Interpretable Policies and the Price of Interpretability in Hypertension Treatment Planning

Gian-Gabriel P. Garcia[1], Lauren N. Steimle[1], Wesley J. Marrero[2], and Jeremy B. Sussman[3]

[1]H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA

[2]Thayer School of Engineering, Dartmouth College, Hanover, NH

[3]Department of Internal Medicine, Michigan Medicine, University of Michigan, Ann Arbor, MI

Problem definition: Effective hypertension management is critical to reducing 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 many complexities in treatment planning. However, model-generated recommendations may be uninterpretable/unintuitive, limiting their acceptability in practice. We address this challenge by investigating interpretable treatment planning.

Methodology/results: We study interpretable treatment planning for Markov Decision Processes (MDPs) and specifically analyze the problems of optimizing *monotone policies*, which increase treatment intensity with worsening patient health, and optimizing *class-ordered monotone policies (CMPs)*, which generalize monotone policies by imposing monotonicity over state and action classes rather than states and actions. We establish that both policies depend on initial state distributions. Furthermore, optimal monotone policies can be generated tractably for many treatment planning problems. Next, we propose exact formulations for optimizing interpretable policies broadly. Then, we define and analyze the *price of interpretability (PI)*, proving that the CMP's PI does not exceed the monotone policy's. Finally, we formulate, parameterize, and evaluate MDPs for 10-year hypertension treatment planning using a large, nationally representative dataset of the United States population. We compare the structure and performance of optimal monotone policies and CMPs to 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 CMPs never de-escalate treatment, reflecting clinical intuition. Across 66.5 million patients, optimized monotone policies and CMPs outperform clinical guidelines, saving over 3,246 quality-adjusted life years per 100,000 patients, while paying low PIs. Sensitivity analysis further reveals that monotone policies and CMPs are robust to various definitions of "interpretability."

Managerial implications: Interpretable policies can be tractably optimized, drastically outperform existing guidelines, and pay low PIs – potentially increasing the acceptability of decision-analytic approaches in practice.

Key words: Markov decision processes, healthcare applications, medical decision-making, interpretability, cardiovascular disease, personalized treatment planning

1. Introduction

Atherosclerotic cardiovascular disease (ASCVD), which constitutes coronary heart disease (CHD) and stroke, is a leading cause of death in the United States (Kochanek et al. 2019). Recent reports show CHD and stroke account for 42.6% and 17.0% of deaths due 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 which affects 45.6% of adults in the United States (Whelton et al. 2018) and puts these individuals at a higher risk of experiencing CHD 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 who 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, have been met with substantial backlash (Ioannidis 2018). In contrast, decision-analytic approaches may better capture these risks, benefits, and uncertainties, and 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 counter-intuitive (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 for healthcare (Tjoa and Guan 2021). Yet, interpretable decision-analytic methods remain under-developed. To increase the acceptability of decision-analytic models in clinical practice, it is critical to extend these methods for *interpretable treatment planning*. In interpretable treatment planning, a decision maker (DM) 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. Due to their ubiquity and wide adoption in the medical decision-making literature, we focus our attention on interpretable treatment planning using Markov Decision Processes (MDPs).

1.1. MDPs and Interpretable Policies

Surveys by Denton et al. (2011), Capan et al. (2017), and Saville et al. (2018) comprehensively review many MDPs and Partially Observable MDPs (POMDPs) in the medical decision-making literature. Recent examples include research by Chen et al. (2018), Hicklin et al. (2018), Suen et al. (2018), Cevik et al. (2018), Agnihothri et al. (2018), Lee et al. (2018), Ayer et al. (2019), Boloori et al. (2020), Skandari and Shechter (2021), Steimle et al. (2021), Marrero et al. (2021), Tunç et al. (2022) and Bonifonte 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. Moreover, our methods can handle complex treatment options (e.g., multiple medications and dosages, see §3.1), extending beyond the binary treatments (e.g., initiate treatment/wait) considered in many prior works.

Among POMDP applications in treatment planning, Chen et al. (2018) and Cevik et al. (2018) design interpretable policies for POMDPs in the context of cancer screening. 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 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. While our research focuses on interpretable policies for MDPs, we note that the policy imposed by Cevik et al. (2018) is a special case of our novel class-ordered monotone policy (CMP).

Beyond healthcare, 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 which 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 which guarantee the existence of an optimal policy that is monotone (Puterman 2014, §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 CMP, which can be interpretable when the states and/or actions do not follow a strict ordering.

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 studied the class of *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 DMs depends only on the most recent observation, rather than the entire history of observations and actions.

Perhaps the most closely related works to ours is that of Petrik and Luss (2016) and Serin and Kulkarni (1995) who consider interpretable policies in fully observed MDPs. Petrik and Luss

(2016) and Serin and Kulkarni (1995) 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 DM'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 showed that the optimal policy depends on the initial distribution over the states and proposed an iterative method for finding local optimal solutions. Petrik and Luss (2016) later showed that solving for randomized or deterministic interpretable policies is NP-hard and proposed a mixed-integer program (MIP) to solve this problem. In this article, we propose the CMP, 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 CMPs 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 *CMP* — a new type of interpretable policy that generalizes monotone policies. We find that the optimal CMP 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 policies and CMPs. We show that these methods are amenable to both patient- and population-level treatment planning.
3. We introduce the *price of interpretability (PI)* in MDPs, which measures the difference between the optimal total discounted reward of the MDP and the best interpretable policy. This metric

can guide DMs who wish to know the cost of using the best interpretable policy instead of the optimal policy. We show that under mild conditions, the optimal CMP is guaranteed to perform no worse than the optimal monotone policy under this metric.

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 CMP perform close to optimal (i.e., low PIs) while being more clinically intuitive. Moreover, both types of interpretable policies drastically outperform current guidelines, saving over 3,246 quality-adjusted life years per 100,000 patients.
5. Through a sensitivity analysis in hypertension treatment planning, we show that both the optimal monotone policy and optimal CMP 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 §2, we introduce the interpretable treatment planning problem and specifically analyze optimal monotone policies and CMPs. We then derive an exact MIP formulation to solve the interpretable treatment planning problem, followed by an analysis of the PI. In §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 CMPs compared to the optimal MDP policy and current clinical guidelines. Finally, in §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 CMP. 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 CMPs. Finally, we define the PI and compare the CMP and optimal monotone policy with respect to the PI. Proofs for all technical results are provided in §EC.1 of the e-companion. All of these methods can be flexibly adapted to finite horizon MDPs by adding a temporal component (see §EC.3.1 in 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 §3).

2.1. Treatment Planning with MDPs

Our infinite horizon MDP formulation for the treatment planning problem is as follows. A patient's health is modeled as a Markov Chain with health states $\mathcal{S} = \{1, \ldots, S\}$. At each decision epoch $t \in \mathcal{T} = \{0, 1, \ldots\}$, a DM 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 DM 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 DM 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 which 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 (Puterman 2014, Ch. 6.9) that there exists a deterministic and stationary policy that is optimal. Moreover, $J^{\pi^*}(\alpha) = \sum_{s\in\mathcal{S}} \alpha(s)v^*(s)$ where $[v^*(s)]_{s\in\mathcal{S}}$ are given by the optimal solution to the following linear program:

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

where $\mathbf{v} := [v(s)]_{s\in\mathcal{S}}$ is a vector of state value functions. For each $s \in \mathcal{S}$, $\pi^*(s)$ is equal to the action where $v^*(s)$ is tight. Note that π^* is independent of α due to the principle of optimality.

2.2. Interpretable Treatment Planning with MDPs

Suppose that the DM 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

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

As discussed in §1.1, there are many types of interpretable policies. In the following subsections, we focus our attention on monotone policies and the novel CMP.

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 $\pi(s) \le \pi(s')$ for all $s, s' \in \mathcal{S}$ such that $s \le s'$.

Monotone policies are appealing to practitioners because they can be more easily interpreted and implemented compared to 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 §1.1) have aimed to identify sufficient conditions on the MDP data (i.e., P and r) which guarantee that there exists an optimal policy which is monotone, i.e., $\pi^* \in \Pi^M$. In contrast, our aim is to determine the policy $\pi^M \in \Pi^M$ which 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 which 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 since 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 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. Furthermore, 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 risk score, the American College of Cardiology/American Heart Association risk score, 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 number of available medications have increased over time.

Although monotone policies are highly desirable due to 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 multi-dimensional state and action spaces. While the issue of non-orderability is beyond our scope, we recognize that the DM may feel more comfortable by defining orderings on groups of states and groups of actions, e.g., groupings defined by risk categories. In §2.2.2, 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.

2.2.2. Class-ordered Monotone Policies The *class-ordered monotone policy (CMP)* 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 $\mathcal{S}_1, ..., \mathcal{S}_K$ indexed by the set $\mathcal{K} = \{1, \ldots, K\}$ and $\mathcal{A}$ is partitioned into ordered action classes $\mathcal{A}_1, ..., \mathcal{A}_G$ 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 $s' \in S_{k'}$ is "more severe" than any state $s \in S_k$ (with similar interpretation for action classes). States and actions within a class are not required to be ordered. Using this construction, we now define CMPs.

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

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

Let $\Pi^{CM}_{\Theta,\Psi}$ be the set of CMPs with respect to Θ and Ψ. The optimal CMP is the solution to:

$$\pi^{CM} = \underset{\pi \in \Pi^{CM}_{\Theta,\Psi}}{\arg\max}\, J^{\pi}(\alpha). \tag{4}$$

Like π^M, the policy π^{CM} is also dependent on α in general (see §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 CMPs. 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

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

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

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

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

$$\mathbf{x} \in \mathcal{X}_0 \tag{5c}$$

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

Objective (5a) differs from the objective in (1) because of the epigraph constraints (5b). In general, big-M constraints like (5b) weaken MIP formulations. Since 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 $r(s,a) \geq 0$ 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). Steimle et al. (2021) and Meraklı and Küçükyavuz (2020) 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 π^M and π^{CM}. To obtain π^M, we replace constraint (5d) with

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

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

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

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

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

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

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

2.4. The Price of Interpretability

A DM who values interpretability may ask, "what is the cost of implementing the best interpretable policy instead of the best overall policy?" Thus, we introduce the *price of interpretability (PI)*:

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

The PI informs the DM about the cost of implementing a policy from the interpretable policy class Π^I rather than implementing the optimal policy π^*. To facilitate our analysis of the PI for optimal CMPs relative to optimal monotone policies, we make the following assumption:

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

Assumption 1 restricts our attention to the class of CMPs with respect to 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 and thus, a partial ordering may be more appropriate. For example, research suggests that the benefit of antihypertensive treatment is mainly determined by their BP reduction with little effect attributable to drug-specific factors (Law et al. 2009). While 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 CMPs.

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 CMP over a monotone policy; if a strict ordering is ambiguous, and state and/or action classes are more intuitive, then π^{CM} can achieve a lower PI than π^M 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 §EC.2.3 of the e-companion.

3. Numerical Analysis: Personalized Hypertension Treatment

We now apply the optimal monotone policy (π^M) and the optimal CMP (π^{CM}) to the management of atherosclerotic cardiovascular disease (ASCVD). We begin by describing our MDP, model parameters, data sources, and treatment strategies. Next, we present the treatment plans and health outcomes of patients following the optimal policy (π^*), optimal CMP (π^{CM}), optimal monotone policy (π^M), and the current clinical guidelines (π^G). Finally, we discuss the implications of the interpretable hypertension treatment plans at a patient and a population level.

3.1. Markov Decision Process Formulation

Since 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 §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 §EC.3.2 of the e-companion). The elements of our MDP are as follows:

- $\mathcal{T}'$: planning horizon of 10 years; $\mathcal{T}' \coloneqq \{0, 1, \ldots, T\}$. Decision epoch $t \in \mathcal{T}'$ represents the year $[t, t+1)$ and $T-1$ is the year at which physicians select the last action. We use $T = 10$ to represent the effects of treatment on patients' lifetime. This planning horizon is selected based on conversations with our clinical collaborators and following the major clinical guidelines for the management of cardiovascular diseases (Whelton et al. 2018).
- $\mathcal{S}$: state space comprising patients' 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 CHD 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 CHD and stroke but no adverse event in the current year ($h_t = 4$), survival of a CHD 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 CHD 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.
- $\mathcal{A}$: action space composed of 0 to 5 antihypertensive medications of five different drug types at their standard dose. Among the types of antihypertensive medications, we include the following: thiazide diuretics, beta-blockers (BBs), calcium channel blockers (CCBs), angiotensin-converting enzyme (ACE) inhibitors, and angiotensin II receptor blockers (ARBs). Since 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).

- $p_t(s_{t+1}|s_t, a_t)$: transition probability derived from patients' 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 and Xu 2019). Based on communications with clinical collaborators, we assume independence among CHD and stroke events. CHD 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 CHD or stroke events if they have a history of such ASCVD events (Schell et al. 2016). We account for this assumption by adjusting patients' CHD and stroke odds if they have a history of either ASCVD event (Brønnnum-Hansen et al. 2001, Burn et al. 1994).
- $r_t(s_t, a_t)$: patients' reward 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 (Kohli-Lynch et al. 2019, Law et al. 2003). Terminal rewards $r_T(s_T)$ represent patients' total quality-adjusted life years (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 and Xu 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).
- $\alpha(s_t)$: initial state distribution used to represent patients' 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.
- γ: discount factor of the model. 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 EC.3.3.

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 π^* is achieved 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 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 (resp., stage 2) hypertension as a systolic BP of 130-139 mm Hg (resp., at least 140 mm Hg) or diastolic BP of 80-89 mm Hg (resp., 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 QALYs. This optimization model follows formulation (1) with additional constraints to guarantee that the number 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 π^{G}.

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, physical examinations, and 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 leads 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. These estimates are used 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 the states of each patient based on their associated risk for ASCVD events. At each decision epoch t, the states of every patient share their 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 the 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 CHD 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. (2009, 2003). 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, CCBs, thiazides, BBs, 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. But since the difference between the drugs is small, 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 $i = 1, ..., 5$. Note that our initial selection of Θ and Ψ satisfy 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 π^{CM} and π^{M}. We then study the trade-off between optimality and interpretability at a population level by comparing our policies to π^* and π^G. We begin by inspecting the number and type of medications recommended by each treatment strategy. Subsequently, we assess the QALYs saved and ASCVD events prevented by our policies, compared to the clinical guidelines. Lastly, we inspect the PI for π^M, π^{CM}, and π^G. To compare if PIs among the policies are statistically different, we use Wilcoxon Signed Rank Tests with a significance level of 0.05. The significance level and confidence interval (CI) 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. While patients in the NHANES dataset 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 on §EC.3.4 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 π^{CM} and π based on the PI, number of ASCVD events allowed compared to π^*, and the average number of medications recommended. Any record with an optimization model exceeding the time limit in any scenario was excluded from all sensitivity analysis scenarios. The complete dataset 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 subsection, we examine and describe the effect of the interpretable hypertension treatment plans. We provide insights into the patient- and population-level results in §3.7.

3.5.1. Patient-Level Results. We now evaluate π^{CM} and π^M in a series of patient profiles. For comparison purposes, we also determine π^* and π^G for each patient profile. We first obtain treatment plans for the following patient profile: a 45-year-old, non-diabetic, non-smoker 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 π^{CM} and π^M as well as π^* over the health conditions of our selection of patient profiles at 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. Thus, there is no PI associated with this profile.

Increasing the base profile's arterial pressure level to elevated BP or stage 1 hypertension leads to the suggestion of one ARB at standard dose when the patient has no history of ASCVD in all

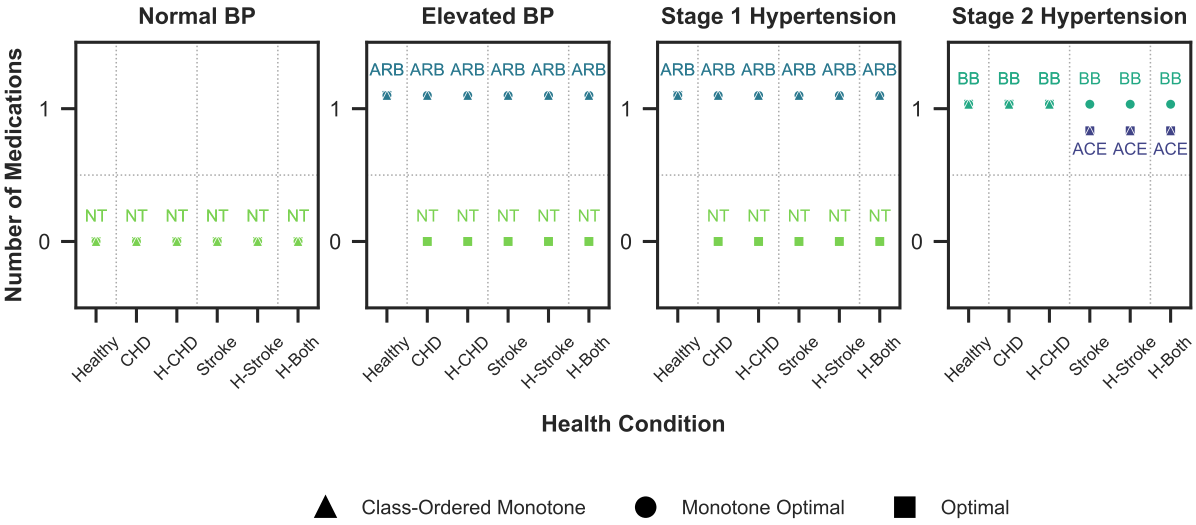

Figure 1 **Treatment policies over the health conditions of selected patient profiles at the last year of our study. The area below and above the horizontal dotted lines represent action classes $\mathcal{A}_0$ and $\mathcal{A}_1$, respectively. The state classes $\mathcal{S}_1$, $\mathcal{S}_2$, $\mathcal{S}_3$, and $\mathcal{S}_4$ are separated by vertical dotted lines. The label "H-" denotes the state classes associated with a history of ASCVD events. NT: No treatment; ACE: ACE inhibitor; ARB: Angiotensin II receptor blockers; BB: Beta blocker.**

policies. The optimal policy π^* 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, π^M and π^{CM} account for interpretability aspects by recommending one ARB across all states. For these patient profiles, there is no considerable consequence for providing interpretability as the PI is less than 0.0001 QALYs.

When the patient profile has stage 2 hypertension, all the strategies recommend a BB at standard dose if the patient has no history for ASCVD events or has ever survived a CHD event. If the patient has ever survived a stroke, π^{CM} and π^* suggest prescribing one ACE inhibitor. Conversely, since ACE inhibitors have a smaller expected risk reduction than BBs, π^M continues to recommend one BB. While there is no PI associated with π^{CM}, the PI of π^M is positive for this patient profile.

Figure 1 excludes π^G to ease the comparison between π^*, π^{CM}, and π^M. The clinical guidelines π^G 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 CHD-related states. The clinical guidelines suggest one ACE inhibitor for the states associated with stroke events.

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 correspond to patients in the first year of our study.

Treatment Recommendations. The distribution of treatment recommended by each policy per BP category in states associated with no history for ASCVD at years 0 and 9 of our study is shown in Figure EC.2 of §EC.3.5 of the e-companion. Other than more intense treatment over time, we note that the distribution of treatment did not change considerably in years 1 through 8.

From the distribution of treatment recommendations, we observe that virtually no patient receives treatment in the normal BP category at any given year. Comparing our interpretable policies to π^* and π^G, we notice that π^{CM} and π^M are often close to optimal. We also note that our interpretable policies are typically more intense than π^G for patients with elevated BP and stage 1 hypertension. For patients with stage 2 hypertension, π^G mostly prescribes two to three medications, whereas our interpretable policies fluctuate more broadly from two to four medications.

Across medications prescribed, the treatment strategies behave similarly from one to three medications at standard dose. For example, the most frequent medication at standard dose is one ARB across all treatment strategies over time. Two doses of an ARB are prescribed more commonly than two doses of any other drug type or two-drug combination. Similarly, three doses of an ARB is prescribed more often than any other three-drug or three-dose combination. The variation among the drug combinations recommended by the treatment strategies is substantially higher when four medications are prescribed. However, the suggestions from π^{CM} and π^M are often close to the recommendations from π^*. For instance, the most common four-drug combinations by our strategies are four doses of an ARB. The most common drug combination by π^G is three doses of an

ARB and one BB at standard dose. Five doses of an ARB are prescribed more regularly than five doses of any other drug type or five-drug combination. The number of times each medication type is prescribed as part of a drug combination is summarized in Figure EC.3 in the e-companion. Overall, the clinical guidelines recommend no treatment much more frequently compared to our policies in patients with elevated BP and stage 1 hypertension. In contrast, our policies frequently prescribe at least one ARB at standard dose across all drug combinations in each BP category.

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 strategy in terms of the number of QALYs saved and ASCVD events prevented, compared to the clinical guidelines. In total, π^*, π^{CM}, and π^M save 3,262.01, 3,246.62 and 3,246.46 QALYs per 100,000 patients over the planning horizon, compared to the clinical guidelines. We notice a similar pattern when comparing the policies in terms of ASCVD events averted. Over the 10-year planning horizon, π^*, π^{CM}, and π^M prevent 305.98, 305.10, and 305.09 ASCVD events per 100,000 patients, compared to π^G.

Patients with stage 2 hypertension receive the greatest benefit from treatment. We note that patients' health outcomes under π^{CM} and π^M are not substantially different for patients with elevated BP or stage 1 hypertension. In people with stage 2 hypertension, π^{CM} saves 0.87 QALYs and averts 0.04 ASCVD events more than π^M per 100,000 patients. The clinical guidelines π^G are outperformed by our treatment strategies in every BP category. Our policies provide the greatest benefit to patients with elevated BP and stage 1 hypertension, when compared to π^G (see Figure EC.4 in §EC.3.5 of the e-companion).

Price of Interpretability. Overall, the PI of π^{CM}, π^M, and π^G are 15.38, 15.55, and 3,262 QALYs per 100,000 patients, respectively — an immediate consequence of the difference in the total QALYs saved between each treatment strategy and π^*. Figure 2 illustrates the PI corresponding to π^{CM} and π^M per BP category, with the normal BP category and outliers above the 99th percentile of the PI in each BP category excluded for illustration purposes. We also show the PI associated with every patient in our dataset following π^{CM} and π^M in Figure EC.5 of §EC.3.5.1 of the e-companion.

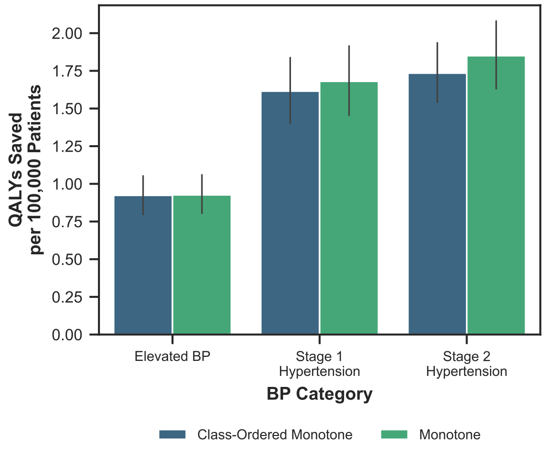

Figure 2 PI associated with our interpretable policies per BP category. Error bars represent the 95% bootstrap confidence intervals around the PI per 100,000 patients using 10,000 replications.

The PI for π^{CM} and π^{M}, and the difference between the two, generally increase with patients' BP. Using Wilcoxon Signed Rank Tests, there is enough evidence to conclude that the PI of π^{M} is significantly greater than the PI of π^{CM} across all patients (95% CI $[1^{-6}, \infty], P = 3^{-8}$). The PI of π^{M} is also significantly greater than the PI of π^{CM} 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). Since we are comparing three BP categories simultaneously, statistical significance was determined using a Bonferroni threshold of 0.02.

The PI of π^{G} is 5,603, 7,392, and 2,451 QALYs per 100,000 patients with elevated BP, stage 1 hypertension, and stage 2 hypertension, respectively (not shown in Figure 2). Similar to our findings in terms of QALYs saved and ASCVD events prevented, we observe that π^{*} provides the greatest benefit over π^{G} for patients with elevated BP and stage 1 hypertension.

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 1. Note that the PI and number of ASCVD events allowed in the base case are different than in our main analysis. This difference is due to 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 due to the time limit, while in our sensitivity analyses we are excluding 13.03 million patients (see §EC.3.5.2 in the e-companion for details). This exclusion allows us to compare the performance of the policies across the sensitivity analysis scenarios.

Table 1 Sensitivity analyses summary. All results are presented as the average per 100,000 patients.

Scenario	PI		ASCVD Events Allowed[a]		Number of Medications[b]	
	CMP	MP	CMP	MP	CMP	MP
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

[a] Results correspond to the first year of our study. Values presented in thousands.
[b] Values represent year 0, year 4, and year 9 of our study, respectively, in thousands.

Ordering the states based on nonincreasing severity, merging the ASCVD events classes into a single class, or categorizing the actions according to their systolic BP reductions do not have substantial effects on the outcomes of the policies. Ordering the states on the basis of nonincreasing severity of the health conditions allows the interpretable policies to mimic π^* more closely, which results in marginal reductions in the PI. Combining the ASCVD events classes into a single class offers increased flexibility to π^{CM}, which results in minor reductions in the PI and the ASCVD events allowed. Grouping the actions based on their systolic BP reductions results in a small decrease in the number of ASCVD events allowed by π^{CM}. None of these scenarios dramatically change the number of medications prescribed over time.

Ordering and categorizing actions based on diastolic BP reductions changes the outcomes of the interpretable policies noticeably. By treating patients less aggressively, π^M results in markedly higher PI and ASCVD events allowed. Ordering the actions in line with their diastolic BP reductions greatly limits the efficacy of π^M. Furthermore, this action ordering and classification leads to worse health outcomes if patients follow π^{CM}. However, the effect of ordering and classifying the actions based on their diastolic BP reductions is much smaller on π^{CM} than on π^M.

Changing the initial state distribution also has sizable consequences on the interpretable policies. Recall that π^M and π^{CM} generally depend on the initial state distribution (see Proposition 1 and Remark EC.1). We find that the PI and number of ASCVD events allowed by both treatment strategies are higher in the scenarios with modified initial state distribution weights. In these scenarios, the PI is considering the rewards associated with ASCVD events, besides the rewards related to the healthy condition. The effect of suboptimal treatment (due to interpretability constraints) is amplified when these rewards are considered. Furthermore, in these scenarios our interpretable policies are more conservative than in the base case. Lower treatment intensity results in considerable increases in the number of ASCVD events allowed.

3.7. Discussion and Implications

In this subsection, we provide potential explanations for our findings and discuss their implications.

3.7.1. Patient-level Implications. We make two key observations at the patient level. First, π^* 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, π^* 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, in turn, the rewards is typically smaller than the treatment-related disutility in the states associated with ASCVD events. Second, π^M typically recommends to keep a constant treatment across all health conditions in each year of the planning horizon. Similarly, π^{CM} 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, π^{CM} generally prescribes a drug combination with lower risk reductions and treatment-related disutilities. For example, π^{CM} may prescribe ACE inhibitors instead of BBs. These patterns align with intuition of physicians in practice. Hence, π^{CM} and π^M provide more intuitive strategies than π^* 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.

3.7.2. Population-level Implications. We find several major trends at the population level. First, π^{CM}, π^{M}, and π^{*} prescribe a similar number of medications in states associated with no history for ASCVD events. A reason for this finding is that although our interpretable policies are constrained to be nondecreasing over time, this restriction is not often violated by the optimal policy. Second, all policies tend to include ARBs in their treatment recommendations, potentially because ARBs have the largest expected ASCVD risk reductions with relatively low treatment-related disutilities. Third, the PI of π^{CM} and π^{M} generally increases with patients' BP. While π^{CM} and π^{M} would never decrease treatment intensity as patients' BP increases, π^{*} may decrease treatment aggressiveness in some situations. Thus, the monotonicity constraints may become more restrictive. The pairwise differences between the PI of π^{CM} and π^{M} tend to grow as patients' BP increases for a similar reason, likely because π^{M} is a more restrictive policy than π^{CM}. Fourth, the restrictiveness of π^{CM} 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 leads to less restrictions in π^{CM}. Finally, we find that π^{CM} offers intuitive treatment strategies to physicians with modest improvements over π^{M}. The optimal CMP π^{CM} results in similar health outcomes to π^{*} with the added benefits of interpretability.

3.7.3. Consequences of Changing Modeling Assumptions. In our sensitivity analyses, 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 PI and number of ASCVD events averted by our interpretable policies. For example, ordering the states in nonincreasing severity of health condition forces π^{CM} and π^{M} to prescribe at most as much medication in the states associated with ASCVD events as they are in the states related to the healthy condition. Recommending less medication to patients with a history of ASCVD events may prompt additional events. Ordering the actions according to their diastolic BP reductions has a larger impact on the health outcomes associated to our interpretable policies. This finding may arise because patients' diastolic BP is unnecessary for calculating risk of ASCVD events. The risk only considers patients' systolic BP, implying that the transition probabilities and rewards do not consider patients' diastolic BP. Even if an action is expected to have a high diastolic BP reduction, it may not lead to a high risk reduction. Changing the classification of the actions may have moderate to large effects. For instance, categorizing actions according to their diastolic BP reductions may restrict π^{CM} to classes with more intense treatment which may not 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.

4. Conclusions

MDPs are a powerful tool capable of capture the risks, benefits, and uncertainties inherent to for optimal treatment planning. Yet, their resulting optimal 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 CMP. Our findings generate many key insights on these problems and their application to hypertension treatment.

Our analysis of the optimal monotone policy and CMP 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, though 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 policy and CMP. These formulations are amenable to both patient- and population-level treatment planning. Further, we defined and analyzed the PI, finding that the CMP pays a lower PI compared to the optimal monotone policy under mild conditions.

In our numerical analysis, we studied the implications of interpretable hypertension treatment plans at a patient and a population level. Our treatment strategies led to better health outcomes compared to the current clinical guidelines across all BP categories, indicating that the clinical guidelines may be under-treating some patients and over-treating others. This finding may be due, in part, because our treatment strategies are informed by risk and consider the patient's expected future health status, while the clinical guidelines are mainly driven by BP levels. At the same time, our interpretable policies better matched clinicians' intuition compared to the optimal MDP policy with only minor negative consequences in a large population of adults in the US.

This work can be extended in several ways. First, our analysis of interpretable policies focused on monotone policies and CMPs. Future work may propose other interpretable policies for MDPs and analyze their respective PIs. Second, we formulated MIPs which can exactly determine optimal interpretable policies, including the monotone policy and CMP. 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 CMP over current guidelines. Additionally, measurement error could limit the accuracy of our policies. One 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 into practice, with applications in medical decision-making and beyond.

References

Agnihothri S, Cui L, Delasay M, Rajan B (2018) The value of mHealth for managing chronic conditions. *Health Care Management Science* .

Arias E, Xu J (2019) United States Life Tables, 2017. *National Vital Statistics Reports* 68(7).

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

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

Bhattacharya A, Kharoufeh JP (2017) Linear programming formulation for non-stationary, finite-horizon Markov decision process models. *Operations Research Letters* 45(6):570–574.

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 Operations Management* (February):msom.2019.0797.

Bonifonte A, Ayer T, Haaland B (2022) An Analytics Approach to Guide Randomized Controlled Trials in Hypertension Management. *Management Science* (February).

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. *Journal of 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):333–7.

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 and Operations 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. *Operations Research* 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? *Annals of Internal Medicine* 168(4):287.

Denton BT, Alagoz O, Holder A, Lee EK (2011) Medical decision making: open research challenges. URL http://dx.doi.org/10.1080/19488300.2011.619157.

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. *Journal of the American College of Cardiology* 63(25 Part B):2935–2959.

Govindan S, Shapiro L, Langa KM, Iwashyna TJ (2014) Death Certificates Underestimate Infections as Proximal Causes of Death in the U.S. *PLoS ONE* 9(5):3–6.

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

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

Kochanek KD, Murphy SL, Xu J, Arias E (2019) Deaths: final data for 2017. *National Vital Statistics Reports* 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 Transactions on Healthcare Systems Engineering* 1(July):49–65.

Lakkaraju H, Rudin C (2017) Learning cost-effective and interpretable treatment regimes. *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, AISTATS 2017* 54.

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):1427–0.

Lee E, Lavieri MS, Volk M (2018) Optimal screening for hepatocellular carcinoma: a restless bandit model. *Manufacturing & Service Operations Management* (January 2019):msom.2017.0697.

Littman ML (1994) Memoryless policies: Theoretical limitations and practical results. *From Animals to Animats 3: Proceedings of the third international conference on simulation of adaptive behavior*, volume 3, 238 (Cambridge, MA).

Lovejoy WS (1987) Some monotonicity results for partially observed markov decision processes. *Operations Research* 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 Science* .

Mason JE, Denton BT, Shah ND, Smith SA (2014) Optimizing the simultaneous management of blood pressure and cholesterol for Type 2 diabetes patients. *European Journal of Operational Research* 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 Transactions* 52(8):811–831.

NCHS (2017) Health, United States, 2016: with chartbook on long-term trends in health. *Center for Disease Control* 314–317.

Neumann P, Sanders G, Russell L, Siegel J (2016) *Cost-effectiveness in health and medicine* (Oxford University Press).

Neumann PJ, Cohen JT (2018) QALYs in 2018—Advantages and Concerns. *JAMA* 319(24):2473.

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. *Journal of the American Medical Association* 314(2):142–150.

Petrik M, Luss R (2016) Interpretable policies for dynamic product recommendations. *32nd Conference on Uncertainty in Artificial Intelligence 2016, UAI 2016*, 607–616, ISBN 9781510827806.

Puterman ML (2014) *Markov decision processes: discrete stochastic dynamic programming* (John Wiley & Sons).

Saville CE, Smith HK, Bijak K (2018) Operational research techniques applied throughout cancer care services: a review. *Health Systems* 6965:1–22.

Schäl M (1976) On the optimality of (s,s)-policies in dynamic inventory models with finite horizon. *SIAM Journal on Applied Mathematics* 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).

Serfozo RF (1976) Monotone optimal policies for markov decision processes. *Stochastic Systems: Modeling, Identification and Optimization, II*, 202–215 (Springer).

Serin Y, Kulkarni VG (1995) Implementable Policies: Discounted Cost Case. *Computations with Markov Chains* 283–306.

Sethi T, Kalia A, Sharma A, Nagori A (2020) *Interpretable artificial intelligence: Closing the adoption gap in healthcare* (Elsevier Inc.).

Skandari MR, Shechter SM (2021) Patient-Type Bayes-Adaptive Treatment Plans. *Operations Research* (March):opre.2020.2011.

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 Transactions* 53(10):1124–1139.

Suen Sc, Brandeau ML, Goldhaber-Fiebert JD (2018) Optimal timing of drug sensitivity testing for patients on first-line tuberculosis treatment. *Health Care Management Science* 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 Transactions on Neural Networks and Learning Systems* 32(11):4793–4813.

Tunç S, Alagoz O, Burnside ES (2022) A new perspective on breast cancer diagnostic guidelines to reduce overdiagnosis. *Production and Operations Management* (608).

Van Der Wardt V, Harrison JK, Welsh T, Conroy S, Gladman J (2017) Withdrawal of antihypertensive medication: A systematic review. *Journal of 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.*

Vlassis N, Littman ML, Barber D (2012) On the computational complexity of stochastic controller optimization in pomdps. *ACM Transactions on Computation Theory (TOCT)* 4(4):1–8.

Wang F, Kaushal R, Khullar D (2020) Should Health Care Demand Interpretable Artificial Intelligence or Accept “Black Box” Medicine? *Annals of Internal Medicine* 172(1):59.

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. *Journal of the American College of Cardiology* 71(19):e127–e248.

E-Companion
Interpretable Policies and the Price of Interpretability in Hypertension Treatment Planning

EC.1. Proofs of Technical Results

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

Proof of Proposition 1 We prove this result by contradiction. Suppose that there exists at least one monotone policy π^M that is optimal regardless of α, i.e., $\pi^M \in \arg\max_{\pi\in\Pi^M} J^\pi(\alpha)$ for all α. Now, consider the MDP in Figure EC.1. If the initial distribution is $\alpha^1 = (0.5, 0, 0.5, 0, 0)$, the set of optimal monotone policies is given by $\Pi^1 = \{(1,1,1,1,1),(1,1,1,1,2),(1,1,1,2,2)\}$. Each $\pi^1 \in \Pi^1$ achieves an expected total discounted reward of $J^{\pi^1}(\alpha^1) = \sum_{t=1}^{\infty} \gamma^t$, with $J^{\pi^1}(\alpha^1) > J^\pi(\alpha^1)$ for any other monotone policy $\pi \notin \Pi^1$. Now, if the initial distribution is $\alpha^2 = (0, 1, 0, 0, 0)$, then the set of optimal monotone policies is given by $\Pi^2 = \{(1,2,2,2,2),(2,2,2,2,2)\}$. Notice that $J^{\pi^1}(\alpha^2) = 0$ for all $\pi^1 \in \Pi^1$, while $J^{\pi^2}(\alpha^2) = \sum_{t=1}^{\infty} \gamma^t > 0$ for all $\pi^2 \in \Pi^2$. Hence, no $\pi^1 \in \Pi^1$ is optimal under initial state distribution α^2 and no $\pi^2 \in \Pi^2$ is optimal under initial state distribution α^1, which is a contradiction. □

Remark EC.1. To show that the optimal CMP also depends on the initial state distribution, it suffices to consider the same example in the proof of Proposition 1 with state classes $\mathcal{S}_1 = \{1, 2\}$ and $\mathcal{S}_2 = \{3, 4, 5\}$. In this case, we do not follow the trivial example of setting the state and action classes to follow strict monotonicity, but still end up with the same result.

Before showing the proof of Theorem 1, we present the following Lemma.

Lemma EC.1. *The number of monotone policies is given by $\binom{S+A-1}{A-1}$.*

Proof of Lemma EC.1. This can be shown directly using the *stars and bars* technique. Each *star* represent a state of the MDP. A deterministic monotone policy can be represented using bars where a bar between two consecutive states, s and s' indicates that $\pi(s) = a$ and $\pi(s') = a + 1$ for some action a. More generally, k bars between consecutive states indicates $\pi(s) = a$ and $\pi(s') = a + k$. k bars before the first state, 1, indicates that $\pi(1) = k$ and k bars after the last state, S, indicates $\pi(S) = A - k$, Thus, the stars and bars technique gives that the number of monotone policies is equivalent to the number of non-negative solutions to $x_1 + x_2 + \ldots + x_A = S$. From combinatorics, it is well-known that this number is $\binom{S+A-1}{A-1}$. □

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.*

Proof of Theorem 1. The first statement follows directly from Lemma EC.1. The second statement follows from the fact that the resulting value in Lemma EC.1 is a binomial coefficient involving the number of states and number of actions. The number of monotone policies for various problem sizes closely relates to the numbers on the diagonals of Pascal's triangle. For a fixed number of states S, the number of monotone policies corresponds to the A^{th} number on the $S + 1$ diagonal of Pascal's triangle. For a fixed number of actions A, the number of monotone policies corresponds to the S^{th} number of the A^{th} diagonal of Pascal's triangle. For each diagonal of Pascal's triangle, the numbers on the diagonal grow as a polynomial function and therefore, the number of monotone policies grows as a polynomial in the number of states for a fixed $\mathcal{S}$ and as a polynomial in the number of actions for a fixed $\mathcal{A}$. □

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 $\tilde{v}(s) = \mathbb{E}^{\pi_{\tilde{\mathbf{x}}}}[\sum_{t=0}^{\infty} \gamma^t r(s_t, \pi_{\tilde{\mathbf{x}}}(s_t))|s_0 = s]$.*

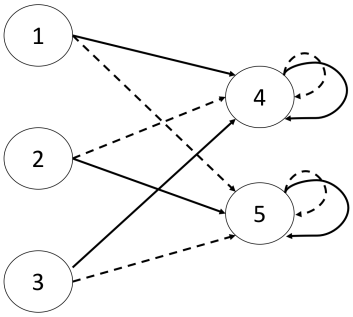

Figure EC.1 **An MDP for which the optimal monotone policy, $\bar{\pi}$, depends on the initial distribution, α. The solid (dashed) lines represent transitions corresponding to taking action 1 (action 2) which occur with probability 1. The rewards are defined as $r(s,a)=0$ for all $s\neq 4, a\in\mathcal{A}$ and $r(4,a)=1$, for all $a\in\mathcal{A}$.**

Proof of Proposition 2. In the following results, we define the sets $\mathcal{S}_0 := \{s\in\mathcal{S} : \sum_{t=0}^{\infty}\mathbb{P}(s_t = s|\pi_{\tilde{\mathbf{x}}},\alpha)=0\}$ and $\mathcal{S}_1 = \mathcal{S}\setminus\mathcal{S}_0$. By definition, we necessarily have $\alpha(s)=0$ for all $s\in\mathcal{S}_0$. For notational convenience, we also define $v^{\pi}(s)=\mathbb{E}^{\pi}[\sum_{t=0}^{\infty}\gamma^t r(s_t,\pi(s_t))|s_0=s]$.

First, we show that $J^{\pi_{\tilde{\mathbf{x}}}}(\alpha)=\sum_{s\in\mathcal{S}}\alpha(s)\tilde{\mathbf{v}}(s)$. For any state $s\in\mathcal{S}_0$, the action $\pi_{\tilde{\mathbf{x}}}(s)$ and value of $\tilde{v}(s)$ are inconsequential to the solution of (5) since s is never reached under policy $\pi_{\tilde{\mathbf{x}}}$. Moreover, for every $s\in\mathcal{S}_1$, we have

$$\tilde{v}(s)=r(s,\pi(s))+\gamma\sum_{s'\in\mathcal{S}_1}P(s'|s,\pi_{\tilde{\mathbf{x}}}(s))\tilde{v}(s').$$

Thus, $\tilde{v}(s)=v^{\pi_{\tilde{\mathbf{x}}}}(s)$ for all $s\in\mathcal{S}_1$ and it follows that

$$J^{\pi_{\tilde{\mathbf{x}}}}(\alpha)=\sum_{s\in\mathcal{S}}\alpha(s)v^{\pi_{\tilde{\mathbf{x}}}}(s)=\sum_{s\in\mathcal{S}}\alpha(s)\tilde{v}(s)=\sum_{s\in\mathcal{S}_1}\alpha(s)\tilde{v}(s).$$

Hence, the result holds.

Next, we show that $\pi_{\tilde{\mathbf{x}}}\in\arg\max_{\pi'\in\Pi^I}J^{\pi'}(\alpha)$. Suppose there exists $\pi'\in\Pi^I$ such that $J^{\pi'}(\alpha)>J^{\pi_{\tilde{\mathbf{x}}}}(\alpha)$. Consider a vector $v'\in\mathbb{R}^{|\mathcal{S}|}$ such that $v'(s)=v^{\pi'}(s)$ for all $s\in\mathcal{S}$ and matrix $\mathbf{x}'\in\mathbb{R}^{|\mathcal{S}|\times|\mathcal{A}|}$ such that $x'(s,a)=1$ if $\pi'(s)=a$ and $x'(s,a)=0$ otherwise. Then, one can verify that $(\mathbf{v}',\mathbf{x}')$ is feasible in (5). Moreover,

$$J^{\pi'}(\alpha)=\sum_{s\in\mathcal{S}}\alpha(s)v'(s)>\sum_{s\in\mathcal{S}}\alpha(s)\tilde{v}(s),$$

which gives us a contradiction on the optimality of $(\tilde{\mathbf{v}},\tilde{\mathbf{x}})$.

Finally, we show that for any state s that is reachable under policy $\pi_{\tilde{\mathbf{x}}}$ and initial distribution α, we have $\tilde{v}(s)=\mathbb{E}^{\pi_{\tilde{\mathbf{x}}}}[\sum_{t=0}^{\infty}\gamma^t r(s_t,\pi_{\tilde{\mathbf{x}}}(s_t))|s_0=s]$. Suppose that $\alpha(s)=0$ for some subset of $s\in\mathcal{S}$ and that $\mathcal{S}_0$ is non-empty. For any reachable state $s\in\mathcal{S}_1$, we have

$$\tilde{v}(s)=r(s,\pi_{\tilde{\mathbf{x}}}(s))=\gamma\sum_{s'\in\mathcal{S}_1}P(s'|s,\pi_{\tilde{\mathbf{x}}}(s))\tilde{v}(s')=\mathbb{E}^{\pi}\Big[\sum_{t=0}^{\infty}\gamma^t r(s_t,\pi_{\tilde{\mathbf{x}}}(s_t))|s_0=s\Big].$$

Hence, we have shown our desired result.

For completeness, we also now show that if s is unreachable under $\pi_{\tilde{\mathbf{x}}}$ and α, then $\tilde{\mathbf{v}}(s)$ may not equal $\mathbb{E}^{\pi_{\tilde{\mathbf{x}}}}[\sum_{t=0}^{\infty}\gamma^t r(s_t,\pi_{\tilde{\mathbf{x}}}(s_t))|s_0=s]$. For any $s\in\mathcal{S}_0$, the values of $\tilde{v}(s)$ are inconsequential to the objective function (5a). Thus, taking

$$v(s)=\begin{cases}v^{\pi_{\tilde{\mathbf{x}}}}(s) & s\in\mathcal{S}_1\\ v^{\pi_{\tilde{\mathbf{x}}}}(s)-\epsilon & s\in\mathcal{S}_0,\end{cases}$$

for arbitrary $\epsilon > 0$ gives us a feasible solution which does not change the objective function value. Hence, $(\tilde{\mathbf{v}}, \tilde{\mathbf{x}})$ is also an optimal solution to (5). $\square$

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

Proof of Proposition 3. Under Assumption 1, we have $\Pi^M \subseteq \Pi^{CMP}_{\Theta,\Psi}$. Hence, $\max_{\pi \in \Pi^M} J^\pi(\alpha) \leq \max_{\pi \in \Pi^{CM}_{\Theta,\Psi}} J^\pi(\alpha)$, which implies $PI(\Pi^M) \geq PI(\Pi^{CM}_{\Theta,\Psi})$. $\square$

EC.2. Practical Consideration

In this section, we provide a brief discussion on the practical implementation of formulations (8) and (9). In auxiliary numerical experiments, we found that formulation (9) may be solved more quickly than formulation (8). We hypothesize that this result may be due to the number of constraints imposed by each formulation; notice that the number of constraints imposed by (7) are

$$G \sum_{k=1}^{K-1} |\mathcal{S}_k||\mathcal{S}_{k+1}| \leq G(K-1)\left(\frac{|\mathcal{S}|}{K}\right)^2 \leq |\mathcal{A}|(|\mathcal{S}|-1),$$

where the final term is the number of constraints in (6). Moreover, the final inequality comes from the fact that there can be at most $|\mathcal{A}|$ action classes and $|\mathcal{S}|$ state classes. We also found that the dual formulations of formulations (8) and (9) can often be solved more quickly than their primal formulations, especially when a warm start solution is provided. We provide details of these dual formulations and the procedure for generating an initial feasible solution in §EC.2.1 and §EC.2.2, respectively.

EC.2.1. Dual Formulations of Formulations (8) and (9)

The dual formulation of formulation (8) is given by

$$\max_{\mathbf{x},\mathbf{y}} \quad \sum_{s\in\mathcal{S}}\sum_{a\in\mathcal{A}} r(s,a)y(s,a) \tag{EC.1a}$$

$$\text{subject to:} \quad \sum_{a\in\mathcal{A}} y(s,a) - \gamma \sum_{s'\in\mathcal{S}}\sum_{a'\in\mathcal{A}} P(s|s',a')y(s',a') = \alpha(s) \text{ for all } s\in\mathcal{S} \tag{EC.1b}$$

$$\sum_{a\in\mathcal{A}} x(s,a) = 1 \text{ for all } s\in\mathcal{S} \tag{EC.1c}$$

$$y(s,a) \leq Mx(s,a) \text{ for all } s\in\mathcal{S},\ a\in\mathcal{A} \tag{EC.1d}$$

$$x(s,a) \leq \sum_{a'\geq a} x(s+1,a') \text{ for all } s\in\mathcal{S}\setminus S, a\in\mathcal{A} \tag{EC.1e}$$

$$y(s,a) \geq 0 \text{ for all } s\in\mathcal{S}, a\in\mathcal{A}. \tag{EC.1f}$$

$$x(s,a) \in \{0,1\} \text{ for all } s\in\mathcal{S}, a\in\mathcal{A}. \tag{EC.1g}$$

In (EC.1d), the parameter M is a large constant. It is well-known that the dual variable $y(s,a)$ represents a discounted "count" of being in state s and performing action a, i.e., $y(s,a) = \sum_{t=0}^{\infty} \gamma^t \mathbb{P}(s_t = s, a_t = a)$. Hence, we can set $M = \frac{1}{1-\gamma} = \sum_{t=0}^{\infty}\gamma^t \geq \sum_{t=0}^{\infty}\gamma^t\mathbb{P}(s_t = s, a_t = a)$ as an upper bound. Note that in the finite horizon case, $y_t(s,a) = \mathbb{P}(s_t = s, a_t = a)$ so setting $M = 1$ suffices. Furthermore, by modifying formulation (EC.1) , the dual formulation for formulation (9) is given by

$$\max_{\mathbf{x},\mathbf{y}} \quad \text{(EC.1a)} \quad \text{subject to:} \quad \text{(EC.1b)-(EC.1d), (EC.1f)-(EC.1g), (7).} \tag{EC.2}$$

EC.2.2. Warm Start via Monotone Policy Iteration

Typically, the computational effort required to solve an MIP can be reduced through a *warm start*, i.e., supplying an initial feasible solution. Here, we modify the classic policy iteration algorithm to identify a policy which is guaranteed to be monotone. Given the importance of the initial distribution α, we rely on α to provide an order by which to prioritize states in the policy. Then, we construct a feasible set of actions for each state based on actions chosen in preceding states. Throughout our algorithm, we denote $\alpha_{[i]}$ as the i^{th} order statistic of α. We initialize our algorithm with some policy $\pi_0(s)$ which is associated with some initial value function v_0. The *Monotone Policy Iteration* algorithm is summarized in Procedure 1.

Procedure 1 Monotone Policy Iteration algorithm

```
Data: $\alpha$, $\mathbf{v}_0$, $\pi_0$, $\epsilon \geq 0$, $t = 1$
Result: Monotone policy $\pi$
while $t = 1$ or $\|\boldsymbol{v}_t - \boldsymbol{v}_{t-1}\| \leq \epsilon$ do
    for $i = 1$ to $S$ do
        Set $s$ as the state corresponding to $\alpha_{[i]}$.
          Set $\mathcal{S}^- \leftarrow \{s' \in \bar{\mathcal{S}} : s' < s\}$ and $\mathcal{S}^+ \leftarrow \{s' \in \bar{\mathcal{S}} : s' > s\}$.
          Set $A_{\min} \leftarrow \begin{cases} 1 & \text{if } |\mathcal{S}^-| = 0 \\ \pi_t(\max \mathcal{S}^-) & \text{otherwise.} \end{cases}$ and $A_{\max} \leftarrow \begin{cases} A & \text{if } |\mathcal{S}^+| = 0 \\ \pi_t(\min \mathcal{S}^+) & \text{otherwise.} \end{cases}$.
          Set $\pi_t(s) \leftarrow \arg\max_{\mathcal{A}_{\min} \leq a \leq \mathcal{A}_{\max}} r(s,a) + \gamma \sum_{s' \in \mathcal{S}} P(s'|s,a) v_{t-1}(s')$.
          Set $\bar{\mathcal{S}} \leftarrow \bar{\mathcal{S}} \cup \{s\}$.
          Perform policy evaluation on $\pi_t$ to obtain $\mathbf{v}_t$. if $\|\boldsymbol{v}_t - \boldsymbol{v}_{t-1}\| \leq \epsilon$ then
            return $\pi = \pi_t$
        else
            Set $t \leftarrow t + 1$.
        end
    end
end
```

Much like the classic policy iteration algorithm, the *Monotone Policy Iteration* algorithm terminates in a finite number of iterations (since there are a finite number of monotone policies). However, it is not guaranteed to provide the optimal monotone policy. Regardless, it is straightforward to verify that *Monotone Policy Iteration* generates a monotone policy. Hence, it can be used to warm start formulation (8). Furthermore, since monotone policies are also CMPs (under Assumption 1), this policy can also be used to warm start formulation (9) .

REMARK EC.2. A monotone policy iteration algorithm is presented in Puterman (2014, Ch. 6.11.2). The main difference between the algorithm presented in Procedure 1 and this previously developed algorithm is the order in which states are visited to obtain the monotone policy. Specifically, our algorithm visits states according to their prominence in the initial state distribution whereas the algorithm described in Puterman (2014) visits the states in ascending order. Nevertheless, both algorithms can be shown to terminate in finitely many iterations and are guaranteed to provide an optimal policy if the optimal policy is indeed monotone in s.

EC.2.3. Creating State and Action Classes

In practice, states or actions that are "adjacent" under strict ordering may be considered "close" in severity. For example, two drugs may have similar risk reduction for a given health outcome, and practically speaking, may be considered nearly identical. We now show that combining adjacent classes will never increase the PI.

PROPOSITION EC.1. *Consider any class functions Θ, Ψ which satisfy Assumption 1 and admit more than two state and action classes, respectively. Let Θ' and Ψ' be state and action class functions which merge two adjacent classes admitted by Θ and Ψ, respectively. Then,*

$$PI(\Pi^{CM}_{\Theta,\Psi}) \geq PI(\Pi^{CM}_{\Theta',\Psi}) \geq PI(\Pi^{CM}_{\Theta',\Psi'}) \text{ and } PI(\Pi^{CM}_{\Theta,\Psi}) \geq PI(\Pi^{CM}_{\Theta,\Psi'}) \geq PI(\Pi^{CM}_{\Theta',\Psi'}).$$

Proof of Proposition EC.1. This result follows directly from Lemma EC.2. □

LEMMA EC.2. *Consider any class functions Θ, Ψ which satisfy Assumption 1.*

1. *Suppose that Θ admits at least two state classes. Let Θ' be a new state class function that merges two classes admitted by Θ, i.e., $\Theta'(s) = \Theta'(s') = k$ for all s, s' where $\Theta(s) = k$ and $\Theta(s') = k+1$ for an arbitrary $k \in \mathcal{K} \setminus \{K\}$. Then $\pi \in \Pi^{CM}_{\Theta,\Psi}$ implies $\pi \in \Pi^{CM}_{\Theta',\Psi}$.*
2. *Suppose that Ψ admits at least two action classes. Let Ψ' be a new action class function that merges two classes admitted by Ψ. Then, $\pi \in \Pi^{CM}_{\Theta,\Psi}$ implies $\pi \in \Pi^{CM}_{\Theta,\Psi'}$*

Proof of Lemma EC.2. To show that Property 1 holds, it suffices to show that for any s^* in the newly merged class and policy $\pi \in \Pi^{CM}_{\Theta,\Psi}$, having $\Theta'(s^*) \geq \Theta'(s)$ implies $\Psi(\pi(s^*)) \geq \Psi(\pi(s))$ and having $\Theta'(s^*) \leq \Theta'(s)$ implies $\Psi(\pi(s^*)) \leq \Psi(\pi(s))$. Take any s^* in the newly merged class and s such that $\Theta'(s^*) \geq \Theta'(s)$. By the construction of Θ', it follows that $\Theta(s^*) \geq \Theta(s)$. Since $\pi \in \Pi^{CM}_{\Theta,\Psi}$, we have $\Psi(\pi(s^*)) \geq \Psi(\pi(s))$. Showing the reverse inequality is similar. Hence, Property 1 has been shown. The proof for showing Property 2 is similar and has been omitted. □

EC.3. Case Study Details

We provide additional methods and results from our case study in this section of the e-companion.

EC.3.1. Finite-Horizon MDP Formulations

In this subsection, we adapt our formulations to the context finite-horizon MDPs for the management of hypertension. Because there is no evidence that hypertension treatment is beneficial at low BP levels, our models do not allow for treatment if patients' systolic BP is below 120 mm Hg or their diastolic BP is below 55 mm Hg. In addition, since many believe hypertension is especially dangerous at high levels, we always offer treatment if patients BP is above 150/90 mm Hg (Schell et al. 2016). We incorporate these clinical restrictions by adding a constraint that establishes that any treatment choice that violates the minimum systolic BP or diastolic BP levels cannot be optimal. The set of treatment choices that leads to "clinically infeasible actions" at state s_t and year t is denoted by $I_t(s_t) \subset \mathcal{A}$. This set of actions for each state is identified before formulating the models based on the estimated effect of antihypertensive drugs on the risk for ASCVD events (Law et al. 2009).

To account for the effect of age in the risk of ASCVD events, we modify the LP formulation in (1) as described in Theorem 1 of Bhattacharya and Kharoufeh (2017). The possibility of "clinically infeasible actions" is incorporated into this formulation by constraining the value functions using the rewards of actions that are clinically feasible. That is, by repeating the constraint in (1) for all $t \in \mathcal{T}' \setminus \{T\}$, $s_t \in \mathcal{S}$, and $a_t \in \mathcal{A} \setminus I_t(s_t)$. This possibility is incorporated into the dual LP formulation by constraining the dual variables $y_t(s_t, a_t)$ associated with the "clinically infeasible actions" as $\sum_{a_t \in I_t(s_t)} y_t(s_t, a_t) = 0$ for all $t \in \mathcal{T}' \setminus \{T\}$ and $s_t \in \mathcal{S}$. Both approaches result in nonbinding constraints at any treatment choice that violates the minimum systolic BP or diastolic BP levels.

Formulation (8) is modified to account for the nonstationarity of the risk for ASCVD events and the plausibility of "clinically infeasible actions" in a similar way. We incorporate a summation over $t \in \mathcal{T}'$ in the objective function, repeat constraint (5b) for all $t \in \mathcal{T}' \setminus \{T\}$, $s_t \in \mathcal{S}$, and $a_t \in \mathcal{A}$, constraint the terminal value functions with the terminal rewards (i.e., $v_T(s_t) \leq r_T(s_T)$ for all $s_T \in \mathcal{S}$), replicate (5c) for all $t \in \mathcal{T}' \setminus \{T\}$ and $s_t \in \mathcal{S}$, repeat constraint (6) for all $t \in \mathcal{T}' \setminus \{T\}$, $s_t \in \mathcal{S} \setminus \{s_t\}$, and $a_t \in \mathcal{A}$, and restrict the binary variables $x_t(s_t, a_t)$ using the following equation:

$$\sum_{a_t \in I_t(s_t)} x_t(s_t, a_t) = 0 \text{ for all } t \in \mathcal{T}' \setminus \{T\}, s_t \in \mathcal{S}. \tag{EC.3}$$

Note that we have added the index t to the binary variables $x_t(s_t, a_t)$ to highlight their dependence on the decision epoch. The dual formulation (EC.1) of Section EC.2.1 of the e-companion is modified as follows:

1. Incorporating a summation over $t \in \mathcal{T}' \setminus \{T\}$ and adding a summation of the product of terminal rewards and dual variables (i.e., $\sum_{s_T \in \mathcal{S}} r_T(s_T) y_T(s_T)$) in (EC.1a).
2. Repeating constraints (EC.1b) and (EC.1c) for all $t \in \mathcal{T}' \setminus \{T\}$ and $s_t \in \mathcal{S}$.
3. Replicating (EC.1d) and (EC.1f) for all $t \in \mathcal{T}' \setminus \{T\}$, $s_t \in \mathcal{S}$, and $a_t \in \mathcal{A}$. Notice that it suffices to take $M = 1$ since $y_t(s,a) = \mathbb{P}(s_t = s, a_t = a)$.
4. Repeating constraint (EC.1e) for all $t \in \mathcal{T}' \setminus \{T\}$, $s_t \in \mathcal{S} \setminus \{S_t\}$, and $a_t \in \mathcal{A}$.
5. Restraining the binary variables $x_t(s_t, a_t)$ using constraint (EC.3).

We adjust the formulation (9) in a similar manner to (8). However, we repeat constraint (7) for all $t \in \mathcal{T}' \setminus \{T\}$, $s_t \in \mathcal{S}_k$, and $s'_t \in \mathcal{S}_{k+1}$ with $k = 1, \ldots, K-1$ instead of (6) for all $t \in \mathcal{T}' \setminus \{T\}$, $s_t \in \mathcal{S} \setminus \{S_t\}$, and $a_t \in \mathcal{A}$. The dual formulation (EC.2) is also similar to the dual formulation (EC.1), with the exception that (EC.1e) is modified to (7).

EC.3.2. Enforcing Monotonicity on Decision Epochs.

To guarantee nondecreasing actions over time, we incorporate the following constraint to formulations (8) and (EC.1):

$$x_t(s_t, a) \le \sum_{a' \ge a} x_{t+1}(s_{t+1}, a') \text{ for all } t \in \mathcal{T}' \setminus \{T\}, s_t, s_{t+1} \in \mathcal{S}, a \in \mathcal{A}.$$

We have dropped the index t in action a to indicate explicitly that they remain constant in the year. We also add the following constraint to formulation (9) and its dual:

$$\sum_{a \in \mathcal{A}_g} x_t(s_t, a) \le \sum_{a' \in \bigcup_{g'=g}^{G} \mathcal{A}_{g'}} x_{t+1}(s_{t+1}, a') \text{ for all } t \in \mathcal{T}' \setminus \{T\}, s_t, s_{t+1} \in \mathcal{S}, g = 1, \ldots, G.$$

EC.3.3. Model Parameters

The clinical parameters used throughout our numerical study are listed in Table EC.1.

EC.3.3.1. Parameters for Transition Probabilities. Based on the national heart disease statistics, 70% of the ASCVD risk is due to CHD events (Virani et al. 2020). The expected reduction in ASCVD risk from treatment is calculated using the equations described in the Methods Section of Law et al. (2009) and the parameters in Table 2 of Law et al. (2003). If a patient has an adverse event, we calculate the probability that the event is fatal by applying fatality likelihoods to the post-treatment risk of CHD and stroke events. The fatality likelihoods are estimated as the ratio of known fatal event rates from the National Center for Health Statistics (NCHS) to the overall event rates in our population predicted by the ASCVD risk, adjusted for age and sex (Goff et al. 2014, NCHS 2017). We incorporate the probability of non-ASCVD mortality by age and sex from Tables 2 and 3 in Arias and Xu (2019).

Based on communications with our clinical collaborators, we assume secondary events are more common than would be predicted by the ASCVD risk score alone (Schell et al. 2016). To account for this difference, we multiply a patient's CHD odds by 3 if the patient has a history of CHD ($h_t = 2, 5$), multiply the odds for stroke by 2 if the patient is at least 60 years old and has a history of stroke ($h_t = 3, 6$), and multiply the stroke odds by 3 if the patient has a history of stroke and is less than 60 years old ($h_t = 3, 6$) (Brønnnum-Hansen et al. 2001, Burn et al. 1994).

EC.3.3.2. Reward Parameters. We obtain the quality of life weights from eTable 4 in Kohli-Lynch et al. (2019). Following previous studies, we assumed a treatment-related disutility (i.e., the harm from each medication) of 0.002 QALYs for a generic BP medication at standard dose (Sussman et al. 2013, Schell et al. 2016). We adjust the treatment harm per medication by the percentage of people with one or more symptoms attributable to the medication type as reported in Table 5 of Law et al. (2003). To assess the lifetime effects of treatment, we compute the total QALYs after the planning horizon from age and sex-specific life expectancy from life tables adjusted by a mortality factor that accounts for the effect of ASCVD events on future mortality

Table EC.1 Model parameters

Parameter Description	Value	Source
Proportion of ASCVD risk due to CHD	70%	Virani et al. (2020)
Relative risk reduction per medication	Varies by patient	Law et al. (2009, 2003)
Fatality likelihood		
CHD	Varies by patient	Goff et al. (2014), NCHS (2017)
Stroke	Varies by patient	
Non-ASCVD Mortality	Varies by patient	Arias and Xu (2019)
Scaling factor to account for history of ASCVD events		
CHD ($h_t = 2, 5$)	3	Brønnnum-Hansen et al. (2001)
Stroke ($h_t = 3, 6$)	2,3	Burn et al. (1994)
Quality of life weight		Kohli-Lynch et al. (2019)
Healthy ($h_t = 1$)	1	
ASCVD events ($h_t = 2, \ldots, 6$)	Varies by patient	
Dead ($h_t = 7, \ldots, 10$)	0	
Treatment harm per medication		
Thiazide diuretics	0.002198	Sussman et al. (2013), Schell et al. (2016), Law et al. (2003)
Beta-blockers	0.002150	
CCB	0.002078	
ACE inhibitors	0.002000	
ARBs	0.002166	
Life expectancy	Varies by patient	Arias and Xu (2019)
Mortality factor	Varies by patient	Pandya et al. (2015)
Discount factor	0.97	Neumann et al. (2016)

and quality of life weights (Arias and Xu 2019, Pandya et al. 2015, Kohli-Lynch et al. 2019). The mortality factors are calculated as the reciprocal of the chronic mortality factors by sex in eTable 2 of Pandya et al. (2015). As per recommendations in the medical literature, all quality of life weights are discounted by 3% (Neumann et al. 2016).

While QALYs provide a method to compare the benefits of health interventions, there are drawbacks associated with their use. For example, they have been controversial in medical research, as some researchers do not consider them to be patient focused. QALYs may also be used as rationing tools by health insurers. Additionally, they may be perceived as dehumanizing for patients because policy makers must assign a number on what people are "worth" (Neumann and Cohen 2018).

EC.3.3.3. Calibration and Validation. We calibrated the number of events predicted by the ASCVD risk calculator to ensure the number of fatal and non-fatal ASCVD events in our model match those of the national statistics. Mortality from secondary events and known overdiagnosis of cardiovascular diseases was accounted for by decreasing the fatal event rates reported by the NCHS by 50% (Govindan et al. 2014). An independent, third-party clinical researcher from the University of Michigan Medical School verified the calibration of our model. Our model of patients' disease progression is built by discussing the parameters and logic with experts in the field. Our co-author, a practicing physician and clinical researcher, helped validate our model.

EC.3.4. Description of Sensitivity Analysis Scenarios

We consider the following sensitivity analysis scenarios:

1. *Nonincreasing severity state order*: In this scenario, we order states by nonincreasing risk for ASCVD events (i.e., $s_t(d_t, c_t, 4), s_t(d_t, c_t, 6), s_t(d_t, c_t, 3), s_t(d_t, c_t, 5), s_t(d_t, c_t, 2)$, and $s_t(d_t, c_t, 1)$).

2. *Single ASCVD events state classes*: We modify the state classes from one class per ASCVD event to one class encompassing all ASCVD events. In this scenario, we combine $\mathcal{S}_2$, $\mathcal{S}_3$, and $\mathcal{S}_4$ into a new class $\mathcal{S}_2'$. Class $\mathcal{S}_1$ remains unchanged.
3. *Action order and classes*: We examine two scenarios by categorizing each treatment in terms of their expected BP reduction.
   (a) *Systolic BP reductions*: We generate five classes using 5 mm Hg increments on the expected systolic BP reduction of each treatment until 25 mm Hg (slightly above the maximum systolic BP reduction with 5 medications). As ordering actions according to their systolic BP reductions is equivalent to ordering them by ASCVD risk reduction, the order of actions within each class remains unchanged.
   (b) *Diastolic BP reductions*: We classify medications into six classes based on 3 mm Hg increments of each treatment's expected diastolic BP reduction until 18 mm Hg (marginally above the maximum diastolic BP reduction with 5 medications). In this scenario, the actions are ordered in terms of their expected diastolic BP reductions as reported by Law et al. (2009).
4. *Initial state distribution*: We evaluate three additional cases.
   (a) *99% in healthy states at year 0*: We assign 99% of the state distribution weight to states associated with healthy conditions at the first year of our study. The remaining 1% of the initial state distribution is uniformly dispersed over the rest of the states and years.
   (b) *99% in year 0*: To represent the influence of time on patients' health, we assign 99% of the state distribution weight uniformly to states at the first year of our study. The leftover 1% of the initial state distribution is uniformly dispersed over remaining states and years.
   (c) *Uniform weight*: We distribute α uniformly over all states and years (i.e., decision epochs).

EC.3.5. Additional Case Study Results

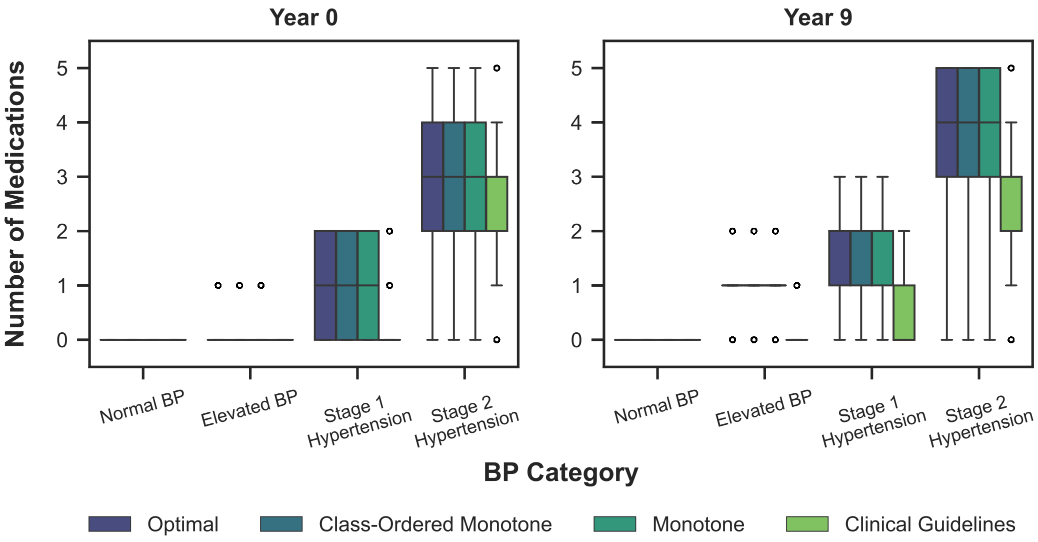

Figure EC.2 Distribution of treatment at year 0 and year 9 of the study. BP categories are made based on patients' characteristics at year 0.

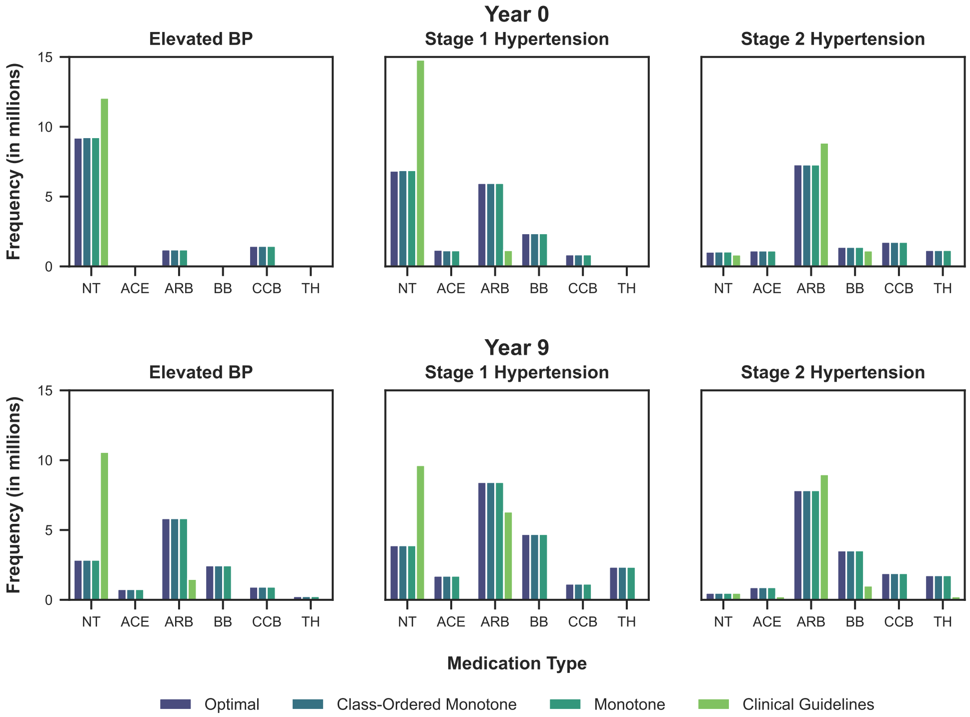

Figure EC.3 Frequency of medication type across all number of medications prescribed per BP category at year 0 (top) and 9 (bottom) of the study. NT: No treatment; ACE: ACE inhibitor; ARB: Angiotensin II receptor blockers; BB: Beta blocker; CCB: Calcium channel blocker; TH: Thiazide.

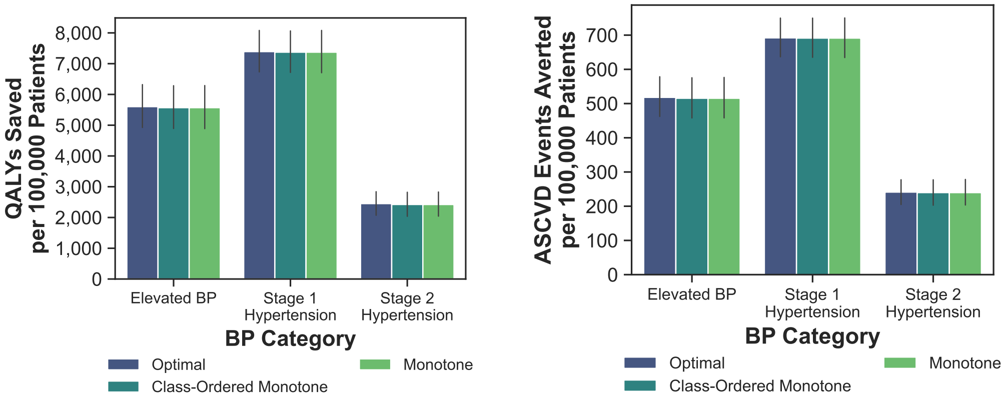

Figure EC.4 QALYs saved (left) and ASCVD events averted (right) by each treatment policy at every BP category per 100,000 patients, compared to the clinical guidelines. Error bars represent the 95% bootstrap confidence intervals around the mean per 100,000 patients using 10,000 replications.

EC.3.5.1. Distribution of PI. In this subsection, we study the PI of π^{CM} and π^{M} across each patient in our population. For comparison purposes, we also examine the pairwise differences between the total discounted reward of π^{CM} and π^{M} for every patient. We inspect our results as a function of each patient's risk for ASCVD events. The ASCVD risk summarizes the health and provides a rich description of the patient in a single number. In Figure EC.5, we find that the PI of

π^{CM} and π^{M} are considerably related. Overall, 0.13% of patients experience a higher PI with π^{M} than with π^{CM}. This translates to 0.00%, 0.12%, and 0.63% for patients with elevated BP, stage 1 hypertension, and stage 2 hypertension. We also note that higher PIs typically appear in lower risk scores at each BP category. Moreover, higher PIs tend to lead to large pairwise differences in PI between the π^{CM} and π^{M}.

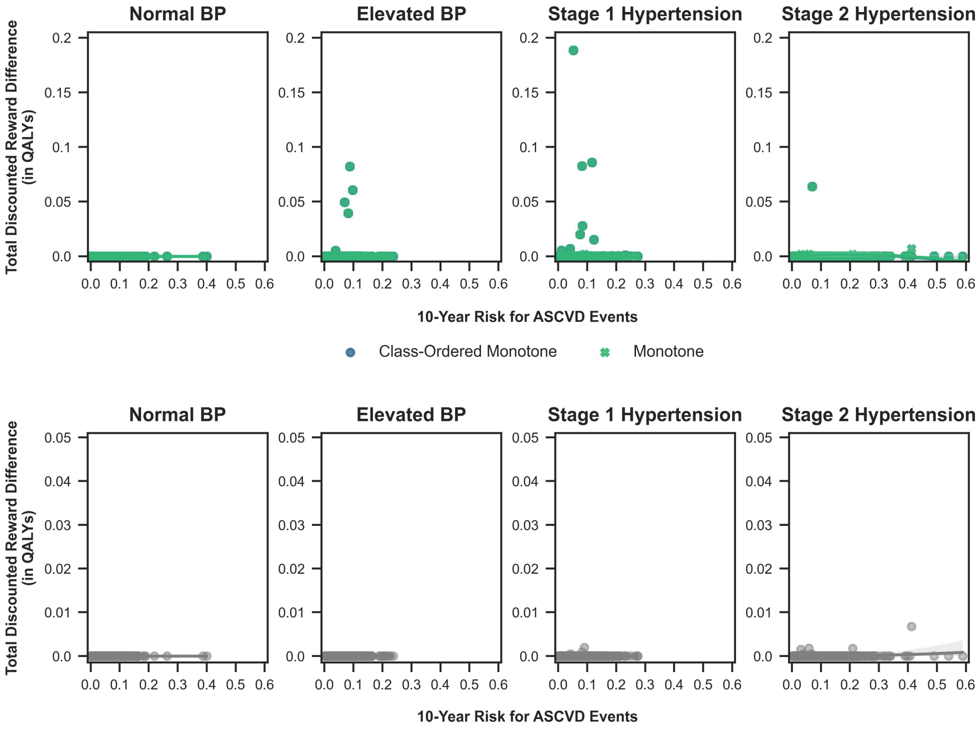

Figure EC.5 **Distribution of PI (top) and pairwise differences between π^{CM} and π^{M} (bottom) over patients' 10-year risk for ASCVD events. Smoothed lines are obtained using second degree local regression. Shaded areas around the smoothed lines represent 95% bootstrapped confidence intervals around the mean using 10,000 replications.**

EC.3.5.2. Effect of Modeling Assumptions on Computational Time. Changing our modeling assumptions affects the computational time of our optimization models, and therefore, the number of patients for whom we obtain optimal solutions. We find that the ordering and classification of the states and actions have minor effects on computational time. In these scenarios, no more than 2 records corresponding to 23,504 patients exceed the time limit.

We also note that focusing 99% of the initial state distribution in healthy states at year 0 of our study does not have a substantial effect in the computational time. In this scenario, 1 record is excluded corresponding to 16,218 patients. The largest increase in computational time is observed in the scenarios where a greater proportion of the initial state distribution is uniformly allocated across multiple states. We exclude 832 records corresponding to 10.04 million patients in the scenario in which 99% of the initial state distribution is uniformly dispersed across the states representing the first year of our study. A total of 902 records associated with 11.25 million patients exceeded the 60-minute time limit when the initial state distribution is uniformly allocated across all states. Overall, we note that the more uniform the dispersion of the initial state distribution across the states is, the greater the computational time. Combining all the scenarios, a total of 1,042 records corresponding to 13.03 million patients are excluded due to the time-limit restrictions in our sensitivity analyses.