Manufacturing & Service Operations Management

Interpretable Policies and the Price of Interpretability in Hypertension Treatment Planning

Gian-Gabriel P. Garcia, Lauren N. Steimle, Wesley J. Marrero, Jeremy B. Sussman

Interpretable Policies and the Price of Interpretability in Hypertension Treatment Planning

Gian-Gabriel P. Garcia,[a,*] Lauren N. Steimle,[a] Wesley J. Marrero,[b] Jeremy B. Sussman[c]

[a] H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332; [b] Thayer School of Engineering, Dartmouth College, Hanover, New Hampshire 03755; [c] Department of Internal Medicine, Michigan Medicine, University of Michigan, Ann Arbor, Michigan 48109
*Corresponding author

Contact: giangarcia@gatech.edu, https://orcid.org/0000-0001-9315-0195 (G-GPG); steimle@gatech.edu,
https://orcid.org/0000-0002-4073-6165 (LNS); wesley.marrero@dartmouth.edu, https://orcid.org/0000-0002-7092-2292 (WJM);
jeremysu@med.umich.edu, https://orcid.org/0000-0002-2158-3618 (JBS)

Abstract. *Problem definition*: Effective hypertension management is critical to reducing the consequences of atherosclerotic cardiovascular disease, a leading cause of death in the United States. Clinical guidelines for hypertension can be enhanced using decision-analytic approaches capable of capturing complexities in treatment planning. However, model-generated recommendations may be uninterpretable/unintuitive, limiting their clinical acceptability. We address this challenge by investigating interpretable treatment plans. *Methodology/results*: We formulate interpretable treatment plans as Markov decision processes (MDPs) and analyze the problems of optimizing *monotone* policies, which prohibit decreasing treatment intensity for sicker patients, and *class-ordered monotone policies*, which generalize monotone policies. We establish that both policies depend on initial state distributions and that optimal monotone policies can be generated tractably for many treatment planning problems. Next, we propose exact formulations for optimizing interpretable policies broadly. Then, we analyze the price of *interpretability*, proving that the class-ordered monotone policy's price of interpretability does not exceed the monotone policy's price of interpretability. Finally, we formulate and evaluate MDPs for hypertension treatment planning using a large nationally representative data set of the U.S. population. We compare the structure and performance of optimal monotone policies and class-ordered monotone policies with optimal MDP-based policies and current clinical guidelines. At the patient level, optimal MDP-based policies may be unintuitive, recommending more aggressive treatment for healthier patients than sicker patients. Conversely, monotone policies and class-ordered monotone policies never deescalate treatment, reflecting clinical intuition. Across 66.5 million patients, optimized monotone policies and class-ordered monotone policies outperform clinical guidelines, saving over 3,246 quality-adjusted life years per 100,000 patients, with both policies paying a low price of interpretability. Sensitivity analysis illustrates that monotone policies and class-ordered monotone policies are robust to various definitions of "interpretability." *Managerial implications*: Interpretable policies can be tractably optimized, drastically outperform existing guidelines, and perform near optimally—potentially increasing the acceptability of decision-analytic approaches in practice.

1. Introduction

Atherosclerotic cardiovascular disease (ASCVD), mainly composed of myocardial infarction and stroke, is a leading cause of death in the United States (Kochanek et al. 2019). Recent reports show coronary heart disease (which is largely manifested by myocardial infarction) and stroke account for 42.6% and 17.0% of deaths attributed to cardiovascular diseases in the United States, respectively (Virani et al. 2020). High blood pressure (BP; i.e., *hypertension*) is a major risk factor for ASCVD that affects 45.6% of adults in the United States (Whelton et al. 2018) and puts these individuals at a higher risk of experiencing myocardial infarction or stroke. Effective management of hypertension is critical to reducing adverse outcomes related to ASCVD.

Clinical practice guidelines (Whelton et al. 2018) play an important role in guiding hypertension management decisions. These guidelines are typically formed on the judgment of expert panels that aim to synthesize the most recent research and clinical evidence available. However, guidelines designed by expert consensus may fail to capture all of the risks, benefits, and uncertainty inherent to treatment planning. As such, they may be deemed as subjective (Cohen and Townsend 2018, Solberg and Miller 2018), and in the past, they have been met with substantial backlash (Ioannidis 2018). In contrast, decision-analytic approaches may better capture these risks, benefits, and uncertainties, and they have been shown to outperform clinical guidelines in simulation studies (Denton et al. 2009, Kurt et al. 2011, Mason et al. 2014, Schell et al. 2016, Steimle et al. 2021, Bonifonte et al. 2022). However, these models may generate complex decision guidelines that are not easily interpretable or may appear counterintuitive (Lakkaraju and Rudin 2017). Despite their apparent effectiveness in reducing adverse ASCVD outcomes, a lack of *interpretability* in decision-analytic approaches can limit their acceptability in clinical practice (Sethi et al. 2020, Wang et al. 2020).

Significant efforts have been put forth in creating interpretable prediction models (Rudin et al. 2022). Yet, interpretable decision-analytic methods remain underdeveloped. To increase the acceptability of decision-analytic models in clinical practice, it is critical to extend these methods for *interpretable treatment planning*. In this setting, a decision maker aims to optimize dynamic treatment rules for patients based on their health status, health risks, predicted future health trajectories, and costs and benefits of treatment—all while restricting attention to treatment plans within a specified class of *interpretable treatment policies* that are amenable to human intuition, cognition, and pattern recognition. Our focus on interpretable treatment policies does not refer to treatment policies that are inherently explainable or transparent, as may be case in the artificial intelligence literature (Tjoa and Guan 2021). Because of their ubiquity and wide adoption in the medical decision-making literature, we focus on interpretable treatment planning using Markov decision processes (MDPs), with a special emphasis on *monotone policies* and *class-ordered monotone policies* (i.e., treatment policies that increase in intensity as a patient's health worsens).

1.1. MDPs and Interpretable Policies

Early work on interpretability for MDPs has focused on structured policies that are optimal for specific applications of MDPs (e.g., inventory control (Bellman et al. 1955, Schäl 1976)) and on monotone policies (Serfozo 1976), where optimal policies prescribe actions that are monotone in the system state. In these applications, monotone policies are considered interpretable when the MDP's states and actions follow a strict ordering. Previous research has established sufficient conditions on the MDP parameters that guarantee the existence of an optimal policy that is monotone (Puterman 2014, section 6.11). Our research leverages the natural interpretability inherent in monotone policies and extends this past work by developing methods to derive an optimal interpretable policy when these sufficient conditions are not met. Moreover, we consider a generalization of the monotone policy (i.e., the *class-ordered monotone policy*), which can be interpretable when the states and/or actions do not follow a strict ordering. Policies that follow this type of structure are ideal in healthcare applications.

The literature on MDPs in healthcare applications is rich. Surveys by Denton et al. (2011), Capan et al. (2017), and Saville et al. (2019) comprehensively review many MDPs and partially observable Markov decision processes (POMDPs) in the medical decision-making literature. Recent examples include research by Cevik et al. (2018), Chen et al. (2018), Hicklin et al. (2018), Lee et al. (2018), Suen et al. (2018), Ayer et al. (2019), Boloori et al. (2020), Grand-Clément et al. (2021), Marrero et al. (2021), Skandari and Shechter (2021), Steimle et al. (2021), Bonifonte et al. (2022), and Tunç et al. (2022). A common goal in many of these prior works is to show that treatment plans generated by MDPs are naturally interpretable. For example, in treatment initiation and screening applications, the optimal policy may only initiate treatment or screen if a patient is sicker than some threshold (Denton et al. 2009, Skandari and Shechter 2021, Bonifonte et al. 2022, Tunç et al. 2022). Unfortunately, the sufficient conditions that guarantee an optimal policy with these interpretable structures can be difficult to verify and are often violated to some degree in practice. To build on this prior work, we develop a method to design optimal interpretable treatment plans without requiring these sufficient conditions. Notably, Grand-Clément et al. (2021) design deterministic tree policies for ventilator triage in an MDP framework.

Although these tree policies are different from the monotone and class-ordered monotone policies we study, all of these approaches are dependent on the initial state distribution. However, the types of interpretable policies we consider can be solved exactly with a mixed-integer program (MIP), whereas they design a value iteration-like algorithm to solve their problem approximately. Moreover, our methods can handle complex treatment options (e.g., multiple medications and dosages) (see Section 3.1), extending beyond the binary treatments (e.g., initiate treatment/wait) considered in many prior works.

Among POMDP applications in treatment planning, Cevik et al. (2018) and Chen et al. (2018) design interpretable policies for POMDPs in the context of screening for breast cancer and hepatitis C, respectively. Chen et al. (2018) consider a special form of interpretable policies called "M switch," which enforces that screening schedules must be at regular intervals and that the length of these intervals can only switch M times. The "M-switch" policies are interpretable relative to traditional recommendations from POMDPs for cancer screening, in which the optimal policy is not guaranteed to be of a regular frequency. Likewise, Cevik et al. (2018) consider breast cancer screening under resource constraints and design policies with the property that if it is optimal to screen a patient with a certain risk for breast cancer, then it should also be optimal to screen any patient with greater risk. Although our research focuses on interpretable policies for MDPs, we note that the policy in Cevik et al. (2018) is a special case of our novel class-ordered monotone policy.

Interpretability has also been of interest in POMDPs outside of healthcare. Monotone policies for POMDPs are described in Lovejoy (1987). There has been some investigation of interpretable and implementable, yet potentially suboptimal, policies for POMDPs. Early work in this area dates back to Littman (1994) and Vlassis et al. (2012), who study *memoryless policies* for POMDPs. Although memoryless policies are not guaranteed to be optimal for POMDPs, they are interpretable in the sense that the action taken by the decision maker depends only on the most recent observation rather than the entire history of observations and actions.

Perhaps the most closely related works to ours are those of Serin and Kulkarni (1995) and Petrik and Luss (2016), who consider interpretable policies in fully observed MDPs. Serin and Kulkarni (1995) and Petrik and Luss (2016) consider interpretable policies for MDPs by first partitioning the state space of an MDP into K sets. A policy is considered interpretable if the probability of taking an action is the same for all states in the same set. The decision maker's goal in this setting is to find the best policy among this type of interpretable policy. Serin and Kulkarni (1995) show that

this problem is a special case of finding the best memoryless policy in a POMDP. They also prove that in general, there is no guarantee that a deterministic interpretable policy will be optimal. Further, they show that the optimal policy depends on the initial distribution over the states and propose an iterative method for finding local optimal solutions. Petrik and Luss (2016) later show that solving for randomized or deterministic interpretable policies is NP hard and thus, propose an MIP to solve this problem. In this article, we propose the class-ordered monotone policy, a related type of interpretable policy wherein both the state space and action space are partitioned into classes, and we impose monotonicity on the space of state classes and action classes. We show that class-ordered monotone policies leverage the interpretability of monotone policies while also achieving better performance.

1.2. Contributions

Motivated by interpretable treatment planning, we develop, analyze, and apply new methods for designing interpretable policies using MDPs. We make the following contributions.

1. We formally define the interpretable treatment planning problem. We specifically analyze the problem of finding the optimal monotone policy, showing that the optimal monotone policy depends on the initial state distribution and that this problem can be solved in polynomial time under realistic assumptions for hypertension treatment planning. We also analyze the problem of finding an optimal *class-ordered monotone policy*—a new type of interpretable policy that generalizes monotone policies. We find that the optimal class-ordered monotone policy also depends on the initial state distribution.

2. We formulate and characterize MIP-based exact solution methods for finding optimal interpretable treatment policies, including monotone and class-ordered monotone policies. We show that these methods are amenable to both patient- and population-level treatment planning.

3. We introduce the *price of interpretability* in MDPs, which measures the difference in expected total discounted reward between the optimal policy and the best interpretable policy. We show that under mild conditions, the optimal class-ordered monotone policy is guaranteed to pay a lower price of interpretability than the optimal monotone policy.

4. We apply our methods to personalized hypertension treatment planning. Our study demonstrates that even when the optimal policy is not monotone, both the optimal monotone policy and the optimal class-ordered monotone policy perform close to optimal (i.e., pay a low price of interpretability) while being more clinically intuitive. Moreover, both the monotone policy and class-ordered monotone policy drastically outperform current

guidelines, saving over 3,246 quality-adjusted life years (QALYs) per 100,000 patients.

5. Through a sensitivity analysis in hypertension treatment planning, we show that both the optimal monotone policy and optimal class-ordered monotone policy are robust to changes in initial state distributions and different state and action class definitions—implying that both methods are amenable to context-specific interpretability.

The remainder of this article is organized as follows. In Section 2, we introduce the interpretable treatment planning problem and specifically analyze optimal monotone policies and class-ordered monotone policies. We then derive an exact MIP formulation to solve the interpretable treatment planning problem followed by an analysis of the price of interpretability. In Section 3, we formulate a finite-horizon MDP to derive hypertension treatment plans for the prevention of ASCVD, analyzing the structure and performance of optimal monotone policies and class-ordered monotone policies compared with the optimal MDP policy and current clinical guidelines. Finally, in Section 4, we conclude with a discussion of our findings and directions for future research.

2. Modeling Approach

In this section, we first present an infinite-horizon MDP formulation for optimal treatment planning. The infinite-horizon setting allows us to convey all of the key ideas in our methods and analysis. Next, we formally define the interpretable treatment planning problem, with a focus on the problems of finding the optimal monotone policy and optimal class-ordered monotone policy. Next, we provide an exact MIP formulation to solve the interpretable treatment planning problem for a broad class of interpretable policies, including monotone policies and class-ordered monotone policies. Finally, we define the price of interpretability and compare the class-ordered monotone policy and optimal monotone policy using this metric. Proofs for all technical results are provided in Section EC.1 of the e-companion. All of these methods can be flexibly adapted to finite-horizon MDPs by adding a temporal component (see Section EC.3.1 of the e-companion). Doing so allows for policies that can additionally impose monotonicity on time. We take this approach in our application to hypertension treatment planning (see Section 3).

2.1. Treatment Planning with MDPs

We formulate an infinite-horizon MDP for the treatment planning problem as follows. A patient's health is modeled with health states $\mathcal{S} = \{1, \ldots, S\}$. At each decision epoch $t \in \mathcal{T} = \{0, 1, \ldots\}$, a decision maker observes the state $s \in \mathcal{S}$ and then prescribes a treatment $a \in \mathcal{A} = \{1, \ldots, A\}$. When treatment a is prescribed in state s, the decision maker receives a finite reward $r(s,a)$ (e.g., quality-adjusted life years), and the patient's health transitions to

a new state s' according to a transition probability matrix P with entries

$$P(s'|s,a) = \mathbb{P}(s_{t+1} = s'|s_t = s, a_t = a).$$

Rewards are discounted at a rate $\gamma \in (0,1)$. At the first decision epoch $t = 0$, the patient's health is probabilistically generated by an initial state distribution α, where $\sum_{s \in \mathcal{S}} \alpha(s) = 1$ and $\alpha(s) \geq 0$ for all $s \in \mathcal{S}$.

The decision maker aims to design a treatment plan in order to maximize the total discounted rewards over the planning horizon. A deterministic stationary policy $\pi : \mathcal{S} \to \mathcal{A}$ is a decision rule that maps each state to a treatment. We focus on deterministic policies because they are generally regarded as more interpretable and implementable than randomized policies, especially in medical decision making where deterministic policies can be easily transformed into treatment guidelines or clinical decision aids. We denote by Π the set of all admissible policies. The total discounted reward accrued by an MDP under policy $\pi \in \Pi$ and initial state distribution α is given by

$$J^\pi(\alpha) = \mathbb{E}^\pi \left[\sum_{t=0}^\infty \gamma^t r(s_t, \pi(s_t)) \middle| \mathbb{P}(s_0 = s) = \alpha(s) \right].$$

The optimal policy is given by $\pi^* = \arg \max_{\pi \in \Pi} J^\pi(\alpha)$. It is well known that there exists a deterministic and stationary policy that is optimal (Puterman 2014, chapter 6.9). Moreover, $J^{\pi^*}(\alpha) = \sum_{s \in \mathcal{S}} \alpha(s) v^*(s)$, where $\mathbf{v} = [v^*(s)]_{s \in \mathcal{S}}$ are given by the optimal solution to the following linear program:

$$\min_{\mathbf{v}} \sum_{s \in \mathcal{S}} v(s) \quad \text{subject to} \quad v(s) \geq r(s,a)$$
$$+ \gamma \sum_{s' \in \mathcal{S}} P(s'|s,a)v(s') \quad \text{for all} \quad s \in \mathcal{S}, a \in \mathcal{A}.$$
$$(1)$$

For each $s \in \mathcal{S}$, $\pi^*(s)$ is equal to the action where the constraint on $v(s)$ is tight. Note that π^* is independent of α because of the principle of optimality.

2.2. Interpretable Treatment Planning with MDPs

Suppose that the decision maker aims to design an *interpretable* treatment plan that maximizes the total discounted rewards over the planning horizon. Let $\Pi^I \subset \Pi$ be a set denoting a specific type of interpretable policy. The interpretable treatment planning problem is given by

$$\pi^I = \arg \max_{\pi \in \Pi^I} J^\pi(\alpha). \qquad (2)$$

As discussed in Section 1.1, there are many types of interpretable policies. In the following subsections, we focus our attention on monotone policies and the novel class-ordered monotone policy.

2.2.1. Monotone Policies. In this section, we assume that the sets $\mathcal{S}$ and $\mathcal{A}$ are ordered. We now restrict our

attention to the set of monotone policies denoted by Π^M and defined as follows.

Definition 1 (Monotone Policy). Under ordered states and actions, a monotone policy is a policy $\pi : \mathcal{S} \to \mathcal{A}$ such that $\pi(s) \leq \pi(s')$ for all $s, s' \in \mathcal{S}$ such that $s \leq s'$.

Monotone policies are appealing to practitioners because they can be more easily interpreted and implemented compared with optimal policies that may lack structure. For instance, physicians may find treatment strategies more interpretable if the policies follow a natural order, such as increasing treatment intensity on the severity of a patient's health condition. Many prior works (see Section 1.1) have aimed to identify sufficient conditions on the MDP data (i.e., P and r) that guarantee that there exists an optimal policy that is monotone (i.e., $\pi^* \in \Pi^M$). In contrast, our aim is to determine the policy $\pi^M \in \Pi^M$ that achieves the greatest total discounted reward without requiring any conditions on the MDP data. Formally, our optimization problem is given by

$$\pi^M = \arg\max_{\pi \in \Pi^M} J^\pi(\alpha). \tag{3}$$

We now highlight an important property of the optimal monotone policy π^M.

Proposition 1. *The optimal monotone policy π^M depends on the initial distribution α.*

Proposition 1 implies that the initial state distribution must be known to create an optimal monotone policy π^M. This result contrasts the optimal policy, π^*, which does not depend on α. Moreover, for any $s \in \mathcal{S}$, the state-value function associated with π^M (i.e., $v^M(s)$) depends on α in general, whereas $v^*(s)$ does not. Because π^M and $v^M(s)$ depend on α, "off-the-shelf" algorithms for solving MDPs that do not account for the initial distribution may fail to solve (3). For example, with policy iteration, policy improvement steps for each state can be made in any arbitrary order because selecting an action in one state does not restrict what actions are available in other states. However, available actions in one state depend on selected actions in other states for monotone policies. Hence, these algorithms may need to be modified for this class of policies. Nevertheless, Theorem 1 shows that we can still find π^M in polynomial time when S and A grow independently.

Theorem 1. *If S (or A) is fixed, then the number of monotone policies grows as a polynomial in input size. Moreover, (3) is solvable in polynomial time.*

Theorem 1 shows that the optimal monotone policy π^M can be obtained in polynomial time using complete enumeration. However, the degree of this polynomial is nontrivial; approximation algorithms may still be needed for MDPs with large state and action spaces.

Further, if the states and actions grow together, we do not expect the problem to be solvable in polynomial time. However, there are many examples in treatment planning where the number of actions grows independently of the state space. For example, in hypertension treatment, the Framingham and American College of Cardiology/American Heart Association risk scores and their revised versions have all used similar risk factors (i.e., components of the state space) over time. Yet, with the approval of new drugs over the past few decades, the numbers of available medications have increased.

Although monotone policies are highly desirable because of their interpretability, they require the states and actions to follow a strict ordering. However, in practice, a strict ordering across states and actions may be difficult to define, especially for multidimensional state and action spaces. Although the issue of nonorderability is beyond our scope, we recognize that decision makers may feel more comfortable by defining orderings on groups of states and groups of actions (e.g., groupings by risk categories).

2.2.2. Class-Ordered Monotone Policies. In this section, we introduce a new form of interpretable policy that leverages the interpretability of monotone policies when orderings over groups of states and actions are provided. The *class-ordered monotone policy* generalizes monotone policies where monotonicity holds on ordered classes of states and actions rather than the states and actions themselves. Specifically, suppose that $\mathcal{S}$ is partitioned into ordered state classes $\mathcal{S}_1, \ldots, \mathcal{S}_K$ indexed by the set $\mathcal{K} = \{1, \ldots, K\}$ and $\mathcal{A}$ is partitioned into ordered action classes $\mathcal{A}_1, \ldots, \mathcal{A}_G$ indexed by the set $\mathcal{G} = \{1, \ldots, G\}$. Each state is mapped to exactly one state class through the function $\Theta : \mathcal{S} \to \mathcal{K}$, and each action is mapped to exactly one action class through the function $\Psi : \mathcal{A} \to \mathcal{G}$. The state classes can be interpreted such that for any $k' > k$, any state $s' \in \mathcal{S}_{k'}$ is "more severe" than any state $s \in \mathcal{S}_k$ (with similar interpretation for action classes). States and actions within a class are not required to be ordered. Using this construction, we now define class-ordered monotone policies.

Definition 2 (Class-Ordered Monotone Policy). A policy π is a *class-ordered monotone policy* if $\Theta(s) \geq \Theta(s')$ implies $\Psi(\pi(s)) \geq \Psi(\pi(s'))$.

Although class-ordered monotone policies do not enforce strict monotonicity across states and actions, they retain the natural interpretability inherent in monotone policies. In fact, class-ordered monotone policies generalize monotone policies; when Θ and Ψ are identity functions (i.e., $\Theta(s) = s$ and $\Psi(a) = a$), the resulting set of class-ordered monotone policies is the set of monotone policies. As the number of state and/or action

classes decreases, the corresponding set of class-ordered monotone policies grows. When there is one state class and one action class (i.e., $\Theta(s) = \Theta(s')$ for all $s, s' \in \mathcal{S}$ and $\Psi(a) = \Psi(a')$ for all $a, a' \in \mathcal{A}$), the resulting set of class-ordered monotone policies is equivalent to the set of all Markov deterministic policies.

Let $\Pi_{\Theta, \Psi}^{CM}$ be the set of class-ordered monotone policies with respect to Θ and Ψ. The optimal class-ordered monotone policy is the solution to

$$\pi^{CM} = \arg \max_{\pi \in \Pi_{\Theta, \Psi}^{CM}} J^\pi(\alpha). \qquad (4)$$

Like the optimal monotone policy, π^M, the optimal class-ordered monotone policy, π^{CM}, is also dependent on the initial state distribution, α, in general (see Section EC.1 of the e-companion).

2.3. Exact Solution Methods
In this section, we propose an exact MIP formulation to obtain optimal interpretable treatment plans for a broad class of interpretable policies, including monotone policies and class-ordered monotone policies. Specifically, we modify formulation (1) by adding binary decision variables $\mathbf{x} = [x(s, a)]_{s \in \mathcal{S}, a \in \mathcal{A}}$ to impose the desired interpretable structure within the policy. Define the set

$$\mathcal{X}_0 := \left\{ \mathbf{x} : x(s, a) \in \{0, 1\} \text{ for all } s \in \mathcal{S}, a \in \mathcal{A}, \right.$$
$$\left. \sum_{a \in \mathcal{A}} x(s, a) = 1 \text{ for all } s \in \mathcal{S} \right\}.$$

For any $\mathbf{x} \in \mathcal{X}_0$, we can construct the policy $\pi_\mathbf{x}$ such that $\pi_\mathbf{x}(s) = a$ if $x(s, a) = 1$. Additionally, suppose that there exists a set of constraints $\{f_n(\mathbf{x}) \le 0\}_{n=1}^N$, imposing the structure of Π^I on $\pi_\mathbf{x}$ (e.g., logic constraints). That is, for every $\pi \in \Pi^I$, there exists $\mathbf{x} \in \{\mathcal{X}_0 : f_n(\mathbf{x}) \le 0, n = 1, \ldots, N\}$ such that $\pi_\mathbf{x} = \pi$, and for every $\mathbf{x} \in \{\mathcal{X}_0 : f_n(\mathbf{x}) \le 0, n = 1, \ldots, N\}$, we have $\pi_\mathbf{x} \in \Pi^I$. In this setting, we propose the following MIP to solve (2) exactly:

$$\max_{\mathbf{v}, \mathbf{x}} \quad \sum_{s \in \mathcal{S}} \alpha(s) v(s) \qquad (5a)$$

$$\text{subject to} \quad v(s) \le r(s, a) + \gamma \sum_{s' \in \mathcal{S}} P(s' | s, a) v(s')$$
$$+ M_{s,a}(1 - x(s, a)) \text{ for all } s \in \mathcal{S}, a \in \mathcal{A} \qquad (5b)$$

$$\mathbf{x} \in \mathcal{X}_0 \qquad (5c)$$

$$f_n(\mathbf{x}) \le 0 \text{ for all } n = 1, \ldots, N. \qquad (5d)$$

Objective (5a) differs from the objective in (1) because of the epigraph Constraints (5b). In general, big-M constraints like (5b) weaken MIP formulations. Because any feasible $v(s)$ will never exceed $v^*(s)$, the value of $M_{s,a}$

should be chosen to ensure that the right-hand side of (5b) is bounded below by $v^*(s)$ when $x(s, a) = 0$. For example, if $r(s, a) \ge 0$ for all $s \in \mathcal{S}$ and $a \in \mathcal{A}$, setting $M_{s,a} = v^*(s)$ will satisfy this requirement, where $v^*(s)$ is obtained by solving (1). Meraklı and Küçükyavuz (2020) and Steimle et al. (2021) provide further discussion on how to set $M_{s,a}$. Finally, Constraints (5c) and (5d) ensure that the resulting policy is interpretable and deterministic, respectively. In Proposition 2, we show that (5) generates the optimal interpretable policy π^I.

Proposition 2. *Consider any optimal solution $(\tilde{\mathbf{v}}, \tilde{\mathbf{x}})$ to (5). Then, $J^{\pi_{\tilde{\mathbf{x}}}}(\alpha) = \sum_{s \in \mathcal{S}} \alpha(s) \tilde{v}(s)$, $\pi_{\tilde{\mathbf{x}}} \in \arg \max_{\pi' \in \Pi^I} J^{\pi'}(\alpha)$, and for any state s that is reachable under policy $\pi_{\tilde{\mathbf{x}}}$ and initial distribution α, we have $\tilde{v}(s) = \mathbb{E}^{\pi_{\tilde{\mathbf{x}}}} [\sum_{t=0}^\infty \gamma^t r(s_t, \pi_{\tilde{\mathbf{x}}}(s_t)) | s_0 = s]$.*

Notably, formulation (5) does not require $\alpha(s) > 0$ for all $s \in \mathcal{S}$, unlike variations of formulation (1) that incorporate α. This feature is important because in personalized medicine, we may have $\alpha(s) = 0$ for several $s \in \mathcal{S}$ (e.g., if the patient's initial health state is known). Finally, we remark that solving (5) is not guaranteed to provide the value of the policy for states that are unreachable under $\pi_{\tilde{\mathbf{x}}}$ and α, although they can be obtained by applying policy evaluation to π^I. We now show how formulation (5) can be modified to obtain the optimal monotone policy π^M and the optimal class-ordered monotone policy π^{CM}. To obtain π^M, we replace Constraint (5d) with

$$x(s, a) \le \sum_{a' \ge a} x(s + 1, a') \text{ for all } s \in \mathcal{S} \setminus S, a \in \mathcal{A}. \qquad (6)$$

Likewise, to obtain π^{CM}, we can replace Constraint (5d) with

$$\sum_{a \in \mathcal{A}_g} x(s, a) \le \sum_{a' \in \cup_{g'=g}^G \mathcal{A}_{g'}} x(s', a') \text{ for all } s \in \mathcal{S}_k, \quad s' \in \mathcal{S}_{k+1},$$
$$k = 1, \ldots, K - 1, \quad g = 1, \ldots, G. \qquad (7)$$

Thus, we can solve (3) and (4), respectively, by solving the following linear MIPs:

$$J^{\pi^M}(\alpha) = \max_{\mathbf{v}, \mathbf{x}} \quad \sum_{s \in \mathcal{S}} \alpha(s) v(s) \quad \text{subject to} \quad (5b), (5c), (6) \qquad (8)$$

$$J^{\pi^{CM}}(\alpha) = \max_{\mathbf{v}, \mathbf{x}} \quad \sum_{s \in \mathcal{S}} \alpha(s) v(s) \quad \text{subject to} \quad (5b), (5c), (7). \qquad (9)$$

We discuss practical considerations for implementing (8) and (9) in Section EC.2 of the e-companion.

2.4. The Price of Interpretability
A decision maker who values interpretability may be interested in the cost of implementing the best interpretable policy instead of the optimal policy. Thus, we introduce the *price of interpretability*.

Definition 3 (Price of Interpretability). Let $\Pi^I \subset \Pi$ be a specific class of *interpretable* policies. The price of interpretability for Π^I is defined as $PI(\Pi^I) := J^{\pi^*}(\alpha) - \max_{\pi \in \Pi^I} J^{\pi}(\alpha)$.

The price of interpretability informs the decision maker about the cost of implementing a policy from the interpretable policy class Π^I rather than implementing the optimal policy, π^*. We now compare the optimal monotone and class-ordered monotone policies using the price of interpretability. To facilitate our analysis, we make the following assumption.

Assumption 1. *The set of states and actions is indexed according to a strict monotone ordering, and the class functions Θ and Ψ are nondecreasing in these indices.*

Assumption 1 restricts our attention to the set of class-ordered monotone policies that relax the strict ordering imposed in standard monotone policies. These conditions are natural in cases where a strict ordering on the state and action spaces can be constructed, but some ordering relations may be questionable; thus, a partial ordering may be more appropriate. For example, research suggests that the benefit of antihypertensive treatment is mainly determined by their blood pressure reduction, with little effect attributable to drug-specific factors (Law et al. 2009). Although it may be reasonable to order treatment choices in terms of number of medications, it is less clear how antihypertensive drug types should be ordered given the same number of medications. The following result leverages the fact the monotone policies are a special case of class-ordered monotone policies.

Proposition 3. *If Θ and Ψ satisfy Assumption 1, we have $PI(\Pi^M) \geq PI(\Pi^{CM}_{\Theta,\Psi})$.*

Proposition 3 illustrates the value of using a class-ordered monotone policy over a monotone policy; if a strict ordering is ambiguous and state and/or action classes are more intuitive, then the optimal class-ordered monotone policy can achieve a lower price of interpretability than the optimal monotone policy while providing a more intuitive structure. Of course, determining state and action class functions is not always straightforward. We provide some guidance on creating state and class functions in Section EC.2.3 of the e-companion.

3. Numerical Analysis: Personalized Hypertension Treatment

We now apply our optimized interpretable policies to the management of ASCVD. We begin by describing our MDP, model parameters, data sources, and treatment strategies. Next, we present the treatment plans (i.e., the optimal policy, interpretable policies, and clinical guidelines) and how they are derived. Finally, we analyze the health outcomes of patients following each

of these treatment plans and discuss the implications of the interpretable hypertension treatment plans at a patient level and a population level.

3.1. Markov Decision Process Formulation

Because the risk for ASCVD events is nonstationary with respect to patients' age, we model the process of sequentially determining antihypertensive medications as a finite-horizon MDP. Specifically, we update MDP formulation in Schell et al. (2016) with the latest reliable parameters in the cardiovascular disease management literature.

The adaptation of our formulations to finite-horizon MDPs for the management of hypertension is described in Section EC.3.1 of the e-companion. We add an index t to states, actions, transition probabilities, and rewards to highlight their dependence on the decision epoch, which represents the effect of patients' age on the MDP parameters. As clinicians rarely decrease or discontinue the use of antihypertensive medications to control their patients' BP (Van Der Wardt et al. 2017), we guarantee nondecreasing actions over time by incorporating a temporal component in our state definitions (see Section EC.3.2 of the e-companion). The elements of our MDP are as follows.

● The 10-year planning horizon is given by a set of decision epochs $\mathcal{T}' := \{0, 1, \ldots, T\}$. Decision epoch $t \in \mathcal{T}'$ represents the year $[t, t+1)$, and $T-1$ is the year in which physicians select the last action. We use $T = 10$ to represent the 10-year effects of treatment on the patients' lifetime risk of ASCVD-related death. This planning horizon is reflective of 10-year ASCVD risk considerations in many major clinical guidelines for the management of cardiovascular diseases (Goff et al. 2014, Whelton et al. 2018). Moreover, we choose to use one-year decision epochs because our modeling data are available only in one-year increments (see Section 3.2) and because one-year intervals are consistent with regular annual primary care visits for most adults (Rao et al. 2019).

● The patient's state space $\mathcal{S}$ is comprised of demographic information d_t (i.e., age, sex, race, and smoking status), clinical observations c_t (i.e., systolic BP, diastolic BP, total cholesterol, high-density lipoprotein, and diabetes status), and health condition h_t. The health condition of each patient is one of the following mutually exclusive categories: healthy ($h_t = 1$), history of myocardial infarction (i.e., heart attack) but no adverse event in the current year ($h_t = 2$), history of stroke but no adverse event in the current year ($h_t = 3$), history of myocardial infarction and stroke but no adverse event in the current year ($h_t = 4$), survival of a myocardial infarction event ($h_t = 5$), survival of a stroke event ($h_t = 6$), death from a non-ASCVD–related cause ($h_t = 7$), death from a myocardial infarction event ($h_t = 8$), death from stroke ($h_t = 9$), and dead ($h_t = 10$). We use s_t (d_t, c_t, h_t) to denote specific components of a patient's state. Otherwise, we simply use s_t to denote the patient's state.

- The action space $\mathcal{A}$ is composed of zero to five antihypertensive medications of five different drug types at their standard dose. The selected action indicates the drug combination that should be taken that year. We include the following medications: thiazide diuretics, beta-blockers, calcium channel blockers, angiotensin-converting enzyme (ACE) inhibitors, and angiotensin II receptor blockers (ARBs). Because the simultaneous use of ACE inhibitors and ARBs is potentially harmful (Whelton et al. 2018), we exclude the combination of these two drug types from $\mathcal{A}$. Our action space contains a total of 196 treatment choices. The estimates of the effects of antihypertensive drugs on ASCVD events are derived from Law et al. (2009). The disutilities associated with each medication type are estimated from Law et al. (2003), Sussman et al. (2013), and Schell et al. (2016) (see Section EC.3.3 of the e-companion for details).

- We derive the transition probability $p_t(s_{t+1}|s_t, a_t)$ from the patient's risk for ASCVD events (Goff et al. 2014), the benefit from treatment (Law et al. 2009), fatality likelihoods (Kochanek et al. 2019), and non-ASCVD mortality (Arias 2019). Based on communications with clinical collaborators, we assume independence among myocardial infarction and stroke events. Myocardial infarction events account for 70% of the ASCVD risk, and stroke events account for the remaining 30% (Virani et al. 2020). To be consistent with previous studies, we assume that patients are more likely to have additional myocardial infarction or stroke events if they have a history of such ASCVD events (Schell et al. 2016). We account for this assumption by adjusting patients' myocardial infarction and stroke odds if they have a history of either ASCVD event (Burn et al. 1994, Brønnum-Hansen et al. 2001).

- The patient's reward $r_t(s_t, a_t)$ is given by the quality of life weight associated with health condition h_t minus the treatment-related disutility from medication a_t. The quality of life weights and treatment-related disutilities are obtained from previous studies (Law et al. 2003, Kohli-Lynch et al. 2019). Terminal rewards $r_T(s_T)$ represent patients' total QALYs after treatment over the planning horizon. We assume that the terminal rewards can be computed as the product of the patient's expected lifetime (Arias 2019), a mortality factor that accounts for the effect of ASCVD events on future mortality (Pandya et al. 2015), and a terminal quality of life weight (Kohli-Lynch et al. 2019).

- The initial state distribution $\alpha(s_t)$ is used to represent the patient's health condition. Recall that the index t is incorporated into the state definition and represents the effect of patients' age. We select α based on patients' characteristics and test our assumptions in sensitivity analyses.

- The model's discount factor is given by γ. We use $\gamma = 0.97$ as per recommendations in the medical literature (Neumann et al. 2016).

The parameters used throughout our numerical study are described in more detail in Section EC.3.3 of the e-companion.

3.1.1. Treatment Strategies. We aim to determine the policies that maximize patients' discounted QALYs, a common metric to assess the quality and quantity of life associated with health interventions. The optimal antihypertensive treatment strategy (denoted π^*) is determined by solving the dual of formulation (1). We obtain our class-ordered monotone treatment plans π^{CM} and monotone treatment plans π^M by solving the finite-horizon adaptations of formulations (9) and (8), respectively (see formulations EC.1 and EC.2 in Section EC.2.1 of the e-companion). All these strategies share the same MDP. However, each of them has a different level of restriction in the actions that can be prescribed at a state and decision epoch, as encoded in the constraints of their formulations.

For comparison purposes, we also derive a treatment strategy that follows the recommendations from the 2017 Hypertension Clinical Practice Guidelines (Whelton et al. 2018). These guidelines define stage 1 (stage 2) hypertension as a systolic BP of 130–139 mm Hg (at least 140 mm Hg) or diastolic BP of 80–89 mm Hg (at least 90 mm Hg). They suggest pharmacological treatment for patients with stage 1 hypertension if their 10-year risk for ASCVD exceeds 10%. For patients with stage 2 hypertension, the guidelines recommend treatment until they reach controlled BP levels below stage 1 hypertension. As the clinical guidelines only provide suggestions regarding the number of medications, we formulate a linear program to find the drug type that maximizes each patient's expected total discounted QALYs. This optimization model follows formulation (1) with additional constraints to guarantee that the numbers of medications match the recommendations by the clinical guidelines. The treatment suggestions from the 2017 Hypertension Clinical Practice Guidelines together with each patient's linear program will be referred to as the clinical guidelines.

3.2. Data Source

We use data from the National Health and Nutrition Examination Survey (NHANES) to parameterize our models. NHANES offers large, high-quality, and nationally representative data; it is unique in that it combines interviews and physical examinations, and it administers tests of physical activity and fitness. Our population is composed of adult Caucasian or African-American patients from 40 to 60 years old with no history of ASCVD. This inclusion criteria lead to 4,590 records representing a total population of 66.50 million people. To estimate the progression of patients' risk factors over the planning horizon, we linearly regress systolic BP, diastolic BP, high-density lipoprotein, and

total cholesterol on age, age squared, gender, race, smoking status, and diabetes status. We assume that smoking and diabetes status remain constant throughout the planning horizon. We then use these estimated risk factors as inputs into the American Heart Association/ American College of Cardiology's 10-year ASCVD risk calculator (Goff et al. 2014) to calculate each patient's risk for ASCVD events, which is adjusted if the patient experiences an adverse event. Death from non-ASCVD causes is modeled independently and not considered in the risk factor progression.

3.3. State and Action Ordering

We order each patient's state space based on the associated risk for ASCVD events. At each decision epoch t, each patient's state space has fixed demographic information, d_t, and estimated clinical observation, c_t. Consequently, any difference in risk for ASCVD events among each patient's states is driven by their health condition h_t. As a result, each patient's state ordering is determined by h_t. Excluding health conditions associated with death, we order patients' states according to the severity of their health condition. This leads to the following order of states for each patient at a specific decision epoch: $s_t(d_t, c_t, 1), s_t(d_t, c_t, 2), s_t(d_t, c_t, 5), s_t(d_t, c_t, 3), s_t(d_t, c_t, 6)$, and $s_t(d_t, c_t, 4)$.

The state classes are also made based on h_t. Given patients' demographic information and clinical observations, we define the following state classes: $\mathcal{S}_1 = \{s_t(d_t, c_t, 1)\}, \mathcal{S}_2 = \{s_t(d_t, c_t, 2), s_t(d_t, c_t, 5)\}, \mathcal{S}_3 = \{s_t(d_t, c_t, 3), s_t(d_t, c_t, 6)\}$, and $\mathcal{S}_4 = \{s_t(d_t, c_t, 4)\}$. The first state class includes the states at which patients are healthy, the second class encompasses the states associated with myocardial infarction events, the third class covers the states related to stroke events, and the fourth class comprises the states at which patients have a history of both ASCVD events. We did not consider the states with health conditions associated with death, as no treatment is possible in these states.

Actions are ordered as per the expected risk reductions of their associated drug combinations as described in Law et al. (2003, 2009). Ties in the expected risk reduction among drug combinations were broken arbitrarily. For example, the expected risk reduction associated with one dose of each drug type leads to the following order (from lowest to highest estimated risk reduction): ACE inhibitors, calcium channel blockers, thiazides, beta-blockers, and ARBs. This order is equivalent to sorting drug medications according to their expected systolic BP reductions. In clinical practice, the drug-type selection is often done for patient-specific reasons related to side effects, such as if a patient does not tolerate blood draws or is strongly opposed to leg swelling. Nevertheless, the difference between the drugs is small, so this distinction is likely practically negligible.

We create action classes on the basis of the number of antihypertensive medications being prescribed. The first action class $\mathcal{A}_0$ encompasses the no treatment action, and action class $\mathcal{A}_i$ comprises any combination of i antihypertensive medications at standard dose, for $i = 1, \ldots, 5$. Note that our initial selection of class functions Θ and Ψ satisfies Assumption 1. We study the impact of the state and action classes in our sensitivity analysis.

3.4. Analysis

To understand the implications of interpretable treatment plans at a patient level, we examine the effect of patients' characteristics on the optimal monotone and class-ordered monotone policies. We then study the trade-off between optimality and interpretability at a population level by comparing our policies with the optimal policy and clinical guidelines. We begin by inspecting the number and type of medications recommended by each treatment strategy. Subsequently, we assess the total discounted QALYs saved and ASCVD events prevented by our policies compared with the clinical guidelines. Lastly, we inspect the price of interpretability for the monotone policy, class-ordered monotone policy, and clinical guidelines. To compare if the price of interpretability paid by each policy is statistically different, we use Wilcoxon signed rank tests with a significance level of 0.05. The significance levels and confidence intervals (CIs) of individual statistical tests are adjusted with the Bonferroni correction method when multiple statistical tests are performed simultaneously.

We study the policy implications of each treatment strategy by dividing our population into BP categories. These categories are created based on the 2017 Hypertension Clinical Practice Guidelines: normal BP, elevated BP, stage 1 hypertension, and stage 2 hypertension. Although patients in the NHANES data set have different demographic information d_t and clinical observations c_t at each time period t, they all share the same initial health condition $h_0 = 1$. This initial health condition acknowledges that patients have no history of ASCVD events at the beginning of the planning horizon. To reflect patients' initial health condition in our study, we assign 100% of the initial state distribution to the states associated with healthy conditions at the first year of our study (i.e., $\alpha(s_0(d_0, c_0, h_0 = 1)) = 1$). We limit the total time each formulation spends obtaining an optimal solution to 60 minutes per record. Records exceeding this time limit in any optimization model are excluded from our analysis.

We also perform sensitivity analysis on the treatment strategies by varying our modeling assumptions. Our sensitivity analysis scenarios are described in Section EC.3.5 of the e-companion. These scenarios are selected based on communications with our clinical collaborators and information available in NHANES. In each scenario,

we evaluate each interpretable policy based on the price of interpretability and the number of ASCVD events allowed compared with the optimal policy. Any record with an optimization model exceeding the time limit in any scenario was excluded from all sensitivity analysis scenarios. The complete data set and code used in our analyses are available on GitHub (https://github.com/wesleymarrero/structured_optimal_policies) for review and reproducibility.

3.5. Numerical Results

In this section, we examine and describe the effect of the interpretable hypertension treatment plans along with key insights based on these results at the patient and population levels.

3.5.1. Patient-Level Results.

We now evaluate the optimal monotone and class-ordered monotone policies in a series of patient profiles. For comparison purposes, we also determine the optimal policy and a treatment policy based on the clinical guidelines for each patient profile. We first obtain treatment plans for the following patient profile: a 45-year-old nondiabetic, nonsmoker individual with normal BP and normal cholesterol levels. This patient profile will be referred to as the base patient profile. Note that this patient profile does not have any major clinical risk factors for ASCVD. We modify the BP levels of the patient and examine how the policies change.

Figure 1 shows the optimal, optimal monotone, and optimal class-ordered monotone policies over the health conditions of the patient profiles with stage 1 hypertension

(Figure 1(a)) and stage 2 hypertension (Figure 1(b)) in the last year of our study. The strategies are less intense in earlier years because of our monotonicity restrictions on the actions over time. In the base patient profile, all strategies coincide in recommending no treatment (not shown). Thus, there is no price of interpretability paid by the interpretable policies in this profile.

Increasing the base profile's BP level to elevated BP (not shown) or stage 1 hypertension leads to the suggestion of one ARB at standard dose when the patient has no history of ASCVD in all policies. The optimal policy decreases the intensity to no treatment if the patient profile's state is associated with the survival or history of an ASCVD event. Although optimal, this strategy is not intuitive for physicians or their patients. In contrast, the optimal monotone and class-ordered monotone policies account for interpretability by recommending one ARB across all states. For these patient profiles, there is no considerable consequence for providing interpretability as the price of interpretability paid by these policies is less than 0.0001 QALYs.

When the patient profile has stage 2 hypertension, all the strategies recommend a beta-blocker at standard dose if the patient has no history for ASCVD events or has ever survived a myocardial infarction. If the patient has ever survived a stroke, both the optimal policy and optimal class-ordered monotone policy suggest prescribing one ACE inhibitor. Conversely, because ACE inhibitors have a smaller expected risk reduction than beta-blockers, the optimal monotone policy continues to recommend one beta-blocker. Although the optimal class-ordered monotone policy pays no price

Figure 1. (Color online) Treatment Policies over the Health Conditions of Profiles with (a) Stage 1 Hypertension and (b) Stage 2 Hypertension
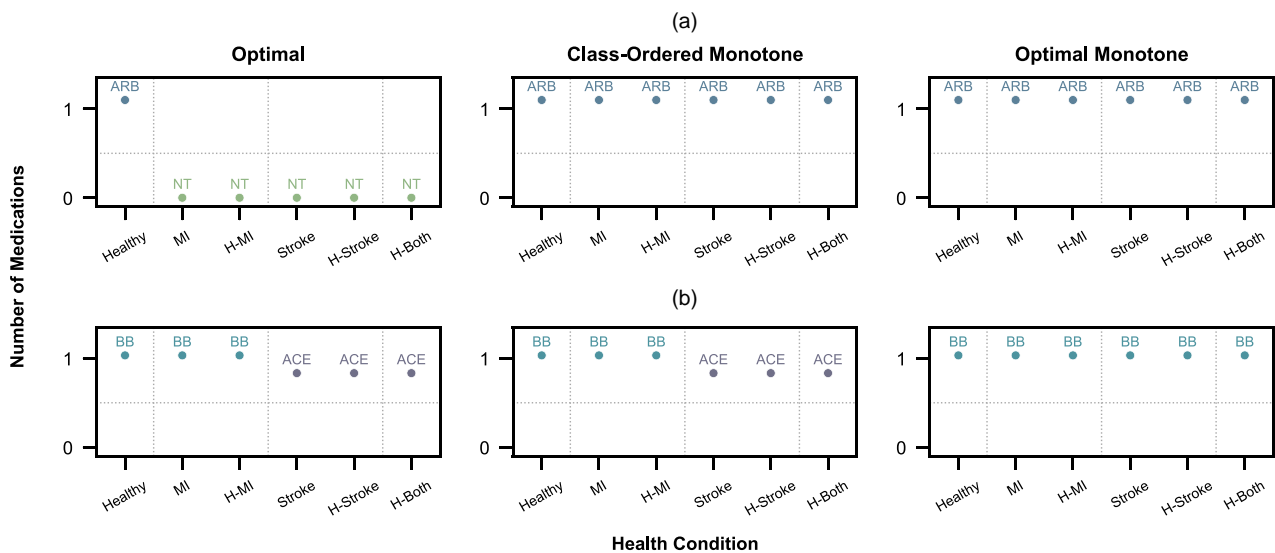
Notes. Action classes $\mathcal{A}_0$ and $\mathcal{A}_1$ are separated by horizontal dotted lines and state classes $\mathcal{S}_1$, $\mathcal{S}_2$, $\mathcal{S}_3$, and $\mathcal{S}_4$ are separated by vertical dotted lines. ACE, ACE inhibitor; BB, beta-blocker; H-, history of; MI, myocardial infarction; NT, no treatment.

of interpretability for this patient profile, the optimal monotone policy does.

Figure 1 excludes the clinical guidelines to ease comparisons between the optimal policy, optimal class-ordered monotone policy, and optimal monotone policy. The clinical guidelines recommend no treatment for normal and elevated BP levels. The profile with stage 1 hypertension is prescribed one ARB if they have no history of ASCVD events or history of a single adverse event and one ACE inhibitor for the state with history of both ASCVD events. Lastly, the profile with stage 2 hypertension is recommended one ARB for the healthy and myocardial infarction-related states. The clinical guidelines suggest one ACE inhibitor for the states associated with stroke events.

We now make two key observations based on these patient-level results. First, the optimal policy tends to suggest less treatment as the severity of the health condition increases. This conduct does not reflect physicians' intuition in practice. A potential explanation for this behavior is that the policy aims to maximize the expected discounted QALYs and not to minimize the total number of ASCVD events. As a result, the optimal policy recommends the most aggressive treatment possible to avoid primary events and maintain patients in the healthy condition. The effect of additional treatment in the transition probabilities and rewards is typically smaller than the treatment-related disutility in the states associated with ASCVD events. Even though the optimal policy is unintuitive, QALYs reflect decision criteria used in clinical practice because they balance between the health benefits and treatment disutilities associated with medication. Different reward structures may greatly alter the optimal policy. For example, if we instead minimize the total number of ASCVD events, the optimal policy may prescribe the maximum allowable dosage of antihypertensive medication. However, this policy might be considered too aggressive because

it ignores the disutilities associated with treatment, especially at high doses, and may still be considered unintuitive. Second, the optimal monotone policy typically prescribes a constant treatment across all health conditions in each year of the planning horizon. Similarly, the optimal class-ordered monotone policy usually keeps the number of medications constant throughout all health conditions at each year of the planning horizon. Rather than reducing the number of medications, the optimal class-ordered monotone policy generally prescribes a drug combination with lower risk reductions and treatment-related disutilities. For example, the optimal class-ordered monotone policy may prescribe ACE inhibitors instead of beta-blockers. These patterns align with the intuition of physicians in practice. Hence, the optimal monotone and class-ordered monotone policies provide more intuitive strategies than the optimal policy with only a small loss in QALYs. However, as the difference among antihypertensive drugs is small, the choice of drug type may be driven by patient-specific reasons instead of QALYs in practice.

3.5.2. Population-Level Results. Across a population of 66.50 million people, 27.35 million (41.13%) have normal BP, 12.49 million (18.78%) have elevated BP, 16.51 million (24.83%) have stage 1 hypertension, and 10.15 million (15.26%) have stage 2 hypertension. These findings coincide with recent age-adjusted hypertension prevalence trends across adults in the United States (Virani et al. 2020). The following results are summarized in Table 1 and correspond to patients in the first year of our study.

3.5.2.1. Health Outcomes. As hardly any patient receives treatment under any of the policies in the normal BP category, we focus on patients with elevated BP, stage 1 hypertension, and stage 2 hypertension. We now evaluate the outcomes of patients under each treatment

Table 1. Population Health Outcomes

BP category	Cohort, %	Total discounted QALYs saved[a]			ASCVD events averted[a]			Price of interpretability[b]		
		π^*	π^{CM}	π^M	π^*	π^{CM}	π^M	π^{CM}	π^M	Guidelines
Overall	100	3,262.01 (0.48%)	3,246.62 (0.00%)	3,246.46 —	305.98 (0.29%)	305.10 (0.00%)	305.09 —	15.38 (99.53%)	15.55 (99.52%)	3,262.01 —
Elevated BP	18.8	5,602.90 (0.60%)	5,569.24 (0.00%)	5,569.24 —	518.16 (0.52%)	515.50 (0.00%)	515.50 —	33.65 (99.40%)	33.66 (99.40%)	5,602.90 —
Stage 1 HTN	24.8	7,391.59 (0.26%)	7,372.68 (0.00%)	7,372.58 —	692.00 (0.10%)	691.34 (0.00%)	691.33 —	18.89 (99.74%)	19.01 (99.74%)	7,391.59 —
Stage 2 HTN	15.3	2,451.24 (1.22%)	2,422.57 (0.04%)	2,421.70 —	241.25 (0.62%)	239.81 (0.02%)	239.77 —	28.66 (98.83%)	29.54 (98.80%)	2,451.24 —

Notes. Values in parentheses represent the percentage improvement over the clinical guidelines in QALYs saved and events averted compared with the optimal monotone policy and the percentage reduction in the price of interpretability compared with the clinical guidelines. HTN, hypertension.
  [a]Per 100,000 patients compared with the guidelines.
  [b]Per 100,000 patients compared with the optimal policy.

strategy in terms of the number of total discounted QALYs saved and ASCVD events prevented compared with the clinical guidelines. In total, the optimal policy, the optimal class-ordered monotone policy, and the optimal monotone policy save 3,262.01, 3,246.62, and 3,246.46 total discounted QALYs per 100,000 patients over the planning horizon compared with the clinical guidelines, respectively. We notice a similar pattern when comparing the policies in terms of ASCVD events averted. Over the 10-year planning horizon, the optimal policy, the optimal class-ordered monotone policy, and the optimal monotone policy prevent 305.98, 305.10, and 305.09 ASCVD events per 100,000 patients, respectively, compared with the clinical guidelines.

Patients with stage 2 hypertension receive the greatest benefit from treatment. We note that patients' health outcomes under the optimal monotone and class-ordered monotone policies are not substantially different for patients with elevated BP or stage 1 hypertension. In people with stage 2 hypertension, the optimal class-ordered monotone policy saves 0.87 total discounted QALYs per 100,000 patients and averts 0.04 ASCVD events per 100,000 patients more than the optimal monotone policy. The clinical guidelines are outperformed by our interpretable policies in every BP category.

3.5.2.2. Price of Interpretability. Overall, the prices of interpretability paid by the optimal class-ordered monotone policy, optimal monotone policy, and clinical guidelines are 15.38, 15.55, and 3,262 total discounted QALYs per 100,000 patients, respectively We also show the price of interpretability associated with every patient in our data set following the optimal monotone and class-ordered monotone policies in Figure EC.2 in Section EC.3.4 of the e-companion.

Using Wilcoxon signed rank tests, there is enough evidence to conclude that the price of interpretability paid by the optimal monotone policy is significantly greater than the price of interpretability paid by the optimal class-ordered monotone policy across all patients (95% CI $[1^{-6}, \infty], p = 3^{-8}$) and in each BP category (98% CI $[4^{-15}, \infty], p = 0.0019$; 98% CI $[3^{-10}, \infty], p = 0.0019$; and 98% CI $[5^{-5}, \infty], p = 0.0002$ for elevated BP, stage 1 hypertension, and stage 2 hypertension, respectively). Because we are comparing three BP categories simultaneously, statistical significance was determined using a Bonferroni threshold of 0.02. Based on these results, we highlight several major insights at the population level. First, the price of interpretability paid by the optimal monotone and class-ordered monotone policies generally increases with patients' BP. Although these interpretable policies would never decrease treatment intensity as patients' BP increases, the optimal policy sometimes does. Thus, the monotonicity constraints may become more restrictive. For a similar reason, the pairwise differences between the price of interpretability paid by the optimal

monotone and class-ordered monotone policies tend to grow as patients' BP increases. We believe that this trend likely arises because the monotone policy is a more restrictive policy than the class-ordered monotone policy. Fourth, the class-ordered monotone policy's restrictiveness normally depends on patients' BP level. Patients with higher BP readings generally receive more treatment; as the number of medications increases, so does the number of potential drug-type combinations. A greater number of medications and drug combinations results in larger action classes, which lead to fewer restrictions in the optimal class-ordered monotone policy. Finally, we find that the optimal class-ordered monotone policy offers intuitive treatment strategies to physicians with modest improvements over the optimal monotone policy. Likewise, the optimal class-ordered monotone policy results in similar health outcomes as the optimal policy with the added benefits of interpretability.

3.6. Sensitivity Analyses

We proceed to study how the treatment strategies are affected by changing our modeling assumptions. The results of our sensitivity analyses are summarized in Table 2, and the sensitivity analysis scenarios are described in detail in Section EC.3.5 of the e-companion. Note that the price of interpretability and the number of ASCVD events allowed in the base case are different than in our main analysis. This difference is because of a larger number of patients exceeding the time limit in the optimization models (60 minutes). In our main analysis, we are excluding 16.22 thousand patients because of the time limit, whereas in our sensitivity analyses, we are excluding 13.03 million patients (see Section EC.3.5.1 of the e-companion for details). This exclusion allows us to compare the performance of the policies across the sensitivity analysis scenarios.

Overall, we find that modifying our modeling assumptions can affect our results in different magnitudes. To a small extent, the ordering and classification of the states can alter the price of interpretability and number of ASCVD events averted by our interpretable policies. For example, ordering the states in nonincreasing severity of health condition forces the interpretable policies to prescribe at most as much medication in the states associated with ASCVD events as they are prescribing in the states related to the healthy condition. Recommending less medication to patients with a history of ASCVD events may prompt additional events. Moreover, ordering the actions according to their diastolic BP reductions has a larger impact on the health outcomes for our interpretable policies. This finding may arise because patients' diastolic BP is unnecessary for calculating risk of ASCVD events. This risk only considers patients' systolic BP, implying that the transition probabilities and rewards are not affected by patients' diastolic BP. Even if an action is expected to have a high diastolic BP reduction, it may not

Table 2. Sensitivity Analyses Summary

Scenario	Price of interpretability		ASCVD events allowed[a]		Number of medications[b]	
	π^{CM}	π^M	π^{CM}	π^M	π^{CM}	π^M
Base case	18.31	18.33	2.01	2.01	19.88, 35.85, 50.84	19.88, 35.84, 50.84
Nonincreasing severity state order	17.98	18.19	2.01	2.01	19.88, 35.85, 50.84	19.88, 35.84, 50.84
Single ASCVD events state class	18.31	18.33	2.01	2.01	19.88, 35.92, 50.89	19.88, 35.84, 50.84
Action order and classes						
Systolic BP reductions	18.32	18.33	2.01	2.01	19.88, 35.85, 50.84	19.88, 35.84, 50.84
Diastolic BP reductions	19.60	71.49	2.15	6.67	19.83, 35.85, 50.84	19.81, 35.76, 50.84
Initial state distribution						
99% in healthy states at year 0	19.33	19.35	2.01	2.01	19.88, 35.85, 50.84	19.88, 35.84, 50.83
99% in year 0	92.01	98.27	29.98	29.77	16.03, 25.77, 36.42	16.47, 25.85, 36.52
Uniform weight	121.36	129.13	29.91	30.07	15.55, 25.92, 36.68	15.8, 25.74, 37.09

Note. All results are presented as the average per 100,000 patients.
  [a]Results correspond to the first year of our study. Values are presented in thousands.
  [b]Values represent year 0, year 4, and year 9 of our study, respectively, in thousands.

lead to a high risk reduction. Additionally, we find that how action classes are constructed may have moderate to large effects. For instance, categorizing actions according to their diastolic BP reductions may restrict the class-ordered monotone policy to classes with more intense treatment, which may not necessarily lead to larger risk reductions. As a consequence, patients may experience worse health outcomes if the action classes are created according to diastolic BP reductions versus the number of medications or systolic BP reductions. We provide additional comments on our sensitivity analysis in EC.3.5.2 of the e-companion.

4. Conclusions

MDPs are a powerful tool capable of capturing the risks, benefits, and uncertainties in optimal treatment planning. Yet, optimal MDP-based policy recommendations may be unintuitive or uninterpretable for clinicians, potentially resulting in a reluctance to implement these policies in practice. To address this issue, we proposed the interpretable treatment planning problem with a focus on finding the optimal monotone policy and the novel class-ordered monotone policy. Our findings generate many key insights on these problems and their application to hypertension treatment.

Our analysis of the optimal monotone and class-ordered monotone policies shows that, in general, both policies depend on the initial state distribution. Moreover, the problem of finding the optimal monotone policy can be solved in polynomial time for many types of medical decision-making problems, although the overall number of policies can still be prohibitively large. To this end, we derived exact MIP-based formulations to identify a broad class of interpretable policies for MDPs, including the optimal monotone and class-ordered monotone policies. These formulations are amenable to both patient- and population-level treatment planning. Further, we defined and analyzed the price of interpretability, finding

that the class-ordered monotone policy pays a lower price of interpretability compared with the optimal monotone policy under mild conditions.

In our numerical analysis, we studied the implications of interpretable hypertension treatment plans at a patient level and a population level. Our treatment strategies led to better health outcomes compared with the current clinical guidelines across all BP categories, indicating that the clinical guidelines may be undertreating some patients and overtreating others. To some degree, these findings may be owed to the facts that our treatment strategies are informed by risk and consider the patient's expected future health status, whereas the clinical guidelines are mainly driven by BP levels. Further, our interpretable policies better matched clinicians' intuition compared with the optimal MDP policy, with only minor negative consequences in a large population of adults in the United States.

This work can be extended in several ways. First, our analysis of interpretable policies focused on monotone policies and class-ordered monotone policies. Future work may propose other interpretable policies for MDPs and analyze the price of interpretability paid by each type of policy. Second, we formulated MIPs that can exactly determine optimal interpretable policies, including the monotone policy and the class-ordered monotone policy. However, these MIPs may be computationally prohibitive for large problem sizes. Future research can investigate computationally efficient algorithms for these problems. The clinical component of this research could be extended by incorporating comorbidities, such as high cholesterol or diabetes. Given its impressive flexibility, we hypothesize that integrating the treatment of multiple conditions will likely increase benefits from our class-ordered monotone policy over current guidelines. Additionally, measurement error could limit the accuracy of our policies. As such, designing policies that are both interpretable and robust to uncertainty in parameter

estimates would be a promising direction for future research. Another crucial form of error is the impact of race and sex on clinical outcomes. Race and gender biases in the calculation of the risk for ASCVD events may alter cardiovascular outcomes, which could propagate to our treatment recommendations. This vital problem is out of the scope of this work and merits follow-up dedicated to addressing it specifically.

Overall, this research provides an optimization-based approach to interpretable treatment policy design with MDPs. We demonstrate that in complex environments, such as personalized hypertension treatment planning, our interpretable policies can drastically outperform existing guidelines while recommending treatments that are clinically intuitive. As such, these policies have great potential to facilitate the implementation of MDP-guided recommendations in practice, with applications in medical decision making and beyond.

Acknowledgments

References

Arias E (2019) United States life tables, 2017. *National Vital Statist. Rep.* 68(7):1–66.

Ayer T, Zhang C, Bonifonte A, Spaulding AC, Chhatwal J (2019) Prioritizing hepatitis C treatment in U.S. prisons. *Oper. Res.* 67(3):853–873.

Bellman R, Glicksberg I, Gross O (1955) On the optimal inventory equation. *Management Sci.* 2(1):83–104.

Boloori A, Saghafian S, Chakkera HA, Cook CB (2020) Data-driven management of post-transplant medications: An ambiguous partially observable Markov decision process approach. *Manufacturing Service Oper. Management* 22(5):1066–1087.

Bonifonte A, Ayer T, Haaland B (2022) An analytics approach to guide randomized controlled trials in hypertension management. *Management Sci.* 68(9):6634–6647.

Brønnnum-Hansen H, Jørgensen T, Davidsen M, Madsen M, Osler M, Gerdes LU, Schroll M (2001) Survival and cause of death after myocardial infarction: The Danish MONICA study. *J. Clinical Epidemiology* 54(12):1244–1250.

Burn J, Dennis M, Bamford J, Sandercock P, Wade D, Warlow C (1994) Long-term risk of recurrent stroke after a first-ever stroke. The Oxfordshire Community Stroke Project. *Stroke* 25(2): 333–337.

Capan M, Khojandi A, Denton BT, Williams KD, Ayer T, Chhatwal J, Kurt M, et al. (2017) From data to improved decisions: Operations research in healthcare delivery. *Medical Decision Making* 37(8):849–859.

Cevik M, Ayer T, Alagoz O, Sprague BL (2018) Analysis of mammography screening policies under resource constraints. *Production Oper. Management* 27(5):949–972.

Chen Q, Ayer T, Chhatwal J (2018) Optimal M-switch surveillance policies for liver cancer in a hepatitis C-infected population. *Oper. Res.* 66(3):673–696.

Cohen JB, Townsend RR (2018) The ACC/AHA 2017 hypertension guidelines: Both too much and not enough of a good thing? *Ann. Internal Medicine* 168(4):287–288.

Denton BT, Alagoz O, Holder A, Lee EK (2011) Medical decision making: Open research challenges. *IIE Trans. Healthcare Systems Engrg.* 1(3):161–167.

Denton BT, Kurt M, Shah ND, Bryant SC, Smith SA (2009) Optimizing the start time of statin therapy for patients with diabetes. *Medical Decision Making* 29(3):351–367.

Goff DC, Lloyd-Jones DM, Bennett G, Coady S, D'agostino RB, Gibbons R, Greenland P, et al. (2014) 2013 ACC/AHA guideline on the assessment of cardiovascular risk: A report of the American College of Cardiology/American Heart Association task force on practice guidelines. *J. Amer. College Cardiology* 63(25 part B):2935–2959.

Grand-Clément J, Chan C, Goyal V, Chuang E (2021) Interpretable machine learning for resource allocation with application to ventilator triage. Preprint, submitted October 21, https://arxiv.org/abs/2110.10994.

Hicklin K, Ivy JS, Payton FC, Viswanathan M, Myerse E (2018) Exploring the value of waiting during labor. *Service Sci.* 10(3): 334–353.

Ioannidis JP (2018) Diagnosis and treatment of hypertension in the 2017 ACC/AHA guidelines and in the real world. *JAMA* 319(2): 115–116.

Kochanek KD, Murphy SL, Xu J, Arias E (2019) Deaths: Final data for 2017. *National Vital Statist. Rep.* 68(9):1–18.

Kohli-Lynch CN, Bellows BK, Thanassoulis G, Zhang Y, Pletcher MJ, Vittinghoff E, Pencina MJ, Kazi D, Sniderman AD, Moran AE (2019) Cost-effectiveness of low-density lipoprotein cholesterol level—Guided statin treatment in patients with borderline cardiovascular risk. *JAMA Cardiology* 4(10):969–977.

Kurt M, Denton BT, Schaefer AJ, Shah ND, Smith SA (2011) The structure of optimal statin initiation policies for patients with type 2 diabetes. *IIE Trans. Healthcare Systems Engrg.* 1(July):49–65.

Lakkaraju H, Rudin C (2017) Learning cost-effective and interpretable treatment regimes. Singh A, Zhu Jerry, eds. *Proc. 20th Internat. Conf. Artificial Intelligence Statist. AISTATS 2017*, vol. 54 (PMLR), 166–175.

Law MR, Morris JK, Wald NJ (2009) Use of blood pressure lowering drugs in the prevention of cardiovascular disease: Meta-analysis of 147 randomised trials in the context of expectations from prospective epidemiological studies. *BMJ* 338:b1665.

Law MR, Wald NJ, Morris JK, Jordan RE (2003) Value of low dose combination treatment with blood pressure lowering drugs: Analysis of 354 randomised trials. *BMJ* 326(7404):1427.

Lee E, Lavieri MS, Volk M (2018) Optimal screening for hepatocellular carcinoma: A restless bandit model. *Manufacturing Service Oper. Management* 21(1):198–212.

Littman ML (1994) Memoryless policies: Theoretical limitations and practical results. *From Animals Animats 3 Proc. Third Internat. Conf. Simulation Adaptive Behav.*, vol. 3 (MIT Press, Cambridge, MA), 238.

Lovejoy WS (1987) Some monotonicity results for partially observed Markov decision processes. *Oper. Res.* 35(5):736–743.

Marrero WJ, Lavieri MS, Sussman JB (2021) Optimal cholesterol treatment plans and genetic testing strategies for cardiovascular diseases. *Health Care Management Sci.* 24(1):1–25.

Mason JE, Denton BT, Shah ND, Smith SA (2014) Optimizing the simultaneous management of blood pressure and cholesterol for type 2 diabetes patients. *Eur. J. Oper. Res.* 233(3):727–738.

Meraklı M, Küçükyavuz S (2020) Risk aversion to parameter uncertainty in Markov decision processes with an application to slow-onset disaster relief. *IISE Trans.* 52(8):811–831.

Neumann P, Sanders G, Russell L, Siegel J (2016) *Cost-Effectiveness in Health and Medicine* (Oxford University Press, Oxford, UK).

Pandya A, Sy S, Cho S, Weinstein MC, Gaziano TA (2015) Cost-effectiveness of 10-year risk thresholds for initiation of statin therapy for primary prevention of cardiovascular disease. *JAMA* 314(2): 142–150.

Petrik M, Luss R (2016) Interpretable policies for dynamic product recommendations. Ihler A, Janzing D, eds. *32nd Conf. Uncertainty Artificial Intelligence 2016 UAI 2016* (AUAI Press), 607–616.

Puterman ML (2014) *Markov Decision Processes: Discrete Stochastic Dynamic Programming* (John Wiley & Sons, New York).

Rao A, Shi Z, Ray KN, Mehrotra A, Ganguli I (2019) National trends in primary care visit use and practice capabilities, 2008–2015. *Ann. Family Medicine* 17(6):538–544.

Rudin C, Chen C, Chen Z, Huang H, Semenova L, Zhong C (2022) Interpretable machine learning: Fundamental principles and 10 grand challenges. *Statist. Surveys* 16:1–85.

Saville CE, Smith HK, Bijak K (2019) Operational research techniques applied throughout cancer care services: A review. *Health Systems (Basingstoke)* 8(1):52–73.

Schäl M (1976) On the optimality of (*s*, *S*)-policies in dynamic inventory models with finite horizon. *SIAM J. Appl. Math.* 30(3): 528–537.

Schell GJ, Marrero WJ, Lavieri MS, Sussman JB, Hayward RA (2016) Data-driven Markov decision process approximations for personalized hypertension treatment planning. *MDM Policy Practice* 1(1):2381468316674214.

Serfozo RF (1976) Monotone optimal policies for Markov decision processes. Wets RJB, eds. *Stochastic Systems: Modeling, Identification and Optimization, II*, Mathematical Programming Studies, vol. 6 (Springer, Berlin), 202–215.

Serin Y, Kulkarni VG (1995) Implementable policies: Discounted cost case. Stewart WJ, eds. *Computations with Markov Chains* (Springer, Boston), 283–306.

Sethi T, Kalia A, Sharma A, Nagori A (2020) *Interpretable Artificial Intelligence: Closing the Adoption Gap in Healthcare* (Elsevier, Amsterdam).

Skandari MR, Shechter SM (2021) Patient-type Bayes-adaptive treatment plans. *Oper. Res.* 69(2):574–598.

Solberg LI, Miller WL (2018) The new hypertension guideline: Logical but unwise. *Family Practice* 35(5):528–530.

Steimle LN, Kaufman DL, Denton BT (2021) Multi-model Markov decision processes. *IISE Trans.* 53(10):1124–1139.

Suen S-C, Brandeau ML, Goldhaber-Fiebert JD (2018) Optimal timing of drug sensitivity testing for patients on first-line tuberculosis treatment. *Health Care Management Sci.* 21(4):632–646.

Sussman J, Vijan S, Hayward R (2013) Using benefit-based tailored treatment to improve the use of antihypertensive medications. *Circulation* 128(21):2309–2317.

Tjoa E, Guan C (2021) A survey on explainable artificial intelligence (XAI): Toward medical XAI. *IEEE Trans. Neural Networks Learn. Systems* 32(11):4793–4813.

Tunç S, Alagoz O, Burnside ES (2022) A new perspective on breast cancer diagnostic guidelines to reduce overdiagnosis. *Production Oper. Management* 31(5):2361–2378.

Van Der Wardt V, Harrison JK, Welsh T, Conroy S, Gladman J (2017) Withdrawal of antihypertensive medication: A systematic review. *J. Hypertension* 35(9):1742–1749.

Virani SS, Alonso A, Benjamin EJ, Bittencourt MS, Callaway CW, Carson AP, Chamberlain AM, et al. (2020) Heart disease and stroke statistics—2020 update: A report from the American Heart Association. *Circulation* 141(9):e139–e596.

Vlassis N, Littman ML, Barber D (2012) On the computational complexity of stochastic controller optimization in POMDPs. *ACM Trans. Comput. Theory* 4(4):12.

Wang F, Kaushal R, Khullar D (2020) Should health care demand interpretable artificial intelligence or accept "black box" medicine? *Ann. Internal Medicine* 172(1):59–60.

Whelton PK, Carey RM, Aronow WS, Casey DE, Collins KJ, Dennison Himmelfarb C, DePalma SM, et al. (2018) 2017 ACC/ AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/ PCNA Guideline for the Prevention, Detection, Evaluation, and Management of High Blood Pressure in Adults: A report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. *J. Amer. College Cardiology* 71(19):e127–e248.