vol}(S^n)$ where S^n is an n-dimensional sphere in $\mathbb{R}^n$ that contains A_i. Equation (10) can be used in assigning the warm-up period for Algorithm 2. Given an upper limit δ on the right-hand side of Equation (10), an appropriate length for the warm-up period is estimated to be j_0 steps, where $j_0 = -n 2^{n-1} \ln \delta / \gamma$. Although this estimate for the warm-up period is exponentially increasing in n, it is required only once in each application of Algorithm 2.

2.3.2. Method 2: Generating probability vectors having a (truncated) normal distribution

For situations in which the rows of a TPM are estimated from sample paths of the target process using the method of maximum likelihood, we let N_i denote the total number of time steps in all those sample paths that started in state s_i for $i = 1, \ldots, n$. Then for large N_i (where $i = 1, \ldots, n$), the MLE $\widehat{\boldsymbol{q}}_i$ of the i^{th} row $\boldsymbol{q}_i$ of the TPM has a singular multivariate normal distribution with mean $\boldsymbol{q}_i$ and covariance matrix

$$\boldsymbol{\Sigma}_i = \frac{1}{N_i}\begin{bmatrix} q_{i,1}(1-q_{i,1}) & -q_{i,1}q_{i,2} & \cdots & -q_{i,1}q_{i,n} \\ -q_{i,2}q_{i,1} & q_{i,2}(1-q_{i,2}) & \cdots & -q_{i,2}q_{i,n} \\ \vdots & \vdots & \ddots & \vdots \\ -q_{i,n}q_{i,1} & -q_{i,n}q_{i,2} & \cdots & q_{i,n}(1-q_{i,n}) \end{bmatrix}; \tag{11}$$

moreover, the random vectors $\{\widehat{\boldsymbol{q}}_i : i = 1,\ldots,n\}$ are mutually independent (Anderson and Goodman (1957), p. 96). If $\widehat{\boldsymbol{\Sigma}}_i$ denotes the estimator of $\boldsymbol{\Sigma}_i$ obtained by replacing each transition probability $q_{i,j}$ in Equation (11) by its MLE $\widehat{q}_{i,j}$, then it is appropriate to model the uncertainty in $\boldsymbol{q}_i$ by generating random samples from $\text{Normal}_n(\widehat{\boldsymbol{q}}_i, \widehat{\boldsymbol{\Sigma}}_i)$, the n-variate normal distribution with mean vector $\widehat{\boldsymbol{q}}_i$ and covariance matrix $\widehat{\boldsymbol{\Sigma}}_i$. Note that both $\boldsymbol{\Sigma}_i$ and $\widehat{\boldsymbol{\Sigma}}_i$ are singular matrices because, for example, the sum of all n columns in each matrix is exactly equal to $\mathbf{0}_n$, the $n \times 1$ vector of zeros; and this multicollinearity is a direct result of the constraints (3) that for both $\boldsymbol{q}_i$ and $\widehat{\boldsymbol{q}}_i$, their entries must sum exactly to one (Draper and Smith (1998), p. 369).

Random sampling from the $\text{Normal}_n(\widehat{\boldsymbol{q}}_i, \widehat{\boldsymbol{\Sigma}}_i)$ distribution requires some care, because $\widehat{\boldsymbol{\Sigma}}_i$ is singular, whereas the usual methods for generating normal vectors are based on the assumption that the associated covariance matrix is nonsingular (Hörmann et al., 2004, p. 250-252). If $\widehat{\boldsymbol{q}}_i^* \equiv [\hat{q}_{i,1},\ldots,\hat{q}_{i,n-1}]$ denotes the reduced row vector consisting of the first $(n-1)$ elements of $\widehat{\boldsymbol{q}}_i$, and if $\widehat{\boldsymbol{\Sigma}}_i^*$ denotes the $(n-1)\times(n-1)$ submatrix in the upper left-hand corner of $\widehat{\boldsymbol{\Sigma}}_i$, then $\widehat{\boldsymbol{\Sigma}}_i^*$ is positive definite with probability one so that we can compute its Cholesky decomposition–that is, an $(n-1)\times(n-1)$ matrix $\boldsymbol{C}$ such that $\widehat{\boldsymbol{\Sigma}}_i^* = \boldsymbol{C}\boldsymbol{C}^\intercal$ (Anderson 2003). If $\boldsymbol{Z}^* = [Z_1,\ldots,Z_{n-1}]$ consists of independent $\text{Normal}(0,1)$ random variables, then given the values of $\widehat{\boldsymbol{q}}_i^*$ and $\widehat{\boldsymbol{\Sigma}}_i^*$, it follows that

$$\boldsymbol{W} = \widehat{\boldsymbol{q}}_i^* + \boldsymbol{Z}^*\boldsymbol{C}^\intercal \tag{12}$$

has the $\text{Normal}_{n-1}(\widehat{\boldsymbol{q}}_i^*, \widehat{\boldsymbol{\Sigma}}_i^*)$ distribution; and if we take

$$\boldsymbol{q}_i = [\boldsymbol{W}, 1 - \boldsymbol{W}\mathbf{1}_{n-1}] = \left[W_1,\ldots,W_{n-1}, 1 - \sum_{j=1}^{n-1} W_j\right], \tag{13}$$

then

$$\boldsymbol{q}_i \sim \text{Normal}_n(\widehat{\boldsymbol{q}}_i, \widehat{\boldsymbol{\Sigma}}_i) \tag{14}$$

as required. To ensure that the nonnegativity condition (3) is satisfied, which is equivalent to requiring that all components of $\boldsymbol{q}_i$ are nonnegative, we repeat the operations in Equations (12) and (13) until finally obtaining a random vector $\boldsymbol{q}_i$ satisfying $\boldsymbol{q}_i \geq \mathbf{0}_n$. This procedure is formalized in Algorithm 3 below.

Algorithm 3 Generating $\boldsymbol{q}_i \sim \text{Normal}_n(\widehat{\boldsymbol{q}}_i, \widehat{\boldsymbol{\Sigma}}_i)$

1: Set $\widehat{\boldsymbol{q}}_i^* \leftarrow$ the first $n-1$ elements of $\widehat{\boldsymbol{q}}_i$, and set $j \leftarrow 0$.
2: Set $\widehat{\boldsymbol{\Sigma}}_i^* \leftarrow$ the upper-left $(n-1) \times (n-1)$ submatrix of $\widehat{\boldsymbol{\Sigma}}_i$.
3: Set $\boldsymbol{C} \leftarrow$ the Cholesky decomposition of $\widehat{\boldsymbol{\Sigma}}_i^*$.
4: **while** $j \leq M$ **do**
5: Generate an $(n-1)$ vector of independent standard normal variates $\boldsymbol{Z}^* = [Z_1, \ldots, Z_{n-1}]$.
6: Set $\boldsymbol{W} \leftarrow \widehat{\boldsymbol{q}}_i^* + \boldsymbol{Z}^* \boldsymbol{C}^\mathsf{T}$.
7: **if** $W_i \geq 0$ for $i = 1, \ldots, n-1$ and $\sum_{i=1}^{n-1} W_i \leq 1$ **then**
8: Set $j \leftarrow j+1$ and $\boldsymbol{q}_i^{(j)} \leftarrow \left[W_1, \ldots, W_{n-1}, 1 - \sum_{i=1}^{n-1} W_i\right]$
9: **end if**
10: **end while**

Remark 2. Algorithm 3 has no warm-up period because each random vector $\boldsymbol{q}_i^{(j)}$, $j = 1, \ldots, M$ delivered by the procedure has the target (truncated) multivariate normal distribution. Moreover if N_i, $i = 1, \ldots, n$ is relatively large, then from Equation (11), we see that steps 6–8 of Algorithm 3 will be performed a relatively small number of times (usually once) on each iteration j of the procedure.

3. Case Study: Evaluating glycemic control treatment guidelines for patients with type 2 diabetes

In this section we demonstrate the use of Algorithms 2 and 3 in a case study concerning an MDP that was designed to evaluate glycemic control treatment guidelines for patients with type 2 diabetes. Glycemic control involves the regulation of blood glucose levels over time with the following goals: (i) avoiding acute daily symptoms of hyperglycemia; and (ii) preventing or delaying the development of diabetes complications associated with high blood glucose levels, including ischemic heart disease, stroke, blindness, renal failure, and amputation. Glycated hemoglobin (HbA1c), which is represented as a percentage of total hemoglobin, is commonly used as a measure of average blood glucose concentration. A patient's HbA1c level is measured using a blood test, typically at time intervals ranging from 3 to 6 months. The American Diabetes Association (ADA) recommends lowering HbA1c to a level below 7%; when a patient's HbA1c exceed 7%, glucose lowering medication is recommended, according to this guideline (American Diabetes Association 2015).

3.1. The glycemic control model

The glycemic control model, adapted from Zhang et al. (2014) for this case study, is a finite-state, finite-horizon, discrete-time MDP. The underlying Markov chain represents the natural progression of a patient's HbA1c level over time. The time horizon is discretized into a set of time epochs indexed by $t \in \mathscr{T} = \{1, 2, \ldots, T\}$, where time epoch 1 refers to the age at diagnosis and the time epoch T refers to a reasonable upper limit on a patient's age (e.g., age 100 years).

The states in the model include HbA1c states, medication states, and absorbing states. A patient's HbA1c level is discretized into a set of 10 HbA1c states, denoted by $\mathscr{L} = \{\ell(1), \ell(2), \ldots, \ell(10)\}$. The HbA1c state at time epoch $t \in \mathscr{T}$ is denoted by $\ell_t \in \mathscr{L}$. 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. Medication states record which medication(s) is (are) already initiated at each time epoch. The medication state at time epoch t is denoted by a binary vector $\boldsymbol{m}_t \in \mathscr{M} \equiv \{(m_{1,t}, m_{2,t}, \ldots, m_{k,t}) : m_{i,t} \in (0,1), \forall i = 1,2,\ldots,k\}$ where $\mathscr{M}$ represents the set of all medication states, and k represents the total number of medications. For any $i = 1,2,\ldots,k$, $m_{i,t} = 0$ represents that the patient is not on medication i at time t, otherwise, $m_{i,t} = 1$. Absorbing states in the model include major diabetes-related events, and are denoted by $d \in \mathscr{D} = \{D^{\mathrm{O}}, E^{\mathrm{macro}}, E^{\mathrm{micro}}\}$, where D^{O} denotes death from other causes, E^{macro} denotes fatal or nonfatal macrovascular events including ischemic heart disease, myocardial infarction, congestive heart failure and stroke, and E^{micro} denotes fatal or nonfatal microvascular events including blindness, renal failure and amputation. The complete set of states in our model is given by $\mathscr{S} = \{\mathscr{L} \times \mathscr{M}\} \bigcup \mathscr{D}$.

At each time epoch t, the action defines which medication(s) to use over the period $(t, t+1]$. We denote the action at time epoch t to be a binary vector $\boldsymbol{\alpha}_t = \{(A_{1,t}(\ell_t, \boldsymbol{m}_t), A_{2,t}(\ell_t, \boldsymbol{m}_t), \ldots, A_{k,t}(\ell_t, \boldsymbol{m}_t)) : A_{i,t}(\ell_t, \boldsymbol{m}_t) = \{0,1\}, i = 1,2,\ldots,k\}$, where $A_{i,t}(\ell_t, \boldsymbol{m}_t) = 1$ represents use of medication i during time period $(t, t+1]$; otherwise $A_{i,t}(\ell_t, \boldsymbol{m}_t) = 0$. We use the model to evaluate the ADA's current treatment guideline (American Diabetes Association 2015) and its consensus algorithm for treating hyperglycemia for type 2 diabetes (Nathan et al. 2009) by setting actions in the model according to the following decision rule: metformin is used as the first-line medication when a patient's HbA1c reaches 7%, sulfonylurea is added as a second-line agent when the HbA1c level reaches 7% again, and if the patient fails to maintain the glycemic control goal of HbA1c $< 7\%$ by using metformin and sulfonylurea combination therapy, then insulin is initiated in place of sulfonylurea and is used for the remainder of the time horizon. Transitions among medication states occur deterministically based on the action taken at the beginning of the time epoch. For any time epoch $t \in \mathscr{T}$, the medication state, $\boldsymbol{m}_t$, is updated as follows:

$$m_{i,t+1} = \begin{cases} 1, & \text{if } A_{i,t}(\ell_t, \boldsymbol{m}_t) = 1, \\ 0, & \text{if } A_{i,t}(\ell_t, \boldsymbol{m}_t) = 0, \end{cases} \tag{15}$$

$\forall\, i = 1,\ldots,k$, $t \in \mathscr{T}$. Medication initiation or intensification results in a proportional decrease in the mean HbA1c level based on the medication effects shown in Table 1. We assume that after insulin is initiated, HbA1c will be maintained at 7%.

The model includes two types of transition probabilities: transition probabilities from HbA1c states to absorbing states, and transition probabilities among HbA1c states. Given that the patient is in the HbA1c state ℓ_t and medication state $\boldsymbol{m}_t$, and takes action $\boldsymbol{\alpha}_t$ at decision epoch t, the probability of transitioning into an absorbing state $d \in \mathscr{D}$ during the period $(t, t+1]$ is denoted by $p_t(d|\ell_t, \boldsymbol{m}_t, \boldsymbol{\alpha}_t)$, where

$$p_t(d|\ell_t, \boldsymbol{m}_t, \boldsymbol{\alpha}_t) = \begin{cases} \pi_t^{\mathrm{O}}, & \text{if } d = D^{\mathrm{O}}, \\ \pi_t^{\mathrm{macro}}(\ell_t, \boldsymbol{m}_t, \boldsymbol{\alpha}_t), & \text{if } d = E^{\mathrm{macro}}, \\ \pi_t^{\mathrm{micro}}(\ell_t, \boldsymbol{m}_t, \boldsymbol{\alpha}_t), & \text{if } d = E^{\mathrm{micro}}, \end{cases} \tag{16}$$

and $p_t(d|d) = 1$ for all $t \in \mathscr{T}$ and $d \in \mathscr{D}$. A macrovascular event (fatal and nonfatal) occurs with probability $\pi_t^{\text{macro}}(\ell_t, \boldsymbol{m}_t, \boldsymbol{\alpha}_t)$, and a microvascular event (fatal and nonfatal) occurs with probability $\pi_t^{\text{micro}}(\ell_t, \boldsymbol{m}_t, \boldsymbol{\alpha}_t)$. These probabilities were determined by the patient's age, gender, ethnicity, smoking status, body mass index, HbA1c level, systolic blood pressure, total cholesterol, and high-density lipoprotein (HDL) cholesterol using the United Kingdom Prospective Diabetes Study (UKPDS) Outcomes Model, which is a proportional hazards rate model (Clarke et al. 2004). Death from other causes occurs with probability π_t^{O}, and this probability was estimated from the Centers for Disease Control and Prevention (CDC) mortality tables (CDC, 2013a). Moreover, we define the total probability of transitioning into absorbing states by

$$p_t^{\mathscr{D}}(\ell_t, \boldsymbol{m}_t, \boldsymbol{\alpha}_t) = \sum_{d \in \mathscr{D}} p_t(d|\ell_t, \boldsymbol{m}_t, \boldsymbol{\alpha}_t). \tag{17}$$

Note that in the model, the medication state and the action affect the transition probabilities of transitioning into an absorbing state by altering a patient's HbA1c trajectory over time based on medication effects.

After diagnosis of diabetes, patients transition among HbA1c states every three months until they transition into an absorbing state or reach age 100. We denote the matrix of transition probabilities among HbA1c states conditional on not transitioning into an absorbing state by

$$P^{\text{HbA1c}} = [p_{\ell_t}^{\text{HbA1c}}(\ell_{t+1})]_{10\times 10}, \tag{18}$$

where $p_{\ell_t}^{\text{HbA1c}}(\ell_{t+1})$ represents the probability of transitioning from HbA1c state ℓ_t to HbA1c state ℓ_{t+1} for any $\ell_t, \ell_{t+1} \in \mathscr{L}$. The MLE of the HbA1c TPM and the uncertainty set generated based on Equation (6) for each HbA1c transition probability (shown in Appendix Tables A1 and A2) were estimated from a retrospective administrative claims data set that included medical claims, pharmacy claims, and laboratory data from a large national U.S. health plan. To qualify for inclusion in the case study, a patient must satisfy the following criteria: (i) be 40 years of age or older; (ii) have been diagnosed with type 2 diabetes between 1995 and 2010; (iii) have received the first prescription for noninsulin glucose-lowering medication at least 6 months after enrollment; and (iv) have accumulated at least 15 HbA1c records within 5 years of continuous enrollment, along with complete pharmacy claims data. Altogether 377 males and 272 females were included in the case study.

We define the probability of a patient being in state $s_{t+1} \in \mathscr{S}$ at epoch $t+1$ given the patient is in state $s_t \in \mathscr{S}$ and takes action $\boldsymbol{\alpha}_t \in \mathscr{A}_t(s_t)$ at epoch t as follows:

$$p_t(s_{t+1}|s_t, \boldsymbol{\alpha}_t) = \begin{cases} (1 - p_t^{\mathscr{D}}(\ell_t, \boldsymbol{m}_t, \boldsymbol{\alpha}_t)) p_{\ell_t}^{\text{HbA1c}}(\ell_{t+1}), & \text{if } s_t = (\ell_t, \boldsymbol{m}_t), s_{t+1} = (\ell_{t+1}, \boldsymbol{m}_{t+1}) \in \mathscr{L} \times \mathscr{M}, \\ p_t(s_{t+1}|\ell_t, \boldsymbol{m}_t, \boldsymbol{\alpha}_t), & \text{if } s_t = (\ell_t, \boldsymbol{m}_t) \in \mathscr{L} \times \mathscr{M}, s_{t+1} \in \mathscr{D}, \\ 1, & \text{if } s_t, s_{t+1} \in \mathscr{D}, \\ 0, & \text{otherwise.} \end{cases} \tag{19}$$

We consider the following two types of one-period reward functions during the time period $(t, t+1]$, $\forall t \in \mathscr{T} \backslash \{T\}$: (i) the medication cost, $r_t^C(\ell_t, \boldsymbol{m}_t, \boldsymbol{\alpha}_t)$, accumulated during the latest three-month period; and

Table 1. Parameter estimates of the glycemic control model for the base case analysis and sensitivity analyses.

Parameter (reference)	Base case value (range)
Patient's characteristics	
Diagnosis age (Centers for Disease Control and Prevention 2013)	Females: 55.2; Males 53.6
Concurrent comorbidity at diagnosis *	No
Blood pressure (mmHg) (American Diabetes Association 2015) †	140 mmHg
Total cholesterol (mg/dL) (NCEP 2002) †	200 mg/dL
HDL (mg/dL) (NCEP 2002) †	40 mg/dL
Body mass index (kg/m^2) (Kramer et al. 2010)	32.6
Glycemic control goal % (American Diabetes Association 2015)	7%
Medication effect on HbA1c (Zhang et al. 2014)	
Metformin	0.0661 (0.0620–0.0703)
Sulfonylurea	0.0937 (0.0852–0.1022)
Insulin	Maintain HbA1c at 7%
Monthly medication cost ($, 2013 U.S. dollars) (Bennett et al. 2011, Yeaw et al. 2012)	
Metformin	81.75 (25.87–181.09)
Sulfonylurea	54.85 (9.31–165.57)
Insulin	245.70 (189.39–327.54)
Medication disutilities (Zhang et al. 2014)	
Metformin	-0.0002 (±25%)
Sulfonylurea	-0.0095 (±25%)
Insulin	-0.0206 (±25%)

* Concurrent comorbidities include peripheral vascular disease, atrial fibrillation, ischemic heart disease, congestive heart failure, and blindness.

† Patients blood pressure, total cholesterol, and HDL were assumed to be well controlled by anti-hypertension and anti-hyperlipidemia medications.

(ii) the QALYs, $r_t^Q(\ell_t, \boldsymbol{m}_t, \boldsymbol{\alpha}_t)$, accumulated during the same period. In the case study, the medication cost is calculated as follows:

$$r_t^C(\ell_t, \boldsymbol{m}_t, \boldsymbol{\alpha}_t) = \sum_{i=1}^{k} 3 \times A_{i,t}(\ell_t, \boldsymbol{m}_t) c_i \tag{20}$$

where c_i is the monthly medication cost of medication i. The cost estimates are given in Table 1. QALYs, a widely used measure for evaluating the health outcomes for treatment and health policy decisions (Gold et al. 2002), represent a patient's survival duration weighted by the quality of health that the patient experiences over that time interval. Each year of life is assigned a value between 0 (death) and 1 (perfect health). Owing to factors such as the burden of treatment and the occurrence of adverse health events, QALYs may be less

than 1. In the case study, QALYs are calculated as follows:

$$r_t^Q(\ell_t, \boldsymbol{m}_t, \boldsymbol{\alpha}_t) = 0.25\left(1 - \sum_{i=1}^{k} A_{i,t}(\ell_t, \boldsymbol{m}_t) D_i\right) \tag{21}$$

where D_i is the disutility associated with taking medication i. The estimates of medication disutilities are given in Table 1. The coefficient of 0.25 is used to calculate QALYs since our case study involves 3-month time periods, though in general the time (in years) between epochs t and $t+1$ can be used in the calculation. The magnitude of the medication disutility depends on documented side effects that a patient might experience, including low blood sugar (hypoglycemia) and weight gain.

We let $v_t^Q(s_t)$ denote the total expected QALYs from time t onward given the patient is in state s_t at time epoch t, and we calculate it as follows:

$$v_t^Q(s_t) = \begin{cases} r_t^Q(\ell_t, \boldsymbol{m}_t, \boldsymbol{\alpha}_t) + \left(1 - p_t^{\mathscr{D}}(\ell_t, \boldsymbol{m}_t, \boldsymbol{\alpha}_t)\right) \sum\limits_{\ell_{t+1} \in \mathscr{L}} p_{\ell_t}^{\text{HbA1c}}(\ell_{t+1}) v_{t+1}^Q(\ell_{t+1}, \boldsymbol{m}_{t+1}), \\ \qquad\qquad \forall s_t \in \mathscr{L} \times \mathscr{M}, t \in \mathscr{T} \backslash \{T\}, \\ R_T^Q(\ell_T, \boldsymbol{m}_T), \qquad\qquad \forall s_T \in \mathscr{L} \times \mathscr{M}, t = T, \\ 0, \qquad\qquad s_t \in \mathscr{D}, t \in \mathscr{T}, \end{cases} \tag{22}$$

for any $\ell_t \in \mathscr{L}, t \in \mathscr{T}$, and $R_T^Q(\ell_T, \boldsymbol{m}_T)$ denotes the end-of-horizon reward which represents the life expectancy after 100 years which were set to be 2.24 for females and 2.05 for males based on a 2008 U.S. life table (Arias 2011).

We let $v_t^C(\ell_t, \boldsymbol{m}_t)$ denote the total expected medical cost from time t onward given the patient is in HbA1c state $\ell_t \in \mathscr{L}$, and medication state $\boldsymbol{m}_t$ at time epoch t, and we calculate it as follows:

$$v_t^C(s_t) = \begin{cases} r_t^C(\ell_t, \boldsymbol{m}_t, \boldsymbol{\alpha}_t) + \lambda\left[1 - p_t^{\mathscr{D}}(\ell_t, \boldsymbol{m}_t, \boldsymbol{\alpha}_t)\right] \sum\limits_{\ell_{t+1} \in \mathscr{L}} p_{\ell_t}^{\text{HbA1c}}(\ell_{t+1}) v_{t+1}^C(\ell_{t+1}, \boldsymbol{m}_{t+1}), \\ \qquad\qquad \forall s_t \in \mathscr{L} \times \mathscr{M}, t \in \mathscr{T} \backslash \{T\}, \\ \left(\sum\limits_{i=1}^{k} 12\, c_i\, m_{i,T}\right) \times R_T^Q(\ell_T, \boldsymbol{m}_T), \qquad \forall \forall s_T \in \mathscr{L} \times \mathscr{M}, t = T, \\ 0, \qquad\qquad s_t \in \mathscr{D}, t \in \mathscr{T}, \end{cases} \tag{23}$$

for any $\ell_t \in \mathscr{L}, t \in \mathscr{T}$ where λ represent the discount factor. In the numerical results, costs were discounted at an annual rate of 3%. We used the finite-horizon policy evaluation algorithm (see p. 80 of Puterman (2005)) to calculate $v_t^Q(s_t)$ and $v_t^C(s_t)$ for all $t \in \mathscr{T}$. In the finite-horizon policy evaluation algorithm, we indexed metformin, sulfonylurea, and insulin to be medication 1, 2, and 3. The action $\boldsymbol{\alpha}_t$ is updated as follows: (1) if the mean HbA1c value for HbA1c state ℓ_t at time epoch t is $\leq 7\%$, then $A_{i,t}(\ell_t, \boldsymbol{m}_t) = m_{i,t}, \forall i = 1,2,3$; (2) if the mean HbA1c value for HbA1c state ℓ_t at time epoch t is above 7%, and $\boldsymbol{m}_t = (0,0,0)$, then $A_{1,t}(\ell_t, \boldsymbol{m}_t) = 1$, and $A_{2,t}(\ell_t, \boldsymbol{m}_t) = A_{3,t}(\ell_t, \boldsymbol{m}_t) = 0$; (3) if the mean HbA1c value for HbA1c state ℓ_t at time epoch t is above 7%, and $\boldsymbol{m}_t = (1,0,0)$, then $A_{1,t}(\ell_t, \boldsymbol{m}_t) = A_{2,t}(\ell_t, \boldsymbol{m}_t) = 1$, and $A_{3,t}(\ell_t, \boldsymbol{m}_t) = 0$; 4) if the mean HbA1c value for HbA1c state ℓ_t at time epoch t is above 7%, and $\boldsymbol{m}_t = (1,1,0)$, then $A_{1,t}(\ell_t, \boldsymbol{m}_t) = 1$, $A_{2,t}(\ell_t, \boldsymbol{m}_t) = 0$, and $A_{3,t}(\ell_t, \boldsymbol{m}_t) = 1$. All other medication states including $(0,1,0), (0,0,1), (0,1,1)$, and $(1,1,1)$ are not permitted based on clinical treatment guidelines.

As described above, the value functions of the model include the expected QALYs and the total medication costs accumulated over the entire time horizon. Given the initial HbA1c distribution $\mathbf{\Pi} = (\Pi(\ell(1)), \ldots, \Pi(\ell(10)))$, estimated from the aforementioned retrospective administrative claims data set, the total expected QALYs accumulated over the entire time horizon, denoted by v^Q, can be calculated as follows:

$$v^Q = \sum_{i=1}^{10} \Pi(\ell(i)) v_1^Q(\ell(i), \mathbf{0}), \tag{24}$$

and the total medication costs accumulated over the entire time horizon, denoted by v^C, can be calculated as follows:

$$v^C = \sum_{i=1}^{10} \Pi(\ell(i)) v_1^C(\ell(i), \mathbf{0}), \tag{25}$$

where $\mathbf{0}$ is a vector of zeros which represents that the patient is not on any medication at the initial time epoch, given that we focus on estimating QALYs and cost for patients newly diagnosed with type 2 diabetes.

3.2. Computation times

For illustrating the efficiency of the proposed TPM sampling algorithms, we collected the computation times for sampling 1000 HbA1c TPMs, (i.e., the computation time used to execute Steps 4-6 of Algorithm 1 where M (the number of samples) was set to be 1000. A computer with an Intel Core Duo CPU (2.26 GHz) CPU and 3.0 GB RAM was used to carry out all the simulation experiments.

The warm-up period for Algorithm 2, estimated based on Equation (10) is conservative, yielding a warm-up period of 23,579 steps based on the values $n = 10$, $\delta = 0.01$, and $\gamma = 1$. For applications of Markov Chain Monte Carlo (MCMC), Geyer (1992) recommends that a warm-up period not exceeding 1% or 2% of the run length is sufficient to yield acceptable accuracy in MCMC-based estimators; and more recently Geyer (2011) recommends using no warm-up period in MCMC applications. In our experience with a broad range of stochastic simulation applications, we have found that in well-designed simulation experiments the warm-up period rarely exceeds 20% of the total simulation run length (Lada et al. 2006, 2007, Lada and Wilson 2006, 2008, Steiger and Wilson 2002, Steiger et al. 2005, Tafazzoli and Wilson 2011, Tafazzoli et al. 2011a,b). On the basis of all these considerations and the structure of Algorithm 2, we set the warm-up period for Algorithm 2 to be 50 steps (i.e., 5% of the sample size beyond the warm-up period) to ensure approximate uniformity in the distribution of TPMs sampled beyond the warm-up period.

Table 2 shows the mean and standard deviation of the computation time of sampling 1000 HbA1c TPMs based on 100 replications for each gender. On average, both algorithms finish sampling 1000 HbA1c TPMs within 2 seconds. Algorithm 3 is slightly faster than Algorithm 2; this is reasonable because Algorithm 3 does not have a warm-up period.

Table 2. Computational time comparison between Algorithms 2 and 3.

Algorithms	Mean (Standard Deviation) of the computation time (in seconds) for sampling 1000 TPMs	
	Males	Females
Algorithm 2	1.90 (0.02)	1.90 (0.02)
Algorithm 3	1.24 (0.04)	1.21 (0.05)

3.3. Numerical Results

We conduct PSA on the HbA1c TPM using the procedure shown in Algorithm 1. As shown in Equation (18), the HbA1c TPM, P^{HbA1c}, is independent of time epochs and actions, therefore, for generating one value function estimate, we only need to sample one HbA1c TPM. In other words, Steps 2-3 in Algorithm 1 can be skipped. To implement Step 5, we used the proposed TPM sampling Algorithms 2 and 3, and to implement Step 6, we used the finite-horizon policy evaluation algorithm (see p. 80 of Puterman (2005)).

In the base case, we used the MLE of the HbA1c TPM. QALY results are presented as years from birth until diagnosis of diabetes plus v^{Q}, i.e., the expected QALYs from diagnosis to the first occurrence of a diabetes-related adverse health event or death. The base case QALYs are 64.40 QALYs for males and 68.42 QALYs for females. The expected total discounted medication costs from diagnosis to the first occurrence of a diabetes-related adverse health event or death is $28,337 for males and $34,293 for females.

Table 3. Minimum, mean, and maximum values of the estimated expected QALYs and the estimated expected total medication costs from diagnosis to the first occurrence of a diabetes-related adverse health event or death based on 1000 TPMs sampled by Algorithm 2 and 3.

		Males			Females		
		Min.	Mean	Max.	Min.	Mean	Max.
(A) TPMs sampled by Algorithm 2	Expected QALYs	64.36	64.40	64.43	68.38	68.42	68.46
	Expected total discounted medication costs ($)	27,483	28,351	29,527	33,137	34,355	35,406
(B) TPMs sampled by Algorithm 3	Expected QALYs	64.37	64.40	64.45	68.38	68.42	68.49
	Expected total discounted medication costs ($)	26,846	28,290	29,266	31,918	34,212	35,397

We first looked at the impact of uncertainty in the TPM on single model outcomes. Table 3 provides two sets of PSA results including the minimum, mean, and maximum values of the estimates of the two value functions (i.e., the expected QALYs and the expected total medication cost) based on 1000 sampled TPMs using the two TPM sampling methods. We define the variation in the value function estimates as the difference between its maximum and minimum simulation-generated values. Results based on Algorithm 2 show that the maximum variation in the expected QALYs is 0.07 QALYs for males and 0.07 QALYs for

females, and the variation in the expected total medication costs is $2,242 for males and $2,269 for females. Results based on Algorithm 3 shows that the variation in the expected QALYs is 0.08 QALYs for males and 0.11 QALYs for females, and the variation in the expected total medication cost is $2,420 for males and $3,478 for females.

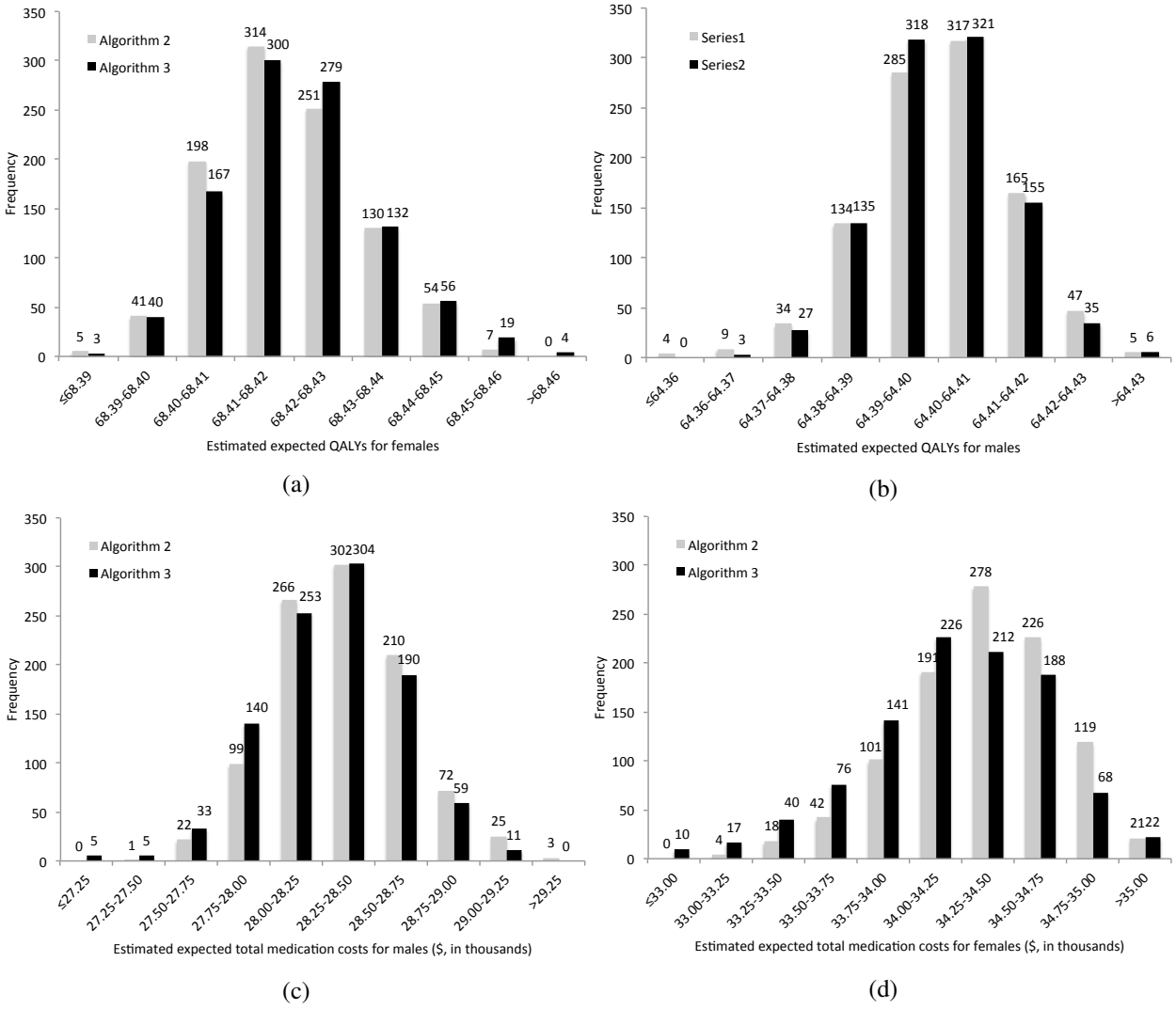

Fig. 1. Histograms of the estimated expected QALYs and the estimated expected total medication costs based on 1000 sampled TPMs: (a) the estimated expected QALYs for males, (b) the estimated expected QALYs for females, (c) the estimated expected total medication costs for males, and (d) the estimated expected total medication costs for females.

Figure 1 shows the histograms of the estimated expected QALYs and the estimated expected total medication costs based on 1000 sampled TPMs. Algorithm 2 samples TPMs uniformly over the uncertainty set while Algorithm 3 samples TPMs according to a truncated normal distribution. We find that in this case study the marginal distributions of the value function estimates are approximately bell-shaped regardless of gender and the PSA methods we used.

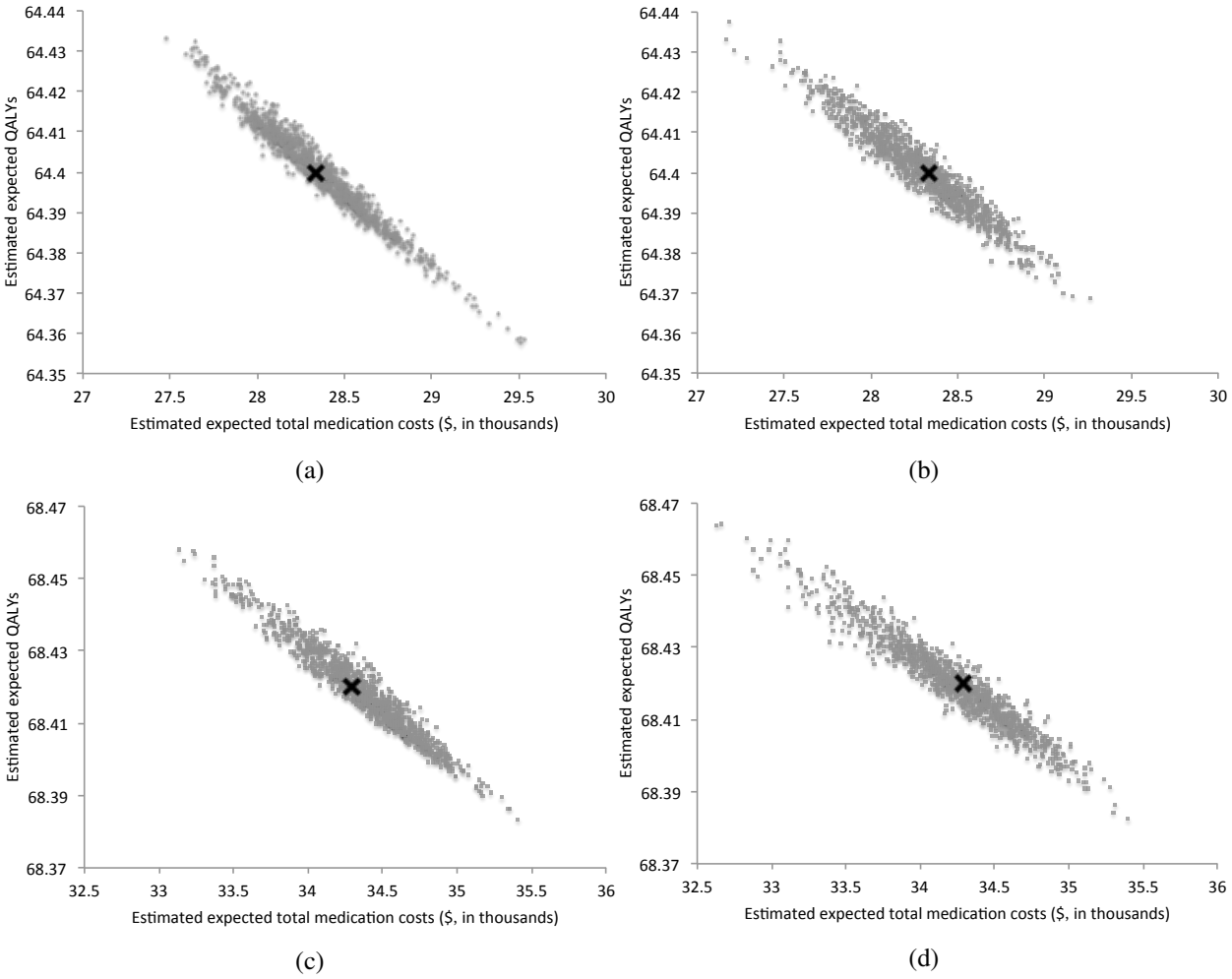

Fig. 2. Scatterplot of the estimated expected QALYs versus the estimated expected total medication costs based on 1000 sampled TPMs: (a) using Algorithm 2 for males, (b) using Algorithm 3 for males, (c) using Algorithm 2 for females, and (d) using Algorithm 3 for females. The base-case results are plotted with a black $\times$.

We also look at the impact of the uncertainty in the TPM on the joint distribution of the two value functions. Figure 2 provides scatter plots of the estimated expected QALYs versus the estimated expected total medication costs for males and females using the two PSA methods. We find that in this case study the two value functions are strongly correlated.

We present tornado diagrams in Figures 3 and 4 to compare sensitivities of the expected QALYs and the expected total medication costs on the HbA1c TPM and other model parameters including medication disutilities, medication effect on HbA1c, and monthly medication costs. The variations in the estimated expected QALYs and the estimated total medication costs, corresponding to other model parameters including medication disutilities, medication effect on HbA1c, and monthly medication costs, are generated by setting each parameter to its lower and upper bounds provided in Table 1. We find that the medication disuility and the HbA1c TPM are the two factors that have the greatest impact on the expected QALYs. For example,

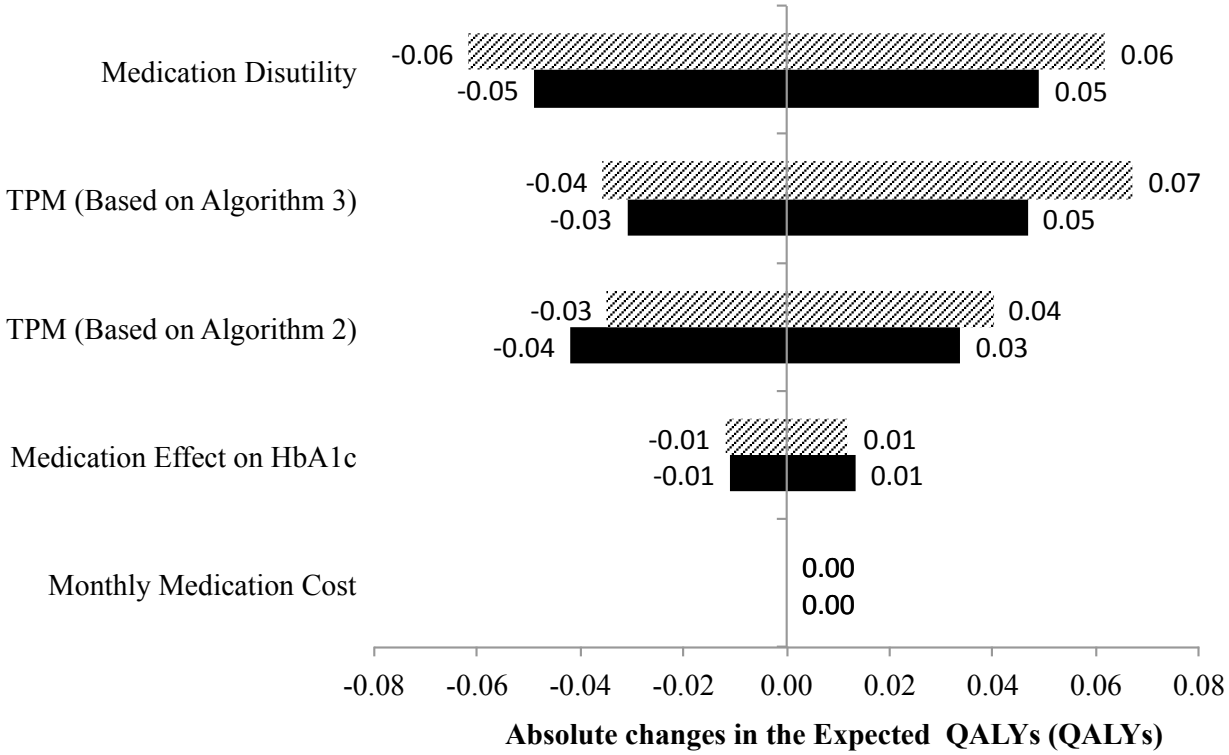

Fig. 3. Sensitivity analysis on the expected QALYs. The x-axis represents the absolution change in the expected QALYs from the base case results (64.40 QALYs for males and 68.42 QALYs for females). The y-axis represents the model parameters. The solid bar represents males, and the hatched bar represents females.

based on Algorithm 3, the absolute changes in the estimated expected QALYs vary from -0.04 QALYs to 0.07 QALYs for females, and -0.03 QALYs to 0.05 QALYs for males. The monthly medication cost has the greatest impact on the expected total medication costs. Compared to the base-case estimates of the expected total medication costs, the relative changes in the estimated expected total medication costs vary from -35.5% to 60.4% for females and -37.1% to 62% for males. The expected total medication costs is also sensitive to the HbA1c TPM. For example, using the TPM sampling Algorithm 2, the relative changes in the estimated expected total medication costs vary from -3.4% to 3.2% for females and -3.0% to 4.2% for males.

4. Conclusions

Algorithms 2 and 3 provide two TPM sampling methods for conducting PSA on the TPM of MDPs. Algorithm 2 can be used for cases in which (i) there is relatively little information for specifying the uncertainty set; or (ii) the decision maker wants to define the uncertainty set subjectively based on expert opinion. Algorithm 3 can be used for cases in which interest is focused on the uncertainty arising from statistical estimation

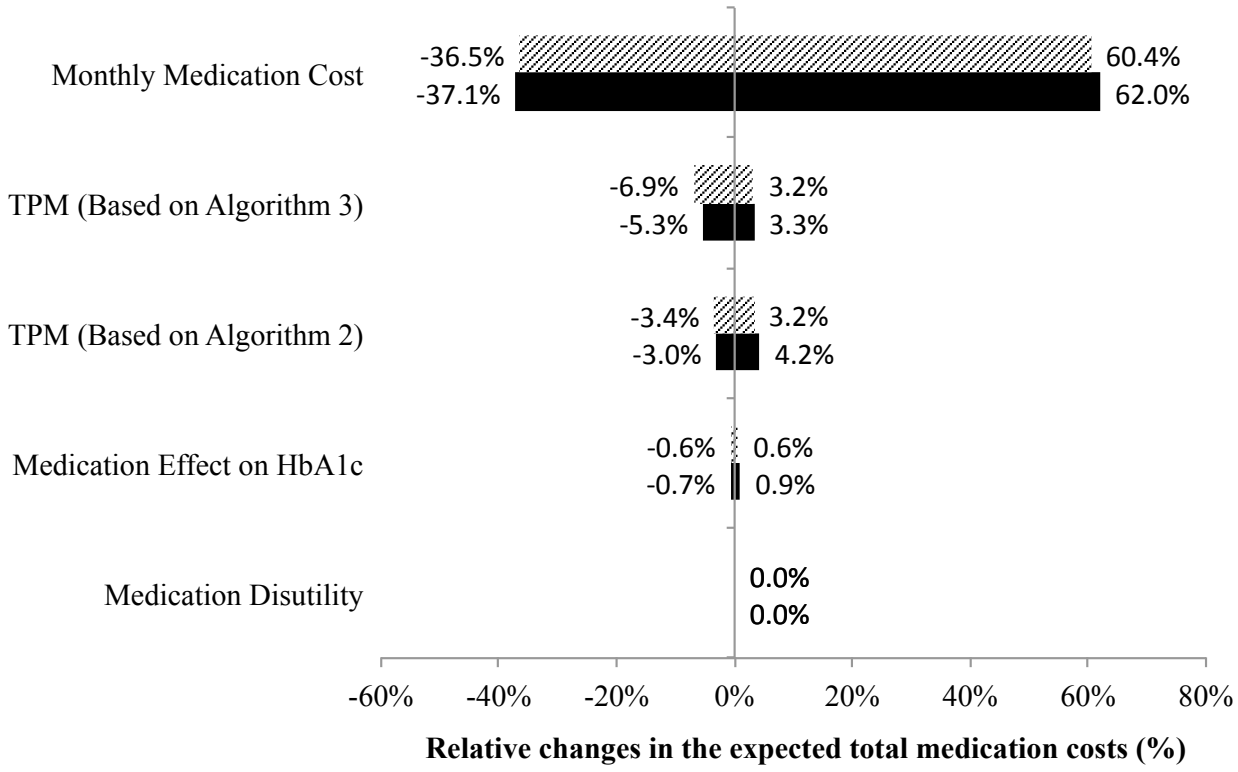

Fig. 4. Sensitivity analysis on the expected total medication costs. The x-axis represents the relative change (%) in the expected total medication cost with respect to the base case results ($28,337 for males and $34,293 for females). The y-axis represents the model parameters. The solid bar represents males, and the hatched bar represents females.

of the TPM. Both methods are easy to implement and can sample TPMs efficiently, as we demonstrate with a practical problem in the context of diabetes treatment.

The case study presented in Section 3 shows that performing PSA can provide useful insights into the reliability and generality of the conclusions derived from an MDP. By using Algorithms 2 and 3, we find that the two value functions (i.e., the expected QALYs and the expected total medication costs) of the glycemic control model are sensitive to the uncertainty in the associated HbA1c TPM for males and females patients.

There are some limitations in this work that provide opportunities for future study. First, Algorithms 2 and 3 are based on the assumption that the rows of the TPM are independent; but in some applications there could be correlations among transition probabilities in different rows, for example, when some of those transition probabilities are based on expert opinion. Second, in the case of Algorithm 2, the uncertainty set is assumed to be convex. If this condition is not satisfied, then Algorithm 2 may generate infeasible results. However, the assumption of convexity is consistent with most commonly used representations of confidence regions or uncertainty sets, such as the regions we consider in this article.

These limitations notwithstanding, the methods that we have proposed and applied demonstrate the ease

with which PSA can be conducted to establish the robustness of findings derived from MDPs.

Acknowledgment

The work of the third author (B.T.D.) was supported in part by the National Science Foundation under grant number CMMI-0969885.

References

American Diabetes Association. 2015. Standards of medical care in diabetes–2015: Summary of revisions. *Diabetes Care* **38**(Suppl.1) S4.

Anderson, T.W. 2003. *An Introduction to Multivariate Statistical Analysis*. 3rd ed. John Wiley & Sons, Hoboken, NJ.

Anderson, T.W., L.A. Goodman. 1957. Statistical inference about Markov chains. *The Annals of Mathematical Statistics* **28**(1) 89–110.

Arias, Elizabeth. 2011. 2007 United States life tables. *National Vital Statistics Reports* **59**(9).

Bennett, W. L., N. M. Maruthur, S. Singh, J. B. Segal, L. M. Wilson, R. Chatterjee, S. S. Marinopoulos, M. A. Puhan, P. Ranasinghe, L. Block, W. K. Nicholson, S. Hutfless, E. B. Bass, S. Bolen. 2011. Comparative effectiveness and safety of medications for type 2 diabetes: An update including new drugs and 2-drug combinations. *Annals of Internal Medicine* **154**(9) 602–13.

Briggs, Andrew H., Milton C. Weinstein, Elisabeth A. L. Fenwick, Jonathan Karnon, Mark J. Sculpher, A. David Paltiel. 2012. Model parameter estimation and uncertainty analysis: A report of the ISPOR-SMDM modeling good research practices task force–6. *Medical Decision Making* **32**(5) 722–732.

Centers for Disease Control and Prevention. 2013. Age at diagnosis of diabetes among adult incident cases aged 18–79 years. URL http://www.cdc.gov/diabetes/statistics/incidence_national.htm.

Cheng, R.C.H. 2013. Fitting statistical models of random search in simulation studies. *ACM Transactions on Modeling and Computer Simulation* **23**(3) article 15.

Chia, Y.L., P.W. Glynn. 2013. Limit theorems for simulation-based optimization via random search. *ACM Transactions on Modeling and Computer Simulation* **23**(3) article 16.

Clarke, P.M., A.M. Gray, A. Briggs, A.J. Farmer, P. Fenn, R.J. Stevens, D.R. Matthews, I.M. Stratton, R.R. Holman. 2004. A model to estimate the lifetime health outcomes of patients with type 2 diabetes: The United Kingdom Prospective Diabetes Study (UKPDS) outcomes model (UKPDS No. 68). *Diabetologia* **47** 17471759.

Craig, A., P.P. Sendi. 2002. Estimation of the transition matrix of a discrete-time Markov chain. *Health Economics* **11** 33–42.

Devroye, L. 1978. The uniform convergence of nearest neighbor regression function estimators and their application in optimization. *IEEE Transactions on Information Theory* **IT-24**(2) 142–151.

Devroye, L. 1986. *Non-uniform random variate generation*. Springer-Verlag, New York.

Doubilet, P., C. B. Begg, M. C. Weinstein, P. Braun, B. J. McNeil. 1985. Probabilistic sensitivity analysis using Monte Carlo simulation: A practical approach. *Medical Decision Making* **5**(2) 157–177. URL http://europepmc.org/abstract/MED/3831638.

Draper, N.R., H. Smith. 1998. *Applied Regression Analysis*. 3rd ed. John Wiley & Sons, New York.

Geyer, C.J. 1992. Practical Markov chain Monte Carlo. *Statistical Science* **7**(4) 473–483.

Geyer, C.J. 2011. *Introduction to Markov Chain Monte Carlo*. CRC Press, Boca Raton, FL.

Gold, M.R., D. Stevenson, D.G. Fryback. 2002. HALYs and QALYs and DALYs, Oh My: Similarities an differences in summary measures in population health. *Annual Review of Public Health* **23** 115–134.

Gold, R.Z. 1963. Tests auxiliary to χ^2 tests in a Markov chain. *Annals of Mathematical statistics* **34**(1) 56–74.

Hörmann, W., J. Leydold, G. Derflinger. 2004. *Automatic Nonuniform Random Variate Generation*. Springer-Verlag, Berlin.

Horn, R.A., C.R. Johnson. 2013. *Matrix Analysis*. 2nd ed. Cambridge University Press, New York.

Kramer, H., G. Cao, L. Dugas, A. Luke, R. Cooper, R. Durazo-Arvizu. 2010. Increasing BMI and waist circumference and prevalence of obesity among adults with type 2 diabetes: The National Health and Nutrition Examination Surveys. *J Diabetes Complications* **24**(6) 368–74.

Lada, E. K., N. M. Steiger, J. R. Wilson. 2006. Performance evaluation of recent procedures for steady-state simulation analysis. *IIE Transactions* **38** 711–727.

Lada, E. K., J. R. Wilson. 2006. A wavelet-based spectral procedure for steady-state simulation analysis. *European Journal of Operational Research* **174** 1769–1801.

Lada, E. K., J. R. Wilson. 2008. Sbatch: A spaced batch means procedure for steady-state simulation analysis. *Journal of Simulation* **2** 170–185.

Lada, E. K., J. R. Wilson, N. M. Steiger, J. A. Joines. 2007. Performance of a waveletbased spectral procedure for steady-state simulation analysis. *INFORMS Journal on Computing* **19**(2) 150–160.

Mannor, S., D. Simester, P. Sun, J.N. Tsitsiklis. 2007. Bias and variance approximation in value function estimates. *Management Science* **53**(2) 308–322.

Nathan, D.M., J.M. Buse, M.B Davidson. 2009. Medical management of hyperglycemia in type 2 diabetes: A consensus algorithm for the initiation and adjustment of therapy: Aconsensus statement of the American Diabetes Association and the European Association of the Study of Diabetes. *Diabetes Care* **32**(1) 193–203.

NCEP. 2002. Third report of the National Cholesterol Education Program (NCEP) Expert Panel on detection, evaluation, and treatment of high blood cholesterol in adults (Adult Treatment Panel III) final report. *Circulation* **106**(25) 3143–3421.

Puterman, M.L. 2005. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. Wiley, New York.

Saltelli, A., K. Chan, E.M. Scott. 2000. *Sensitivity Analysis*. Wiley, New York.

Smith, R.L. 1984. Efficient monte-carlo procedures for generating points uniformly distributed over bounded regions. *Operations Research* **32**(6) 1296–1308.

Steiger, N. M., E. K. Lada, J. R. Wilson, J. A. Joines, C. Alexopoulos, D. Goldsman. 2005. Asap3: A batch means procedure for steady-state simulation output analysis. *ACM Transactions on Modeling and Computer Simulation* **15**(1) 39–73.

Steiger, N. M., J. R. Wilson. 2002. An improved batch means procedure for simulation output analysis. *Management Science* **48**(12) 1569–1586.

Tafazzoli, A., N. M. Steiger, J. R. Wilson. 2011a. N-skart: A nonsequential skewness- and autoregression-adjusted batch-means procedure for simulation analysis. *IEEE Transactions on Automatic Control* **56**(2) 254–264.

Tafazzoli, A., J. R. Wilson, E. K. Lada, N. M. Steiger. 2011b. Performance of skart: A skewness- and autoregression-adjusted batch-means procedure for simulation analysis. *INFORMS Journal on Computing* **23**(2) 297–314.

Tafazzoli, A., J.R. Wilson. 2011. Skart: A skewness- and autoregression-adjusted batch-means procedure for simulation analysis. *IIE Transactions* **43**(2) 110–128.

The MathWorks Inc. 2013. Matlab and simulink version r2013a.

Yeaw, J., W.C. Lee, M. Aagren, T. Christensen. 2012. Cost of self-monitoring of blood glucose in the united states among patients on an insulin regimen for diabetes. *Journal of Managed Care Pharmacy* **18**(1) 21–32.

Zhang, Y., R.G. McCoy, J.E. Mason, S.A. Smith, N.D. Shah, B.T. Denton. 2014. Second-line agents for glycemic control for type 2 diabetes: Are newer agents better? *Diabetes Care* **37**(5) 1338–1345.

Appendix

Table A1. The MLE, sample sizes, and lower/upper bounds of transition probabilities' 99% overall confidence intervals for male patients.

Starting state	Ending state	Sample size	MLE	Lower bound	Upper bound	Starting state	Ending state	Sample size	MLE	Lower bound	Upper bound
1	1	30	0.6667	0.5974	1	6	1	0	0	0	0
1	2	11	0.2444	0.1813	0.5934	6	2	0	0	0	0
1	3	4	0.0889	0.0471	0.32	6	3	2	0.08	0.0265	0.3755
1	4	0	0	0	0	6	4	2	0.08	0.0265	0.3755
1	5	0	0	0	0	6	5	5	0.2	0.1211	0.6357
1	6	0	0	0	0	6	6	9	0.36	0.2654	0.8829
1	7	0	0	0	0	6	7	5	0.2	0.1211	0.6357
1	8	0	0	0	0	6	8	1	0.04	0.0014	0.2535
1	9	0	0	0	0	6	9	0	0	0	0
1	10	0	0	0	0	6	10	1	0.04	0.0014	0.2535
2	1	7	0.1346	0.088	0.3924	7	1	3	0.1071	0.0495	0.4255
2	2	26	0.5	0.4317	0.8777	7	2	0	0	0	0
2	3	19	0.3654	0.2996	0.7291	7	3	1	0.0357	0.0011	0.2267
2	4	0	0	0	0	7	4	3	0.1071	0.0495	0.4255
2	5	0	0	0	0	7	5	2	0.0715	0.0235	0.3365
2	6	0	0	0	0	7	6	4	0.1429	0.0777	0.503
2	7	0	0	0	0	7	7	6	0.2143	0.1378	0.6366
2	8	0	0	0	0	7	8	7	0.25	0.1693	0.6957
2	9	0	0	0	0	7	9	1	0.0357	0.0011	0.2267
2	10	0	0	0	0	7	10	1	0.0357	0.0011	0.2267
3	1	5	0.0794	0.0458	0.2649	8	1	0	0	0	0
3	2	14	0.2222	0.1706	0.5075	8	2	1	0.0833	0.0047	0.5179
3	3	24	0.381	0.3206	0.7142	8	3	0	0	0	0
3	4	18	0.2857	0.2296	0.5957	8	4	1	0.0833	0.0047	0.5179
3	5	0	0	0	0	8	5	3	0.25	0.1268	0.9308
3	6	2	0.0317	0.01	0.1521	8	6	2	0.1668	0.0606	0.7526
3	7	0	0	0	0	8	7	0	0	0	0
3	8	0	0	0	0	8	8	3	0.25	0.1268	0.9308
3	9	0	0	0	0	8	9	1	0.0833	0.0047	0.5179
3	10	0	0	0	0	8	10	1	0.0833	0.0047	0.5179
4	1	1	0.0175	0.0004	0.1123	9	1	1	0.0556	0.0023	0.3496
4	2	5	0.0877	0.0508	0.2918	9	2	1	0.0556	0.0023	0.3496
4	3	16	0.2807	0.222	0.6049	9	3	0	0	0	0
4	4	22	0.386	0.3224	0.7372	9	4	1	0.0556	0.0023	0.3496
4	5	9	0.1579	0.1103	0.421	9	5	3	0.1666	0.0801	0.6451
4	6	4	0.0702	0.0368	0.2545	9	6	2	0.1111	0.0381	0.5146
4	7	0	0	0	0	9	7	2	0.1111	0.0381	0.5146
4	8	0	0	0	0	9	8	4	0.2222	0.1256	0.7559
4	9	0	0	0	0	9	9	1	0.0556	0.0023	0.3496
4	10	0	0	0	0	9	10	3	0.1666	0.0801	0.6451
5	1	0	0	0	0	10	1	0	0	0	0
5	2	2	0.0465	0.0149	0.2214	10	2	0	0	0	0
5	3	9	0.2093	0.1481	0.5472	10	3	2	0.0588	0.019	0.2786
5	4	10	0.2326	0.1691	0.5835	10	4	4	0.1177	0.0632	0.4186
5	5	10	0.2326	0.1691	0.5835	10	5	2	0.0588	0.019	0.2786
5	6	7	0.1628	0.1073	0.4694	10	6	6	0.1765	0.112	0.5326
5	7	4	0.093	0.0494	0.3343	10	7	4	0.1177	0.0632	0.4186
5	8	1	0.0232	0.0006	0.1484	10	8	2	0.0588	0.019	0.2786
5	9	0	0	0	0	10	9	3	0.0882	0.0403	0.3532
5	10	0	0	0	0	10	10	11	0.3235	0.2444	0.7605

Table A2. The MLE, sample sizes, and lower/upper bounds of transition probabilities' 99% overall confidence intervals for female patients.

Starting state	Ending state	Sample size	MLE	Lower bound	Upper bound	Starting state	Ending state	Sample size	MLE	Lower bound	Upper bound
1	1	11	0.6471	0.5328	1	6	1	0	0	0	0
1	2	6	0.3529	0.2387	0.9842	6	2	0	0	0	0
1	3	0	0	0	0	6	3	1	0.037	0.0012	0.235
1	4	0	0	0	0	6	4	5	0.1852	0.1115	0.5924
1	5	0	0	0	0	6	5	5	0.1852	0.1115	0.5924
1	6	0	0	0	0	6	6	8	0.2964	0.2097	0.7749
1	7	0	0	0	0	6	7	6	0.2222	0.1434	0.658
1	8	0	0	0	0	6	8	1	0.037	0.0012	0.235
1	9	0	0	0	0	6	9	1	0.037	0.0012	0.235
1	10	0	0	0	0	6	10	0	0	0	0
2	1	9	0.18	0.1264	0.4759	7	1	0	0	0	0
2	2	26	0.52	0.4504	0.9048	7	2	1	0.0588	0.0026	0.3696
2	3	11	0.22	0.1623	0.5391	7	3	0	0	0	0
2	4	3	0.06	0.0269	0.2429	7	4	1	0.0588	0.0026	0.3696
2	5	1	0.02	0.0005	0.1278	7	5	4	0.2353	0.1339	0.7956
2	6	0	0	0	0	7	6	4	0.2353	0.1339	0.7956
2	7	0	0	0	0	7	7	3	0.1766	0.0853	0.6801
2	8	0	0	0	0	7	8	2	0.1176	0.0406	0.5433
2	9	0	0	0	0	7	9	2	0.1176	0.0406	0.5433
2	10	0	0	0	0	7	10	0	0	0	0
3	1	2	0.0435	0.0138	0.2072	8	1	0	0	0	0
3	2	9	0.1957	0.138	0.5142	8	2	0	0	0	0
3	3	22	0.4783	0.4057	0.8794	8	3	0	0	0	0
3	4	10	0.2174	0.1574	0.5486	8	4	0	0	0	0
3	5	3	0.0651	0.0293	0.2635	8	5	0	0	0	0
3	6	0	0	0	0	8	6	1	0.25	0.0366	1
3	7	0	0	0	0	8	7	1	0.25	0.0366	1
3	8	0	0	0	0	8	8	0	0	0	0
3	9	0	0	0	0	8	9	2	0.5	0.2536	1
3	10	0	0	0	0	8	10	0	0	0	0
4	1	1	0.0192	0.0005	0.123	9	1	0	0	0	0
4	2	3	0.0578	0.0258	0.2338	9	2	0	0	0	0
4	3	13	0.25	0.1908	0.5771	9	3	0	0	0	0
4	4	20	0.3846	0.3181	0.7521	9	4	0	0	0	0
4	5	10	0.1923	0.1384	0.49	9	5	0	0	0	0
4	6	4	0.0769	0.0405	0.2782	9	6	1	0.125	0.0097	0.7619
4	7	1	0.0192	0.0005	0.123	9	7	1	0.125	0.0097	0.7619
4	8	0	0	0	0	9	8	2	0.25	0.0991	1
4	9	0	0	0	0	9	9	3	0.375	0.2063	1
4	10	0	0	0	0	9	10	1	0.125	0.0097	0.7619
5	1	1	0.0323	0.001	0.2051	10	1	0	0	0	0
5	2	0	0	0	0	10	2	0	0	0	0
5	3	6	0.1935	0.1236	0.58	10	3	0	0	0	0
5	4	9	0.2903	0.21	0.7344	10	4	1	0.05	0.002	0.3154
5	5	7	0.2258	0.1518	0.6348	10	5	3	0.15	0.0713	0.5849
5	6	6	0.1935	0.1236	0.58	10	6	1	0.05	0.002	0.3154
5	7	1	0.0323	0.001	0.2051	10	7	2	0.1	0.0339	0.4654
5	8	0	0	0	0	10	8	0	0	0	0
5	9	1	0.0323	0.001	0.2051	10	9	4	0.2	0.1118	0.6872
5	10	0	0	0	0	10	10	9	0.45	0.3403	1