

Monitoring With Limited Information

Dan A. Iancu * Nikolaos Trichakis † Do Young Yoon ‡

April 11, 2019

Abstract

We consider a system with an evolving state that can be stopped at any time by a decision maker (DM), yielding a state-dependent reward. The DM does not observe the state except for a limited number of monitoring times, which he must choose, in conjunction with a suitable stopping policy, to maximize his reward. Dealing with this type of stopping problems, which arise in a variety of applications from healthcare to finance, often requires excessive amounts of data for calibration purposes, and prohibitive computational resources. To overcome these challenges, we propose a robust optimization approach, whereby adaptive uncertainty sets capture the information acquired through monitoring. We consider two versions of the problem – static and dynamic – depending on how the monitoring times are chosen. We show that, under certain conditions, the same worst-case reward is achievable under either static or dynamic monitoring. This allows recovering the optimal dynamic monitoring policy by resolving static versions of the problem. We discuss cases when the static problem becomes tractable, and highlight conditions when monitoring at equi-distant times is optimal. Lastly, we showcase our framework in the context of a healthcare problem (monitoring heart-transplant patients for Cardiac Allograft Vasculopathy), where we design optimal monitoring policies that substantially improve over the status quo recommendations.

1 Introduction

Consider a system with a randomly evolving state that can be stopped at any time by a decision maker (DM), yielding a state-dependent reward. The DM does not observe the system’s state except for a specific limited number of times when he chooses to monitor the system, perhaps at

*Graduate School of Business, Stanford University, daniancu@stanford.edu. (Corresponding author.)

†Massachusetts Institute of Technology, ntrichakis@mit.edu

‡Uber Freight

some cost. To maximize his reward, the DM needs to devise dynamic policies that prescribe when to monitor the system and whether to continue or stop, based on all acquired information.

Such problems have received a lot of attention in the stochastic optimization literature: the common approach is to view these as optimal stopping problems in which the uncertain evolution is characterized by a precise probabilistic model, and to rely on dynamic programming techniques to derive policies that maximize the expected reward.

However, many practical considerations that arise in application domains ranging from health-care to finance can render such approaches unsuitable. First, calibrating joint probability distribution functions for all possible time points in the planning horizon typically requires a prohibitively large amount of data that might not be available in practice, or relies on structural assumptions on state evolution, such as independence or lack of auto-correlation, that might not conform with the true dynamics of the system. Second, the system’s complexity is often such that it requires a high-dimensional state representation, resulting in “curse of dimensionality” issues. Third, modeling the learning process that emerges from acquiring information through system monitoring usually requires augmenting this state beyond computational reach.

To see this in a concrete example, consider patients suffering from Cardiac Allograft Vasculopathy (CAV), who need to appropriately time their main viable treatment option, namely heart-transplant: performing the transplant too early reduces its success chances, while performing it too late comes with the risk of CAV progression and health deterioration. Monitoring requires invasive and potentially expensive procedures, and is necessary given the lack of readily observable biomarkers for disease progression. Also lacking are credible ways to model and predict disease progression: indeed, very little data is available from prior studies of CAV patients, a problem that is compounded by the disease’s pathology, which is known to be very complex (see §4). How should CAV patients and their physicians make monitoring and treatment decisions facing this considerable uncertainty?

Similarly, consider a lending institution that extends commercial loans backed by working assets, such as inventory. For a given loan, the lender can occasionally monitor the value of the pledged collateral through costly field visits and inspections, and decide whether to request early repayment or force collateral liquidation, to avert future possible losses. Making predictions about future collateral value, however, could be a very challenging task, as it would involve tracking the value of a considerable number of assets classes that are present in or correlated with the pledged assets. Furthermore, these assets may be illiquid and the lender may lack expertise in valuing them, such as with specialized types of inventory (CH 2014). How should asset-based lenders make monitoring

and liquidation decisions when they need to track a multitude of highly uncertain assets?

To address such questions, we develop a new solution methodology that overcomes some of the challenges discussed above. In particular, we rely on the robust optimization paradigm to develop a model where the DM’s information concerning future state evolution is captured through multi-dimensional uncertainty sets. Rather than being exogenously specified, the uncertainty sets governing future state values are adaptive, and depend on the state values observed at past monitoring times. We study the DM’s problem of finding a policy for choosing the monitoring times and for stopping the system so as to maximize his worst-case reward. Our framework alleviates some of the aforementioned practical challenges: uncertainty sets can be more readily calibrated from limited data, and the robust optimization problems that need to be solved for finding optimal policies cope more favorably with “curse of dimensionality” issues.

We make the following contributions.

- We consider two versions of the problem: static, where all the monitoring times are committed to upfront, and dynamic, where monitoring times can be adjusted depending on acquired information. We show that when the reward is monotonic and the uncertainty sets have a lattice structure, the same worst-case reward is achievable under either static or dynamic monitoring. This allows recovering an optimal dynamic monitoring policy by re-solving static versions of the monitoring problem, drastically simplifying the decision problem.
- We then discuss sufficient conditions under which the static monitoring and stopping problem can be efficiently solved. We focus on generalizations of box-type (Ben-Tal et al. 2009) and Central-Limit-Theorem-type (Bandi and Bertsimas 2012) uncertainty sets, wherein past measurements impose lower and upper bounds on future state values, giving rise to an “uncertainty envelope.” We show that the curvature of this uncertainty envelope is critical. When the envelope is generated by bounding functions that are convex in the elapsed time, a single monitoring opportunity suffices for recovering the worst-case reward, and the optimal time can be found by solving a one-dimensional optimization problem. When the bounding functions are concave, all monitoring times are needed for reducing the uncertainty, and it is optimal for the DM to distribute these times uniformly over the horizon. We discuss conditions relying on complementarity properties (e.g., supermodularity) of the reward function and uncertainty envelopes that allow solving the problem through combinatorial algorithms.
- We leverage our approach to devise monitoring and transplantation policies for patients suffering from Cardiac Allograft Vasculopathy (CAV). We calibrate our model using data from

published medical papers, and show how optimal monitoring and stopping policies can be found by solving mixed-integer linear programming problems. Simulation results suggest that our policies generate life year gains that dominate those provided by existing medical guidelines, with a substantial increase in lower percentiles and a slight increase in median and higher percentiles.

Finally, we consider two extensions of our base model. In the first, we consider the case when monitoring comes at some given cost, and we discuss the problem of finding the optimal number of monitoring times. In the second, we consider a decision process that is more general than stopping, and monitoring allows the DM to adjust the state values by extracting rewards or incurring a cost.

1.1 Literature Review

Our assumption that information is captured through uncertainty sets relates this paper to the extensive literature in robust optimization (RO) and robust control (see, e.g., Ben-Tal and Nemirovski 2002, Bertsimas et al. 2011 and references therein for the former, and Dullerud and Paganini 2005, Zhou and Doyle 1998 for the latter). Typical models in RO consider uncertainty sets that are exogenous to the decision process, and focus on the computational tractability of the resulting optimization problems. Poss (2013) and Nohadani and Sharma (2016) consider decision-dependent uncertainty sets, and propose mixed-integer programming reformulations for the resulting non-convex optimization problems (also see Jaillet et al. 2016 for similar concepts applied in a model with satisficing objectives and constraints). Similarly, Zhang et al. (2016) study robust control problems where the decision maker *chooses* the uncertainty set from a family of structured sets, and incurs a penalty for “smaller” choices (in terms of radius, volume, etc.); since the dynamic decision problems are generally intractable, the paper resorts to policy approximations through affine decision rules, which are numerically shown to deliver good performance. Bertsimas and Vayanos (2017) consider a pricing problem where the demand function coefficients are unknown, and propose an approximation scheme that requires solving mixed-binary conic optimization problems. Different from these papers, our uncertainty sets depend on a monitoring policy chosen by the DM, which reduces future uncertainty under well defined information updating rules.

Closest to our paper is Nohadani and Roy 2017 (N&R), who study decision-making over a discrete time horizon and formalize time-dependent uncertainty sets that capture increased uncertainty as information ages with time. In particular, problem parameters are assumed to lie

in uncertainty sets whose progression over time mimics the notion of the event horizon from the theory of relativity in physics. Allowing for one intermediate observation time, N&R formulate a two-stage problem and derive a tractable robust counterpart. N&R also consider the problem of optimizing the observation schedule when additional observations are allowed. In a general framework where information degrades with time and observations “reset uncertainty,” N&R show that equidistant observations provide the highest value for the inferred information. Finally, N&R apply their framework to cases in radiation therapy and prostate cancer.

There are similarities and differences between our work and N&R. Both papers study time-dependent uncertainty sets, and the uncertainty envelope in our generalized box- and CLT-type uncertainty sets is similar to the event horizon idea from N&R. But these sets are then leveraged in different decision-making contexts: we study stopping problems under multiple monitorings, whereas N&R study linear problems under one intermediate monitoring. Further, both papers also address the problem of optimizing the observation/monitoring schedule, albeit using different approaches. N&R decouple the decision-making problem from the monitoring optimization problem, under the rationale that the former is likely to benefit the most from monitoring that yields the most valuable inferred information. They also consider the problem of finding an optimal fixed, stationary monitoring schedule. In contrast, we consider the stopping problem and the monitoring optimization problem jointly, and we deal with both stationary and dynamic monitoring schedules. Another difference lies in that N&R use conic geometry to model the uncertainty envelope, whereas we use bounding functions. The latter approach enables us to consider alternative uncertainty evolution processes, which could exhibit non-stationarities or uncertainty that grows at either increasing or decreasing rate. Under stationary and increasing uncertainty, we shall see that in our model observations also reset uncertainty in the verbiage of N&R, and we also find equidistant observations to be optimal. Under non-stationary or decreasing uncertainty, however, we find that optimal observations need not be equidistant, in particular, they might exhibit increasing frequency.

Our work is also related to a subset of papers in RO that discuss the optimality of static policies in dynamic decision problems. Ben-Tal et al. (2004) prove this for a linear program where the uncertain parameters affecting distinct constraints are disjoint, i.e., the uncertainty is “constraint-wise.” Marandi and den Hertog (2017) extend this condition to two-stage problems where constraints are convex in decisions and concave in uncertainties, and Bertsimas et al. (2014) prove static optimality for two-stage linear programs when a particular transformation of the uncertainty set is convex. We extend this literature by proving optimality for static monitoring rules in a broad class

of multi-period problems, provided the objectives possess certain monotonicity properties, and the uncertainty sets have ordering properties (e.g., they are lattices).

Our work is also related to the broad literature on stopping problems. These are typically formulated under a fully specified probability distribution for the stochastic processes of interest (see, e.g., Snell (1952), Taylor (1968), Karatzas and Shreve (2012), and references therein). Some of this literature also incorporates robustness by allowing for a set of possible distributions chosen by nature, and extending the martingale approach to characterize stopping in the (continuous-time) game between the DM and nature (see, e.g., Riedel (2009), Bayraktar and Yao (2014), Bayraktar et al. (2010), and references therein). In contrast, we consider all measures with a given support (i.e., the uncertainty set, as in classical RO), and restrict the stopping decision to a finite number of times, chosen by the decision maker. This provides a more sharp characterization of the optimal stopping policy, as well as computationally tractable procedures for finding it.

Finally, our applications of interest also relate our paper (albeit more loosely) to an extensive literature studying monitoring and stopping problems in healthcare or finance. For the former stream, we direct the interested reader to Shechter et al. (2008) (studying the optimal time to initiate HIV treatment), Maillart et al. (2008) and Ayer et al. (2012) (assessing breast cancer screening policies), Denton et al. (2009) (optimizing the start of statin therapy for diabetes patients), Lavieri et al. (2012) (optimizing the start of radiation therapy for prostate cancer), and the numerous references therein. This line of work typically relies on models based on (partially observable) Markov Decision Processes, and unique probability distributions for transitions and rewards. Instead, we adopt a robust model where limited information is available; this allows a sharper characterization of optimal policies, which may not be possible when insisting on Bellman-optimal policies, as required under uniquely specified stochastic processes and risk-neutral objectives (also see Delage and Iancu 2015 and Iancu et al. 2013 for a discussion).

For the latter stream, we mention the extensive work on pricing exotic options, of which Bermudan options—which can only be exercised on a pre-determined set of dates—are perhaps closest to our framework (see Ibanez and Zapatero (2004) and Kolodko and Schoenmakers (2006) for Bermudan options, and Wilmott et al. (1994) and Karatzas et al. (1998) for a broader overview). In contrast to our work, stopping times are typically exogenous, and exact computations of the liquidation frontier are generally too complex, so that solution methods typically rely on discretized numerical integration and/or simulation-based approximations. Also related are papers in finance that deal with debt monitoring and repayment/amortization schedules (see, e.g., Leland (1994), Leland and Toft (1996), Gorton and Winton (2003), Rajan and Winton (1995), and references

therein). Such papers typically consider stylized settings (e.g., two or three-periods, a single, uniquely defined stochastic process of interest, etc.), since the focus is characterizing the optimal capital structure, rather than optimizing monitoring schedules or liquidation policies.

1.2 Notation and Terminology

To distinguish vectors from scalars, we denote the former by bold fonts, e.g., $\mathbf{x} \in \mathbb{R}^n$. For a set of vectors $\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_k$ in \mathbb{R}^d , we use $\mathbf{x}^{\{k\}} \stackrel{\text{def}}{=} [\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_k] \in \mathbb{R}^{d \times (k+1)}$ to denote the matrix obtained by horizontal concatenation. Similarly, the operator $;$ is used to denote vertical concatenation of matrices with the same number of columns. For a set of matrices $\mathcal{U} \subset \mathbb{R}^{d \times k}$ and indices $1 \leq i_1 \leq i_2 \leq d$ and $1 \leq j_1 \leq j_2 \leq k$, we define

$$\begin{aligned} \Pi_{i_1:i_2, \mathcal{U}} &= \left\{ Y \in \mathbb{R}^{(i_2-i_1+1) \times k} : \exists X \in \mathbb{R}^{(i_1-1) \times k}, Z \in \mathbb{R}^{(d-i_2) \times k} : [X; Y; Z] \in \mathcal{U} \right\}, \\ \Pi_{:,j_1:j_2} \mathcal{U} &= \left\{ Y \in \mathbb{R}^{d \times (j_2-j_1+1)} : \exists X \in \mathbb{R}^{d \times (j_1-1)}, Z \in \mathbb{R}^{d \times (k-j_2)} : [X, Y, Z] \in \mathcal{U} \right\}, \end{aligned}$$

as the projections along a subset of rows and a subset of columns, respectively. To simplify notation, we also use $\mathcal{U}_j \stackrel{\text{def}}{=} \Pi_{:,j:j} \mathcal{U}$ to denote the latter projection along a single column.

We use the terms “increasing” and “decreasing” in a weak sense, and we refer to functions of multiple arguments as being increasing (decreasing) instead of isotone (antitone), which is the established lattice terminology (Topkis 1998). We compare vectors in \mathbb{R}^d using the typical component-wise order, i.e., $\mathbf{x} \succeq \mathbf{y}$ for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$ if and only if $x_i \geq y_i$, $1 \leq i \leq d$. Similarly, we compare matrices by first viewing them as vectors, and applying the component-wise order above. For a set of vectors in \mathbb{R}^d , “increasing” and “decreasing” are understood in the usual induced set order¹ (Topkis 1998), unless explicitly stated otherwise.

2 Model

Consider a system whose state is characterized by a finite number of exogenous processes evolving over a finite time horizon. A decision maker (DM) starts with limited information on the future process values and can observe (or monitor) the system at a finite number of times. These times, which we refer to as *monitoring times*, are chosen by the DM. At every monitoring time, the DM

¹We first define the operator “ \sqsubseteq ” as follows: for any non-empty subsets $A, B \subseteq \mathbb{R}^d$, we have $A \sqsubseteq B$ if $\min(\mathbf{x}, \mathbf{y}) \in A$ and $\max(\mathbf{x}, \mathbf{y}) \in B$ holds for any $\mathbf{x} \in A, \mathbf{y} \in B$. We can then define the notion of “increasing” subsets of \mathbb{R}^d . Suppose T is a non-empty subset of \mathbb{R}^p , and A_t is a sequence of non-empty subsets of \mathbb{R}^d parameterized by $\mathbf{t} \in T$. Then A_t is “increasing” (“decreasing”) in \mathbf{t} if $A_{\mathbf{t}_1} \sqsubseteq A_{\mathbf{t}_2}$ ($A_{\mathbf{t}_2} \sqsubseteq A_{\mathbf{t}_1}$), for any $\mathbf{t}_1, \mathbf{t}_2 \in T$ with $\mathbf{t}_2 \succeq \mathbf{t}_1$.

observes the state of the system, updates his information about the future process values, and decides whether to stop the system or allow it to continue evolving until the next monitoring time. If at some monitoring time the DM chooses to stop the system, he collects a reward that depends on the stopping time and the state. If he decides to continue until the end, he collects the reward associated with the terminal state. We focus on a system with high uncertainty, where the DM’s information on the future state evolution is described by uncertainty sets rather than a complete probability law. In this setting, the DM’s problem is to find a policy for choosing the monitoring times and for stopping the system so as to maximize his worst-case reward.

We next introduce notation and provide a formal model description.

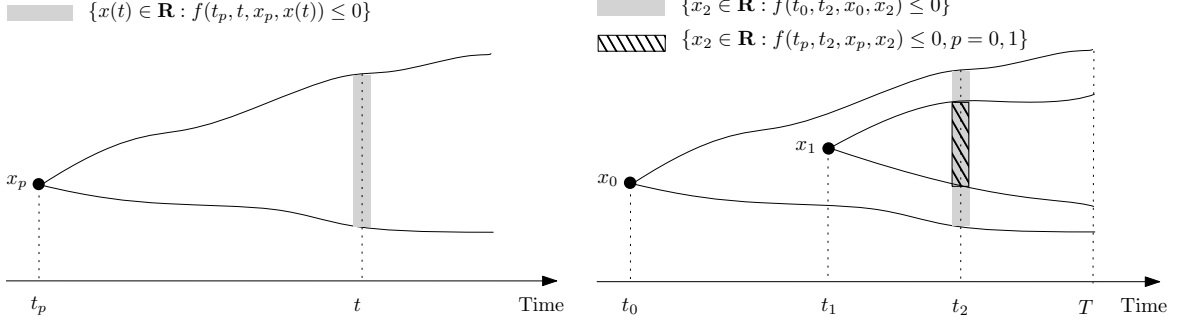
2.1 State Evolution and Uncertainty Set

The system evolves over a continuous time frame $[0, T]$, and is characterized by d real-valued processes. Let $\mathbf{x}(t) \in \mathbb{R}^d$ be a vector whose components are the system’s processes, for all t in $[0, T]$. We refer to $\mathbf{x}(t)$ as the state of the system at time t .

The DM has knowledge of the initial state, i.e., $\mathbf{x}(0)$, and can monitor the system at most n times throughout the planning horizon, in addition to the default monitoring times of 0 and T . Let t_1, \dots, t_n denote these n monitoring times with $0 \leq t_1 \leq \dots \leq t_n \leq T$. To ease notation, we also let $t_0 \stackrel{\text{def}}{=} 0$ and $t_{n+1} \stackrel{\text{def}}{=} T$.

At each monitoring time t_p , $0 \leq p \leq n + 1$, the DM observes the state $\mathbf{x}_p \stackrel{\text{def}}{=} \mathbf{x}(t_p)$ and updates his information on the future possible state values. More precisely, each observed state \mathbf{x}_p imposes restrictions on the feasible set of values $\mathbf{x}(t)$ at every future time $t > t_p$. These restrictions are summarized through m constraints, written compactly as $f(t_p, t, \mathbf{x}_p, \mathbf{x}(t)) \leq 0$, where $f : \mathbb{R}^+ \times \mathbb{R}^+ \times \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^m$. For an illustration, see Figure 1a.

We assume that the DM reconciles new information (from new measurements) with old information (from existing measurements) by considering all the restrictions imposed. To make this precise and introduce some notation, consider the k -th monitoring time t_k . At that time, the DM has made k observations at monitoring times t_1, \dots, t_k , in addition to the initial observation at t_0 , resulting in an observation matrix $\mathbf{x}^{\{k\}} \stackrel{\text{def}}{=} [\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_k]$. Suppose the DM is considering some future monitoring times t_{k+1}, \dots, t_r satisfying $t_k \leq t_{k+1} \leq \dots \leq t_r \leq t_{n+1}$, for $r > k$. Then, in view of observations $\mathbf{x}^{\{k\}}$, the process values at t_{k+1}, \dots, t_r , together with the terminal process value at



(a) Consider the case $d = 1$. Each observation x_p imposes a restriction on the future process values $x(t)$ for all $t > t_p$, which is identified by $f(t_p, t, x_p, x(t))$.

(b) With two observations $\mathbf{x}^{\{1\}} = [x_0, x_1]$, the set of feasible state values x_2 at time t_2 , given by the projection of $\mathcal{U}(\mathbf{t}^{\{2\}}, \mathbf{x}^{\{1\}})$ on x_2 (dashed), is the intersection of the projection of $\mathcal{U}(\mathbf{t}^{\{2\}}, \mathbf{x}^{\{0\}})$ on x_2 (gray) and the restriction imposed by the second observation x_1 .

Figure 1: Illustration of how observed state values constrain future values for $d = 1$.

t_{n+1} , lie in the following uncertainty set:

$$\mathcal{U}(\mathbf{t}^{\{r\}}, \mathbf{x}^{\{k\}}) = \left\{ [\mathbf{x}_{k+1}, \dots, \mathbf{x}_r, \mathbf{x}_{n+1}] \in \mathbb{R}^{d \times (r-k+1)} : f(t_p, t_q, \mathbf{x}_p, \mathbf{x}_q) \leq 0, \right. \\ \left. \forall p, q \in \{0, 1, \dots, r, n+1\}, p < q \right\}, \quad (1)$$

where $\mathbf{t}^{\{r\}} \stackrel{\text{def}}{=} [t_0, t_1, \dots, t_r]$ denotes the vector of the first $r + 1$ monitoring times (including the initial time t_0). As such, information updating in our model corresponds to taking intersections of the corresponding uncertainty sets.²

To illustrate this, consider the example in Figure 1b, where a system with a single process ($d = 1$) is monitored twice ($n = 2$), at t_1 and t_2 . At time t_0 , the DM starts with an initial information on the future values x_1, x_2 and $x_3 = x_{n+1}$, represented by $\mathcal{U}(\mathbf{t}^{\{2\}}, \mathbf{x}^{\{0\}}) = \{[x_1, x_2, x_3] \in \mathbb{R}^3 : f(t_0, t_q, x_0, x_q) \leq 0, q = 1, 2, 3\}$. Then, with an observation x_1 at t_1 , the updated set $\mathcal{U}(\mathbf{t}^{\{2\}}, \mathbf{x}^{\{1\}})$ becomes the intersection of $\{[x_2, x_3] \in \mathbb{R}^2 : f(t_0, t_q, \mathbf{x}_0, \mathbf{x}_q) \leq 0, q = 2, 3\}$ and $\{[x_2, x_3] \in \mathbb{R}^2 : f(t_1, t_q, x_1, x_q) \leq 0, q = 2, 3\}$, which are the restrictions imposed by x_0 and x_1 , respectively. Figure 1b depicts the set of feasible values for x_2 , which is the projection of $\mathcal{U}(\mathbf{t}^{\{2\}}, \mathbf{x}^{\{1\}})$ on the first dimension.

²Taking intersection also implies that uncertainty sets updated after new observations can only include values that were already considered plausible at time t_0 . This allows the DM to behave in a dynamically consistent fashion, and ensures that having additional information is always preferable (Machina 1989, Hanany and Klibanoff 2009). Note that considering a large enough set of future plausible state values at time t_0 always suffices for this to hold.

Note that in some cases, the process values at later periods may depend solely on the most recent observation, as opposed to all past observations — for instance, this occurs in the conic uncertainty sets of Nohadani and Roy 2017. In that case, the description in (1) could be simplified by omitting past observations \mathbf{x}_q , $q < k$. To exemplify, in the one-dimensional setting considered in Figure 1b, if

$$f(t_p, t_q, x_p, x_q) = \begin{bmatrix} x_q - x_p - h(t_q - t_p) \\ -x_q + x_p - h(t_q - t_p) \end{bmatrix},$$

and the function h is convex, it can be readily shown that past observations can be omitted. If h is concave, however, this is no longer true and one would need to follow the model in (1). Including all past observations in the uncertainty evolution thus enables us to deal with more general bounding functions. We re-visit these issues in Example 1 and §3.3.1.

Dealing with arbitrary uncertainty sets defined as in (1) could lead to intractable models in general. We thus impose the following assumption, which bears intuitive interpretations and is compatible with several families of uncertainty sets considered in the literature.

Assumption 1. For any $0 \leq k < r \leq n$ and given $\mathbf{t}^{\{r\}}$ and $\mathbf{x}^{\{k\}}$,

- (i) (Lattice) $\mathcal{U}(\mathbf{t}^{\{r\}}, \mathbf{x}^{\{k\}})$ is a lattice;
- (ii) (Monotonicity) $\mathcal{U}(\mathbf{t}^{\{r\}}, \mathbf{x}^{\{k\}})$ is increasing in $\mathbf{x}^{\{k\}}$;
- (iii) (Dynamic Consistency) $\Pi_{:,i:j}\mathcal{U}(\mathbf{t}^{\{r\}}, \mathbf{x}^{\{k\}}) = \Pi_{:,i:j}\mathcal{U}(\mathbf{t}^{\{r'\}}, \mathbf{x}^{\{k\}})$, $\forall i \leq j < r - k + 1, \forall r \leq r' \leq n$, and $\forall t_{r'} \geq t_{r'-1} \geq \dots \geq t_r$.

The lattice requirement is primarily of technical nature.³ The monotonicity requirements are guided by practical considerations. At an intuitive level, these are akin to “better past performance” being “good news” for the future, and “worst past performance” being “bad news” for the future. Mathematically, “better past performance” means higher $\mathbf{x}^{\{k\}}$, which leads to larger future state values, i.e., good news for the future. Conversely, “worse past performance” means lower $\mathbf{x}^{\{k\}}$, which leads to lower future state values, i.e., bad news for the future. This makes sense in chronic disease monitoring, such as CAV, for example: better (worse) medical history is often indicative of good (bad) outcomes in the future. Finally, dynamic consistency corresponds to the natural requirement that committing to additional monitoring times in the future, $t_{r+1}, \dots, t_{r'}$, does not affect the set of feasible state values at monitoring times before t_r .

In practice, requirements (i) and (ii) are relatively easy to check, as our later examples will illustrate. To facilitate checking (iii), we introduce the following sufficient condition.

³Strictly speaking, our results only require \mathcal{U} to be a meet-semilattice.

Proposition 1. *Suppose that for any t_p and \mathbf{x}_p , and any t, t' satisfying $t_p \leq t < t'$,*

$$\{\mathbf{x} \in \mathbb{R}^d : f(t_p, t, \mathbf{x}_p, \mathbf{x}) \leq 0\} \subseteq \{\mathbf{x} \in \mathbb{R}^d : f(t_p, t', \mathbf{x}_p, \mathbf{x}) \leq 0\}. \quad (2)$$

Then, Assumption 1(iii) holds.

The sufficient condition (2) requires that all possible process values at time t remain possible values at time $t' > t$. This is a very mild requirement if \mathbf{x} corresponds to a “zero-mean” (noise) process, since it is natural to expect information for more distant future times to become “less accurate”/“more noisy” in such settings. But the assumption remains innocuous even when \mathbf{x} exhibits predictable trends or seasonalities: one can decompose the process into a component satisfying (2) and a predictable time-varying component, whose effect can be captured through the time-dependency of the reward function.

2.2 Examples of Uncertainty Sets

We next present two examples of uncertainty sets satisfying our requirements.

Example 1 (Lattice with Cross-Constraints). *For $\beta \geq \alpha \geq 0$, $\mathcal{M} \subseteq \{1, \dots, d\}^2$, and functions $\ell : \mathbb{R}^2 \rightarrow \mathbb{R}_-$ and $u : \mathbb{R}^2 \rightarrow \mathbb{R}_+$ such that ℓ decreases in its second argument and u increases in its second argument, consider a set:*

$$\begin{aligned} \mathcal{U}(\mathbf{t}^{\{r\}}, \mathbf{x}^{\{k\}}) = & \left\{ [\mathbf{x}_{k+1}, \dots, \mathbf{x}_r, \mathbf{x}_{n+1}] \in \mathbb{R}^{d \times (r-k+1)} : \right. \\ & \alpha \cdot x_p^m + \ell(t_p, t_q - t_p) \leq x_q^{m'} \leq \beta \cdot x_p^m + u(t_p, t_q - t_p), \\ & \left. \forall (m, m') \in \mathcal{M}, \forall p \in \{0, 1, \dots, k\}, \forall q \in \{k+1, \dots, r\} \right\}. \quad (3) \end{aligned}$$

Here, \mathcal{M} specifies a group of processes that are dependent on each other. When $(m, m') \in \mathcal{M}$, an observation with value x_p^m for the m -th process at time t_p would impose restrictions on the future process value $x_q^{m'}$ for the m' -th process at time t_q , in the form of a lower bound $\alpha x_p^m + \ell(t_p, t_q - t_p)$ and an upper bound $\beta x_p^m + u(t_p, t_q - t_p)$. The bounds depends on the time of the past observation, t_p , and on the elapsed time between the observations, $t_q - t_p$. That $-\ell$ and u increase in the elapsed time (i.e., their second argument) reflects more variability in process values due to passage of time.

To see that Assumption 1(i) is satisfied, note that the uncertainty set in (3) is given by a finite collection of bimonotone linear inequalities,⁴ and is therefore a lattice (Queyranne and Tardella

⁴A linear inequality in variables \mathbf{x} is bimonotone if it can be written as $a_i x_i + a_j x_j \leq c$, for some i, j , with $a_i \cdot a_j \leq 0$.

2006). Assumption 1(ii) is readily satisfied since both $\alpha x + \ell(t_p, t_q - t_p)$ and $\beta x + u(t_p, t_q - t_p)$ are increasing in x . Lastly, condition (2)—and thus Assumption 1(iii)—is also satisfied since $-\ell$ and u are increasing in their second argument.

Sets such as (3) are a natural extension of the classical box uncertainty model (Ben-Tal et al. 2009) to our dynamic setting where bounds depend on observed information and monitoring times. Further, the time dependency in (3) also resembles that in Nohadani and Roy (2017), who model uncertainty progression using a conic description based on a general norm. The uncertainty progression in N&R is stationary, i.e., the uncertainty cone opens up in the same fashion independent of the monitoring time. In contrast, by allowing the bounding functions ℓ and u to depend on both elapsed time and monitoring time, our model can capture non-stationary uncertainty progression, which we revisit in §3.3.2.

Finally, the example generalizes to bounding functions that depend on the pair of components (i.e., we have $\alpha^{m,m'}$, $\beta^{m,m'}$, $\ell^{m,m'}$ and $u^{m,m'}$), and constraints written more generally as $\tilde{\ell}(x_p^m, t_p, t_q - t_p) \leq x_q^{m'} \leq \tilde{u}(x_p^m, t_p, t_q - t_p)$, where $\tilde{\ell}$ is convex, \tilde{u} is concave, and both functions increase in their first argument. We omit more details for simplicity of exposition.

Example 2 (CLT-Budgeted Uncertainty Sets). *Consider the case where the d processes correspond to random walks with independent increments. In particular, consider $d = 1$, and assume that increments corresponding to a time length Δt have mean $\mu\Delta t$ and standard deviation $\sigma\sqrt{\Delta t}$. By expressing the difference between two observations x_q and x_p (for $q > p$) as a sum of intermediate independent increments, i.e., $x_q - x_p = \sum_{i=p+1}^q (x_i - x_{i-1})$, one can then rely on the ideas introduced by Bandi and Bertsimas (2012) to formulate the following Central-Limit-Theorem-type uncertainty set:*

$$\mathcal{U}(\mathbf{t}^{\{r\}}, \mathbf{x}^{\{k\}}) = \left\{ [x_{k+1}, \dots, x_r, x_{n+1}] \in \mathbb{R}^{r-k+1} : -\Gamma \leq \frac{x_q - x_p - (t_q - t_p)\mu}{\sqrt{t_q - t_p}\sigma} \leq \Gamma, \right. \\ \left. \forall p, q \in \{0, 1, \dots, r, n+1\}, p < q \right\},$$

for some $\Gamma > 0$. Note that this set can be reformulated as a lattice uncertainty set with $\alpha = \beta = 1$, $\ell(t_p, t_q - t_p) = (t_q - t_p)\mu - \Gamma\sqrt{t_q - t_p}\sigma$, and $u(t_p, t_q - t_p) = (t_q - t_p)\mu + \Gamma\sqrt{t_q - t_p}\sigma$. Assumption 1(ii) is immediate, and by Proposition 1, Assumption 1(iii) is satisfied when the budget is sufficiently high ($\Gamma\sigma \geq 2\mu\sqrt{T}$).

2.3 Decision Problem

We consider two models of the problem—*static* and *dynamic*—depending on how the monitoring times are chosen. In the static model, the DM selects all monitoring times t_1, \dots, t_n at time t_0 . The choice of monitoring times remains fixed throughout the planning horizon, while the stopping decision remains dynamic, adapted to available information; more formally, a stopping policy at the k -th monitoring time, π_k^S , is a mapping from the σ -algebra induced by the choice of times t_1, \dots, t_{n+1} and the observation matrix $\mathbf{x}^{\{k\}}$ to the set $\{\text{Stop}, \text{Continue}\}$. The DM’s *static monitoring problem* thus entails choosing the monitoring times $\mathbf{t}^{\{n\}}$ and a set of stopping policies, $\pi^S \stackrel{\text{def}}{=} \{\pi_0^S, \pi_1^S, \dots, \pi_n^S\}$.

In the dynamic model, the DM only needs to select and commit to the nearest future monitoring time. More precisely, at time t_k , the DM simultaneously chooses whether to stop the process or not, and in the latter case selects the time t_{k+1} when to next monitor the process. In the *dynamic monitoring problem*, the DM therefore needs to choose a monitoring and stopping policy $\pi^D = \{\pi_0^D, \pi_1^D, \dots, \pi_n^D\}$, where each π_k^D is a mapping from the σ -algebra induced by $(\mathbf{t}^{\{k\}}, \mathbf{x}^{\{k\}})$ to $\{\text{Stop}, \text{Continue}\} \times [t_k, T]$.

In both models, the DM collects a reward $g(t, \mathbf{x})$ when stopping at time t with state \mathbf{x} .⁵ His objective is to maximize his worst-case stopping reward over the possible monitoring and stopping policies.

As with the uncertainty set definition, dealing with arbitrary reward functions could lead to intractable models. Therefore, we introduce the following assumption.

Assumption 2. $g(t, \mathbf{x})$ is monotonic in each x_j , $\forall j \in \{1, \dots, d\}$.

As we shall see in the next section, reward monotonicity is consistent with our motivating applications and several practical considerations. Without loss, we henceforth assume that $g(t, \mathbf{x})$ is increasing in all states (this can always be achieved through a change of variables, by replacing x_j with $-x_j$ for any j such that g is decreasing in x_j).

2.4 Applications

We conclude by showing how the model captures the two motivating applications in §1.

Healthcare. Consider a patient suffering from a chronic condition for which the preferred initial approach is passive, by regularly monitoring disease progression through office visits and/or various tests. Such monitoring is often invasive, requiring exposure to toxic substances (e.g., radioactive

⁵That the reward is collected only when stopping is without loss of generality here, since we could always append an additional process to the state, accumulating all the intermediate (discounted) rewards.

agents in imaging studies) or even micro-surgical procedures (e.g., collecting tissues for biopsy). In addition, it can also be expensive in monetary terms, generating significant costs for to the patient. These considerations limit the number of monitoring chances, and require physicians to make judicious use of them. It thus becomes critical to understand when to monitor and—depending on observed outcomes—whether to switch to an active treatment, such as a disease modifying agent or a surgical procedure.

In this context, the state $\mathbf{x}(t)$ can capture the health measurements related to disease progression, which can be collected during an office visit. For instance, when monitoring heart-transplanted patients for CAV, $\mathbf{x}(t)$ could consist of the degree of angiographic lesion, the number of acute rejections, the CAV stage, and the age. The stopping reward $g(t, \mathbf{x})$ depends on the focus, but can generally capture the patient’s total cumulative quality-adjusted life years (QALY) from switching to an active treatment.

Our assumptions have natural interpretations for CAV. Assumption 1(ii) holds since higher CAV stages in the past lead to higher stages in the future (see §4). Additionally, Assumption 2 holds since rewards are often monotonic in the health state; for instance, the total expected QALY for a CAV patient receiving re-transplantation decrease with age, the number of acute rejections and CAV stage (see the model of rewards in §4).

Lending. Consider a lender issuing a loan with collateral consisting of several working assets, e.g., accounts receivable, inventory, equipment, etc. Since lenders often lack extensive in-house expertise for evaluating collateralized working assets, it is customary to contract with third party liquidation houses to obtain periodic appraisals (see, e.g., Foley et al. 2013 and CH 2014, p.16). Based on these appraisals and on certain advance rates (i.e., hair-cuts) applied to each collateral type,⁶ the lender then calculates a borrowing base, which essentially corresponds to the liquidity-adjusted collateral value (CH 2014, p.15-20). This borrowing base is critical in determining whether the borrower is “over-advanced,” i.e., whether the outstanding loan is not supported by the collateral value.

Relating to our model, the state $\mathbf{x}(t)$ can capture the value of the d assets pledged as collateral, which fluctuates over time. With $a_i \in [0, 1]$ denoting the advance rate assigned to the i -th collateral type, each monitoring opportunity then allows the lender to observe the current value of the borrowing base, $\mathbf{a}^T \mathbf{x}(t)$. The reward $g(t, \mathbf{x})$ can be modeled in several ways, depending on the remedial actions imposed by the lender when the borrower is over-advanced (e.g., requiring early repayment, renegotiating the contract, or even liquidating the assets), but it would typically be a

⁶In practice, advance rates for accounts receivable range from 70% to 90%, and advance rates for inventory are at most 65% of the book value or 80% of the net orderly liquidation value (see CH 2014, p.17-19).

function of t and \mathbf{x} that increases in \mathbf{x} .

As for our assumptions, several empirical studies show that pledged asset values are positively autocorrelated over a short time horizon (Cutler et al. 1991), which validates Assumption 1(ii). That the lender's payoff upon a remedial action typically grows with the collateral's value (CH 2014) means that Assumption 2 also likely holds in practice.

3 Analysis

We first formulate the two problems—dynamic and static—as dynamic programs.

3.1 Dynamic Programming Formulations

Dynamic Monitoring. Let $J_k(\mathbf{t}^{\{k\}}, \mathbf{x}^{\{k\}})$ denote the worst-case value-to-go at time t_k , with the first k observations $\mathbf{x}^{\{k\}}$ made at times $\mathbf{t}^{\{k\}}$. The Bellman recursions become:

$$J_k(\mathbf{t}^{\{k\}}, \mathbf{x}^{\{k\}}) = \max\left(g(t_k, \mathbf{x}_k), \max_{t_{k+1} \in (t_k, T]} \min_{\mathbf{x}_{k+1} \in \mathcal{U}_{k+1}(\mathbf{t}^{\{k+1\}}, \mathbf{x}^{\{k\}})} J_{k+1}(\mathbf{t}^{\{k+1\}}, \mathbf{x}^{\{k+1\}})\right), \quad (4)$$

where $J_{n+1}(\mathbf{t}^{\{n+1\}}, \mathbf{x}^{\{n+1\}}) = g(t_{n+1}, \mathbf{x}_{n+1})$. For an intermediate monitoring time t_k , $J_k(\mathbf{t}^{\{k\}}, \mathbf{x}^{\{k\}})$ is obtained as the maximum of the reward from stopping at t_k and the worst-case continuation value obtained by (optimally) picking the next monitoring time t_{k+1} . The DM's problem is then to find a set of monitoring and stopping policies $\pi^D = \{\pi_k^D\}_{k=0}^n$, where π_k^D consists of a stopping policy at time t_k and a policy $\tau_k^D(\mathbf{t}^{\{k\}}, \mathbf{x}^{\{k\}})$ for choosing the next monitoring time t_{k+1} . More precisely,

$$\tau_k^D(\mathbf{t}^{\{k\}}, \mathbf{x}^{\{k\}}) = \arg \max_{t_{k+1} \in (t_k, T]} \min_{\mathbf{x}_{k+1} \in \mathcal{U}_{k+1}(\mathbf{t}^{\{k+1\}}, \mathbf{x}^{\{k\}})} J_{k+1}(\mathbf{t}^{\{k+1\}}, \mathbf{x}^{\{k+1\}}).$$

Let $J = J_0(t_0, \mathbf{x}^{\{0\}})$.

Static Monitoring. Suppose all monitoring times were chosen as $\mathbf{t}^{\{n+1\}}$. Let $V_k(\mathbf{t}^{\{n+1\}}, \mathbf{x}^{\{k\}})$ denote the worst-case value-to-go function at time t_k with the first k observations $\mathbf{x}^{\{k\}}$ made. The Bellman equations under static monitoring become:

$$V_k(\mathbf{t}^{\{n+1\}}, \mathbf{x}^{\{k\}}) = \max\left(g(t_k, \mathbf{x}_k), \min_{\mathbf{x}_{k+1} \in \mathcal{U}_{k+1}(\mathbf{t}^{\{n+1\}}, \mathbf{x}^{\{k\}})} V_{k+1}(\mathbf{t}^{\{n+1\}}, \mathbf{x}^{\{k+1\}})\right), \quad (5)$$

where $V_{n+1}(\mathbf{t}^{\{n+1\}}, \mathbf{x}^{\{n+1\}}) = g(t_{n+1}, \mathbf{x}_{n+1})$. The DM's problem is to choose a set of stopping policies $\pi^S = \{\pi_k^S\}_{k=0}^n$ and a vector of monitoring times $\mathbf{t}^S \in \arg \max_{\mathbf{t}^{\{n+1\}}} V_0(\mathbf{t}^{\{n+1\}}, \mathbf{x}^{\{0\}})$. Let $V = V_0(\mathbf{t}^S, \mathbf{x}^{\{0\}})$.

3.2 Optimal Policy Under Dynamic Monitoring

We will show that the optimal policy under dynamic monitoring can be recovered by dynamically (re)solving a sequence of static monitoring problems. The first step is to show that the optimal worst-case reward achieved under dynamic monitoring is the same as under static monitoring.

Theorem 1. *Under Assumption 2 and Assumption 1, we have*

$$J = V, \quad t_1^S \in \tau_1^D(t_0, \mathbf{x}_0).$$

Although static monitoring is able to achieve the same worst-case reward as dynamic monitoring and to generate an optimal initial monitoring time t_1^S , the subsequent optimal static monitoring times t_2^S, \dots, t_n^S are not necessarily dynamically consistent. In particular, if the process values observed at t_1^S do not correspond to nature's worst-case actions, the DM may prefer to adjust the second monitoring time to a value that differs from t_2^S (and the same rationale applies at subsequent monitoring times). However, this issue can be addressed by applying the static policy in a non-committal fashion, i.e., by implementing the first monitoring time and re-solving a static monitoring problem over the remaining horizon, using the updated information. This intuition is formalized in the following result.

Corollary 1. *The following algorithm yields an optimal dynamic monitoring policy:*

For all $k = 0, 1, \dots, n$:

- 1. At the k -th monitoring time t_k , having observed \mathbf{x}_k , find an optimal static monitoring policy over the remaining time horizon $[t_k, t_{n+1}]$, with $n - k$ monitoring chances, initial state \mathbf{x}_k , and observation matrix $\mathbf{x}^{\{k\}}$. Let $V(\mathbf{t}^{\{k\}}, \mathbf{x}^{\{k\}})$ denote the worst-case optimal reward, and $t_1^S(\mathbf{t}^{\{k\}}, \mathbf{x}^{\{k\}})$ denote the first monitoring time under this policy.*
- 2. Stop if $g(t_k, \mathbf{x}_k) \geq V(\mathbf{t}^{\{k\}}, \mathbf{x}^{\{k\}})$. Otherwise continue and $\tau_k^D(\mathbf{t}^{\{k\}}, \mathbf{x}^{\{k\}}) \leftarrow t_1^S(\mathbf{t}^{\{k\}}, \mathbf{x}^{\{k\}})$.*

This result underscores the importance of solving the static monitoring problem, which we analyze next.

3.3 Optimal Policy Under Static Monitoring

Our approach is to first characterize nature's optimal actions and the DM's optimal stopping strategy for a given set of monitoring times. This will allow us to then reformulate (and simplify) the DM's problem to one of choosing only the monitoring times.

Proposition 2. Consider a fixed set of monitoring times $\mathbf{t}^{\{n+1\}}$. Nature's optimal (worst-case) response when the DM continues in period t_k can be recovered recursively as:

$$\underline{\mathbf{x}}_k(\mathbf{t}^{\{n+1\}}) \stackrel{\text{def}}{=} \min \left\{ \boldsymbol{\xi} \in \mathbb{R}^d : \boldsymbol{\xi} \in \mathcal{U}_k(\mathbf{t}^{\{n+1\}}, [\mathbf{x}_0, \underline{\mathbf{x}}_1(\mathbf{t}^{\{n+1\}}), \dots, \underline{\mathbf{x}}_{k-1}(\mathbf{t}^{\{n+1\}})]) \right\}, 1 \leq k \leq n+1. \quad (6)$$

A given set of monitoring times thus induces a particular (predictable) worst-case path, and the DM's stopping problem reduces to choosing when to stop along that path. The DM's problem (5) can thus be re-formulated as:

$$V = \max_{\mathbf{t}^{\{n+1\}}} \max_{k \in \{1, \dots, n+1\}} g(t_k, \underline{\mathbf{x}}_k(\mathbf{t}^{\{n+1\}})). \quad (7)$$

Our next result further simplifies this problem, proving that there exists an optimal set of monitoring times for which, along the resulting worst-case path, it is optimal either to stop at the last monitoring time t_n or to continue until the end.

Theorem 2. The optimal value in (7) can be obtained as

$$V = \max_{\mathbf{t}^{\{n+1\}}} \max_{k \in \{n, n+1\}} g(t_k, \underline{\mathbf{x}}_k(\mathbf{t}^{\{n+1\}})). \quad (8)$$

Theorem 2 implies that the static problem entails solving two optimization problems. Since $t_{n+1} \equiv T$ is fixed, this can be done provided we can find a solution for the problem:

$$\max_{\mathbf{t}^{\{n+1\}}} g(t_n, \underline{\mathbf{x}}_n(\mathbf{t}^{\{n+1\}})). \quad (9)$$

Solution approaches critically depend on the structure of the reward function g and the uncertainty sets. For our subsequent analysis, we focus on the class of lattice uncertainty sets with cross constraints introduced in Example 1, which subsumes many interesting sets considered in the literature (e.g., sets with box constraints and CLT-budgeted sets, per our discussion in §2.2). For simplicity, we take $\alpha = \beta = 1$, so that:

$$\mathcal{U}(\mathbf{t}^{\{r\}}, \mathbf{x}^{\{k\}}) = \left\{ [\mathbf{x}_{k+1}, \dots, \mathbf{x}_r, \mathbf{x}_{n+1}] \in \mathbb{R}^{d \times (r-k+1)} : \right. \\ \left. \begin{aligned} x_p^m + \ell(t_p, t_q - t_p) &\leq x_q^{m'} \leq x_p^m + u(t_p, t_q - t_p), \\ \forall (m, m') \in \mathcal{M}, \forall p \in \{0, 1, \dots, k\}, \forall q \in \{k+1, \dots, r\} \end{aligned} \right\}. \quad (10)$$

To rule out degenerate cases where the uncertainty sets could be empty, we require that $\ell(\cdot, 0) = u(\cdot, 0) = 0$. Let $\underline{x}_0^{m'} \stackrel{\text{def}}{=} \max_{m: (m, m') \in \mathcal{M}} x_0^m, \forall m' \in \{1, \dots, d\}$.

3.3.1 Stationary Bound Functions.

We first discuss the case where ℓ and u only depend on their second argument, i.e., the elapsed time between monitoring opportunities; with a slight overload of notation, we write $\ell(t_q - t_p)$ and $u(t_q - t_p)$. This already captures many uncertainty sets of practical interest, such as the CLT-budgeted sets of Example 2.

Theorem 3. *Under Assumption 2 and for the uncertainty set in (10),*

(i) *if $\ell(\cdot)$ is convex, then*

$$V = \max_{t \in [0, T]} g(t, \mathbf{x}_0 + \ell(t)\mathbf{1}), \quad t_1^S \in \arg \max_{t \in [0, T]} g(t, \mathbf{x}_0 + \ell(t)\mathbf{1}), \quad (11)$$

and t_2^S, \dots, t_n^S can be chosen arbitrarily from $[t_1^S, T]$;

(ii) *if $\ell(\cdot)$ is concave, then*

$$V = \max \left(\max_{t \in [0, T]} g(t, \phi_n(t)), g(T, \phi_{n+1}(T)) \right), \quad (12)$$

where $\phi_k(t) \stackrel{\text{def}}{=} \mathbf{x}_0 + k\ell(t/k)\mathbf{1}, \forall k \in \{n, n+1\}$. Moreover, if $g(t, \mathbf{x})$ is jointly concave, then (12) reduces to solving a convex optimization problem.

According to Theorem 3, when the bound functions have well-defined curvatures, solving the static monitoring problem reduces to optimizing a one-dimensional function.

When ℓ is convex, it is worst-case optimal for the DM to stop at the first monitoring time, found by solving (11), which renders the choice of subsequent monitoring times irrelevant.⁷ Intuitively, this occurs because a convex lower bound function corresponds to a decreasing rate of uncertainty growth over time. This results in the initial estimates from time t_0 providing the least conservative lower bound estimate of the future process values, and makes any additional observations irrelevant for the DM's problem of "learning" the worst-case path. An example of such convex bound functions is the CLT-budgeted uncertainty set in Example 2, where $\ell(\delta) = \delta\mu - \Gamma\sqrt{\delta}\sigma$.

When ℓ is concave, it is worst-case optimal for the DM to evenly space the monitoring times, stopping either at the last monitoring time t_n (chosen to maximize $g(t, \phi_n(t))$) or at the end of the planning horizon, T . Intuitively, a concave lower bound induces an increasing rate of uncertainty growth over time, so that worst-case estimates of future process values quickly degrade. Thus,

⁷Note that this remains in line with Theorem 2—simply pick $t_1^S = t_2^S = \dots = t_n^S$ and we then have, according to Theorem 3, an optimal set of monitoring times for which it is optimal to stop at the first monitoring time, and, because they coincide, also at the last monitoring time, satisfying Theorem 2.

new observations can substantially reduce the uncertainty, and the optimal policy avails itself of all monitoring chances, distributing them uniformly over time so as to maximally reduce the uncertainty and improve worst-case outcomes.

Finally, the above discussion reinforces our remark in §2.1 that when ℓ is concave, the uncertainty evolution depends only on the most recent observation, and not on older observations. In that sense, the information we have about the underlying processes “resets” every time an observation is made, and we also recover the result that uniform monitoring is optimal, in line with Nohadani and Roy (2017). In contrast, when ℓ is convex, the information is not simply reset and past observations matter, making uniform monitoring potentially suboptimal. Another case where this might arise is when the bound function ℓ is not stationary, as we explore next in this section, and also in our CAV study in §4.

3.3.2 Non-stationary Bound Functions.

We now treat the more general case in which the bounds depend on both monitoring time and elapsed time, i.e., $\ell(t_p, t_q - t_p)$ and $u(t_p, t_q - t_p)$. Allowing the bounds to depend on the monitoring time enables us to capture seasonality or non-stationary patterns for the uncertainty evolution, for instance, non-stationary asset price volatilities (in a financial setting) or age-dependent uncertainty in disease progression (in a healthcare setting). Such patterns would typically occur at a macro-scale, as opposed to the micro-scale evolution captured by the dependence on the elapsed time.

As expected, additional conditions are required to characterize the solutions in the static monitoring problem. Our first result parallels Theorem 3, highlighting the importance of the bounds’ curvature.

Theorem 4. *Under Assumption 2 and for the uncertainty set in (10),*

(i) *if $\ell(t, \delta)$ is decreasing in t and convex in δ , then*

$$V = \max_{t \in [0, T]} g(t, \mathbf{x}_0 + \ell(t_0, t - t_0)\mathbf{1}), \quad t_1^S \in \arg \max_{t \in [0, T]} g(t, \mathbf{x}_0 + \ell(t_0, t - t_0)\mathbf{1}),$$

and t_2^S, \dots, t_n^S can be chosen arbitrarily from $[t_1^S, T]$;

(ii) *if $\ell(t, \delta)$ is increasing in t and concave in δ , then*

$$V = \max \left(\max_{t_n \in [0, T]} g(t_n, \mathbf{x}_0 + \xi_n(t_n)\mathbf{1}), g(T, \mathbf{x}_0 + \xi_{n+1}(T)\mathbf{1}) \right), \quad \text{where} \quad (13a)$$

$$\xi_k(t_k) \stackrel{\text{def}}{=} \max_{0 \leq t_1 \dots \leq t_{k-1} \leq t_k} \left[\sum_{i=1}^k \ell(t_{i-1}, t_i - t_{i-1}) \right], \quad \forall k \in \{n, n+1\}. \quad (13b)$$

Furthermore, if ℓ is jointly concave, then (13b) is a convex optimization problem; and if $g(t, \mathbf{x})$ is jointly concave, then (13a) reduces to solving a convex optimization problem.

Part (i) shows that when the bound $\ell(t, \delta)$ is convex in δ and also *decreasing in t* , it would again be worst-case optimal for the DM to stop at the first monitoring time, found by solving a one-dimensional optimization problem; as before, the choice of subsequent monitoring times is irrelevant for maximizing the worst-case reward. Intuitively, that bound functions would decrease in t if the forecasting technology improved over time, or the processes became “more predictable.”

Part (ii) shows that when $\ell(t, \delta)$ is concave in δ and also *increasing in t* , it would be worst-case optimal for the DM to rely on all monitoring chances, which are picked so as to reduce the uncertainty as much as possible, per (13b). When ℓ and g satisfy additional (mild) requirements, the requisite problems have a convex structure, and are tractable.

In the absence of curvature information about ℓ , Problem (9) can also be viewed from a combinatorial optimization perspective. In particular, if the function $g(t_n, \mathbf{x}_n(\mathbf{t}^{\{n+1\}}))$ is a supermodular function of $\mathbf{t}^{\{n+1\}}$, Problem (9) would involve maximizing a supermodular function over a lattice—a problem that can be tackled through tractable combinatorial algorithms (see, e.g., Fujishige (2005), Schrijver (2003), and references therein). Although it is difficult to exactly characterize when $g(t_n, \mathbf{x}_n(\mathbf{t}^{\{n+1\}}))$ becomes supermodular, we provide a set of sufficient conditions in the following result.

Proposition 3. *Under Assumption 2 and for the uncertainty set in (10), if:*

(i) $g(t, \mathbf{x})$ is supermodular in \mathbf{x} , convex in every component x_k , and

(ii) ℓ is supermodular and decreasing in (t_p, t_q) on the lattice $[0, T]^2$,

then $g(t_k, \mathbf{x}_k(\mathbf{t}^{\{k\}}))$ is a supermodular function of $\mathbf{t}^{\{k-1\}}$ for any fixed t_k , for $k \in \{n, n+1\}$.

To solve Problem (9), one should thus follow a two-stage approach.⁸ In the second stage, for a fixed t_n , one can rely on Proposition 3 to find t_1, \dots, t_{n-1} , by maximizing the supermodular function $g(t_n, \mathbf{x}_n(\mathbf{t}^{\{n\}}))$ over the lattice $0 \leq t_1 \leq t_2 \leq \dots \leq t_{n-1} \leq t_n$. In the first stage, one would solve a one-dimensional optimization problem over t_n .

In practice, this combinatorial approach could be quite useful, since t_n may be required to be integral or a multiple of a base planning period like a month, a week, a day, etc. Some of the premises of Proposition 3 are reasonable. Having rewards g that are supermodular is natural in

⁸In case $g(t, \mathbf{x})$ is also increasing in t and (jointly) convex in (t, \mathbf{x}) , then this two-step approach is not needed, since $g(t_n, \mathbf{x}_n(\mathbf{t}^{\{n\}}))$ is actually supermodular in $\mathbf{t}^{\{n\}}$, by Proposition 2.

many healthcare applications, where the gains from improving one biometric can be larger when other biometrics are also improved. In CAV monitoring, for example, the number of acute rejections and the CAV stage act in such a complementary fashion. Component-wise convexity of g suggests increasing returns to scale in each component of \mathbf{x} , which could be expected particularly for low values of biometrics in healthcare applications.

Finally, when none of the approaches above is applicable, one may still be able to formulate the static monitoring problem as a mixed-integer linear program, and rely on modern MIP solvers to determine the monitoring times. This approach works particularly well when the primitives (i.e., the functions g and ℓ) are piece-wise linear. Since the resulting model heavily depends on the problem specifics, we omit a general-purpose formulation and instead direct the interested reader to the next section for a concrete example.

4 Robust Monitoring of Cardiac Allograft Vasculopathy

We leverage our approach to devise monitoring and re-transplantation policies for patients suffering from Cardiac Allograft Vasculopathy (CAV). As discussed in the Introduction, CAV patients face disease progression that is highly uncertain and therefore could benefit from policies prescribed by our method. We use real data to calibrate our model and then evaluate the performance of our proposed monitoring policies vis-a-vis established guidelines in the medical community. Simulation suggests that our policies provide a QALY distribution that dominates that provided by existing guidelines, with a substantial increase in lower percentiles and a slight increase in median and higher percentiles. The results showcase the efficacy and robustness of our policies.

Background. Heart transplantation often represents the only viable option for patients suffering from refractory or end-stage heart failure. Although post-operative outlooks have improved over the last decades, with 5-year survival rates reaching 72.5% (Wilhelm 2015), heart transplantation could still result in serious long-term complications, such as CAV.

CAV is caused by a thickening and hardening of coronary arteries, which obstructs the blood circulation through the heart; this can cause various cardiac problems, from abnormal heart rhythms (arrhythmias) to heart attacks, heart failure, or even sudden cardiac death. CAV is the primary limiting factor for the long-term survival in heart transplantation, accounting for 17% of deaths by the third post-transplant year and 30% by the fifth year. When early stages are included, CAV affects up to 75% of heart-transplanted patients during the three years after transplantation (Ramzy et al. 2005).

To prevent the onset of CAV, physicians rely on a combination of measures to control cardiovascular risk factors such as hypertension, diabetes, hyperlipidemia, obesity and smoking, as well as drugs to prevent cyclomegalovirus infections and organ rejection (see Kobashigawa (2000), Costanzo et al. (2010), Kobashigawa et al. (2013), Cheng and Kobashigawa (2016) for more details).

The pathology of CAV presents two main challenges in managing the disease. First, CAV is a form of chronic rejection that lacks prominent symptoms, and it is particularly difficult to predict its progression. To obtain information about CAV development, patients must undergo periodic monitoring procedures that involve examinations through coronary angiography or intravascular ultrasonography. These procedures are very invasive: they require the introduction of a catheter through the patient’s artery and the use of dyes or ultrasound to visualize the blood vessels, and patients must subsequently rest for several hours to recover (Mayo Clinic 2019). In addition, each procedure costs thousands of dollars in the US (Healthcare Bluebook 2019).

Second, after CAV detection, re-transplantation is the main viable treatment option.⁹ This itself introduces a difficult trade-off. On the one hand, re-transplantation in the early disease stages is not advisable, because of the substantial evidence that its success is negatively correlated with the interval between transplants (Schnetzler et al. 1998, Johnson 2007, Saito et al. 2013, John et al. 1999, Vistarini et al. 2010). However, as the disease progresses, the quality of life before and after re-transplantation and the survivability prognosis significantly worsen.

In order to manage CAV in an effective way then, patients and their physicians must make joint decisions concerning monitoring and re-transplantation that account for all the available information and the relevant trade-offs. Unfortunately, given the complexity of the disease and the scarcity of studies quantifying the onset and progression process, the International Society for Heart and Lung Transplantation (ISHLT) only provides crude recommendations for re-transplantation and monitoring, which are not personalized (see Costanzo et al. 2010 for details).

Our goal in this section is to show how our framework can be adopted to inform these monitoring and re-transplantation decisions. Although this study is limited in scope and should not be interpreted as a complete medical recommendation, we believe its value lies in setting the first steps towards a quantitative process for personalizing such decisions; we discuss the limitations and future research directions in §4.5.

⁹Re-vascularization by percutaneous coronary intervention or coronary artery bypass have also proven effective in some patients (Cheng and Kobashigawa 2016, Costanzo et al. 2010), but there is limited experience and very limited data associated with these procedures, so we do not consider them in our study.

4.1 Data Description

There is very limited data that can be used to guide CAV disease management decisions. To calibrate CAV development, we use a dataset from the Papworth Hospital that was compiled by monitoring 622 heart-transplanted patients on a yearly basis and recording their disease progression, starting from the transplant surgery.¹⁰ The length of the monitoring horizon across all patients varies from a few days to 20 years, with an average of 3.85 years and a standard deviation of 3.34 years. At each monitoring time, the patient’s CAV stage and the cumulative number of acute rejection episodes are recorded. CAV progression is classified into three stages—stage 1 (Not Significant), stage 2 (Mild) and stage 3 (Severe)—depending on the observed degree of angiographic lesion and allograft dysfunction. Overall, the dataset comprises 2,846 monitoring observations.

To calibrate survival after re-transplantation, we obtained data from Tjang (2007), Copeland et al. (2011), Schnetzler et al. (1998), Novitzky et al. (1985), in the form of 37 observations of survivability outcomes after the second transplantation, for different intervals between the two transplants.

4.2 Robust Monitoring

Consider a patient (“he”) who had a heart transplantation at $t = 0$, at an age of A . The patient and his physician need to decide when to monitor for CAV progression and when to perform a re-transplantation throughout a horizon T . To match our robust monitoring framework to this problem, we must characterize the rewards collected upon stopping (i.e., at re-transplantation) and the states and corresponding uncertainty sets.

These primitives critically depend on the degree of CAV progression, which we model first. Consistent with medical studies and with our available monitoring data, we assume that the patient’s disease progression at time t is determined by the CAV stage $i \in \{1, 2, 3\}$ and by the number of acute rejection episodes. We quantify the latter through an indicator $j \in \{L, H\}$, where “H” denotes that the patient suffered from a “High” number of acute rejection episodes – defined as four or more – and “L” denotes the complementary event.¹¹ This results in six disease stages, as depicted in Figure 2. We allow all “forward” transitions, i.e., from a state (i, j) to a state (i', j') with $i' \geq i$ and with $j' = H$ whenever $j = H$.

¹⁰Data is publicly available in an R package called ‘msm.’

¹¹We considered alternative quantizations for the number of rejections. Any quantization with more than two levels posed a very significant estimation burden given the available data, so we defined our binary variable to obtain the highest statistical power (see Appendix C).

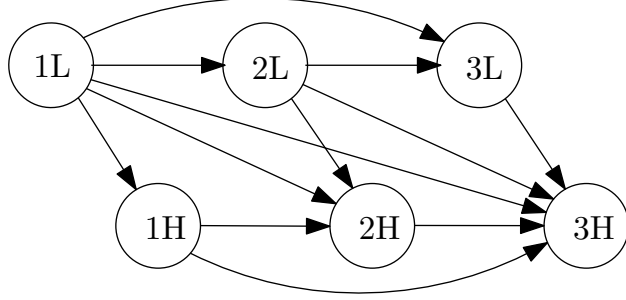


Figure 2: Disease progression model. The number $i \in \{1, 2, 3\}$ captures the CAV stage, and the letter $j \in \{L, H\}$ indicates whether the patient had a “Low” (< 4) or “High” (≥ 4) number of acute rejection episodes.

Rewards

We calculate the rewards from re-transplanting (i.e., “stopping”) at time t as the total Quality-Adjusted Life Years (QALY) provided that the patient is alive at t :

$$\begin{aligned} \text{reward at } t &= (\text{total QALY when re-transplanted at } t) \times (\text{Survival rate at } t) \\ &= ((\text{QALY until } t) + (\text{QALY after re-transplant at } t)) \times (\text{Survival rate at } t). \end{aligned}$$

By multiplying the total QALY with the survival rate, we incorporate the possibility that the patient may not have an option for re-transplantation due to death.¹²

Quality of life prior to re-transplantation is assumed to depend primarily on CAV stage. If we let $Z^i(t)$ denote the amount of time the patient spent in each CAV stage $i \in \{1, 2, 3\}$ by time t , and $LY(t)$ denote the expected life-years after re-transplantation at t , the total QALY when re-transplanting at time t can be expressed as:

$$\text{total QALY when re-transplanted at } t = \sum_i QOL_i \cdot Z^i(t) + QOL_{re} \cdot LY(t). \quad (14)$$

Here, the factors QOL_i and QOL_{re} adjust for quality of life at stage i and after re-transplantation, respectively. Table 2 in the Appendix reports the quality of life factors we use, which we approximated based on the related medical literature (Long et al. 2014, Feingold et al. 2015, Grauhan et al. 2010, Montazeri 2009, Johnson 2007). Note that consistent with intuition, disease progression (captured by higher CAV stages) induces a lower quality of life.

¹²To embed an additional degree of risk aversion in the formulation, our expression for QALYs does not include any reward if the patient is *not* alive at time t . Note that this death event could have also been equivalently captured by including an additional “Death” state in the disease progression model, with suitable transitions.

To estimate the life-years following re-transplantation, we rely on a substantial body of medical literature providing consistent evidence that the interval between transplants is the main statistically significant predictor (Schnetzler et al. 1998, John et al. 1999, Vistarini et al. 2010, Saito et al. 2013), and that re-transplants done too soon after the primary transplant tend to result in worse survival rates compared to those performed later (Saito et al. 2013, Johnson et al. 2007, Goldraich et al. 2016). In view of this, we consider the following piecewise linear model for the life-years after re-transplantation:

$$LY(t) = \beta_0 + \beta_1 * t + \beta_2 * (t - b)^+ + \epsilon, \quad (15)$$

which we fit using regression based on our re-transplantation data, to obtain:¹³

$$\hat{\beta}_0 = 2.1635, \hat{\beta}_1 = 1.0356(.), \hat{\beta}_2 = -1.7727, \hat{b} = 5.060. \quad (16)$$

Note that, consistent with the medical literature, the expected life-years first increase with t up to a time point \hat{b} , and then decrease.

Lastly, we estimate the survival rate through a linear regression model calibrated with our CAV monitoring data. More precisely, we regress the patient’s survival at post-transplant year t (with outcomes 1 = “alive” and 0 = “death”) on the patient’s starting age A and disease stage at the time of re-transplant (summarized by the CAV stage $i \in \{1, 2, 3\}$ and the indicator for acute rejections $j \in \{L, H\}$).¹⁴ The results are summarized in Table 3 of the Appendix.

Consistent with medical data as well as intuition, a patient who is older, has a more advanced form of CAV or suffered from 4 or more acute rejections has a worse survival rate. More importantly, Table 3 also shows that the survival rate decrease more as a result of an increase in the CAV stage than as a result of a jump in the number of acute rejections. Since this observation will be useful subsequently, we summarize it in the following remark.

Remark 1. *Given any state (i, L) with $i \leq 2$, a transition to $(i+1, L)$ always decreases the survival rate more than a transition to (i, H) .*

States and Uncertainty Sets for CAV Progression

To model disease progression, we assume that transitions between disease stages follow a Cox model, which we calibrate using our data. Although Cox models are among the most popular tools

¹³We use the following convention for p-values: ‘***’ for < 0.001 , ‘**’ for < 0.01 , ‘*’ for < 0.05 , and ‘.’ for < 0.1 .

¹⁴Technically, survival at t could depend also on how long the patient spent at each stage, before arriving at the stage he is at t . Calibrating such a path-dependent survival model, however, was not viable with our limited data, and we are therefore implicitly assuming here that survival depends only on stage at t .

for survival analysis in healthcare, their goodness of fit may be questionable in our setting: the underlying assumption that a unit increase in a covariate has a constant multiplicative effect on the transition time might not be reasonable for CAV, and the very limited data may yield very noisy estimates. For these reasons, we will only use the estimates of the Cox models to construct uncertainty sets.

In particular, we assume that the time until a patient transitions from one disease stage to another is exponentially distributed, with a mean that depends linearly on the patient’s age and on the time elapsed from the first transplant. More precisely, when a patient who received a transplant at an age of A is observed to be in stage (i, j) at the current (post-transplant) time t , the time until he progresses to stage (i', j') , denoted by $L_{ij'j'}(t, A)$, is exponentially distributed with rate $\lambda_{ij'j'}(t, A)$, where:

$$1/\lambda_{ij'j'}(t, A) = \beta_0^{ij'j'} + \beta_1^{ij'j'} * A + \beta_2^{ij'j'} * t. \quad (17)$$

Table 4 in the Appendix reports the coefficients obtained by fitting this model to our CAV data using maximum likelihood estimation, for all transitions in Figure 2. As expected, several of the coefficient estimates are quite noisy, which further justifies our robust approach.

To capture the disease progression in our non-probabilistic framework, we rely on this Cox model to construct confidence intervals for the amount of time spent in each disease state before a particular transition. More precisely, recalling that for any $\rho \in [0, 1]$, the ρ -quantile for an exponential random variable with rate λ is $\frac{-\ln(1-\rho)}{\lambda}$, we can construct a confidence interval for $L_{ij'j'}(t, A)$ at confidence level ρ of the form $[\frac{-\ln(\rho)}{\lambda_{ij'j'}(t, A)}, \infty)$.

Based on the progression model and the reward structure, a sufficient state for a given patient would be given by the patient’s personal history of disease progression.¹⁵ To bring this closer to our framework, we define the *state* at time t to consist of the number of years that the patient has spent (by time t) in each disease node (i, j) and above; we denote these quantities with $x^{ij}(t)$, for $i \in \{1, 2, 3\}, j \in \{L, H\}$. For example, referring to Figure 2, $x^{2H}(t)$ would be the number of years spent in stages 2H or 3H up to time t .

Since all patients are classified as CAV stage 1 and are free of organ rejections immediately after heart transplantation, we have $x^{1L}(t) = t$ and $x^{ij}(0) = 0, \forall i, j$, by definition.

In this context, it can be seen that writing the disease dynamics could become complicated, since certain nodes could lead to multiple possible *disease progression paths*, i.e., sequences of nodes connected by arcs in the graph of Figure 2. The following important observation will allow us to

¹⁵Since the patient’s starting age A does not change over time, it can be omitted from the state.

simply the formulation considerably.

Proposition 4. *Consider a patient with starting age $A \in [33, 62]$. Then:*

- (i) *the worst-case disease progression path does not involve nodes 1H and 2H;*
- (ii) *the variables $\mathbf{Z}(t)$ in (14), the CAV stage, and the indicator for acute rejections relate to the state variables $x^{ij}(t)$ as: $Z^1(t) = t - x^{2L}(t)$, $Z^2(t) = x^{2L}(t) - x^{3L}(t)$, $Z^3(t) = x^{3L}(t)$, CAV stage = $1 + \mathbf{1}_{x^{2L}(t)>0} + \mathbf{1}_{x^{3L}(t)>0}$, $\mathbf{1}_{j=H} = \mathbf{1}_{x^{3H}(t)>0}$.*

In view of Proposition 4, a sufficient state in our formulation is given by $(x^{2L}(t), x^{3L}(t), x^{3H}(t))$. Additionally, we can express the rewards for a patient who received an initial transplant at age $A \in [33, 62]$ and receives a re-transplant at time t as:

$$g(t, \mathbf{x}) = (0.8583t - 0.1445x^{2L}(t) - 0.1364x^{3L}(t) + 0.6456(2.1635 + 1.0356t - 1.7727(t - 5.060)^+)) \times [1.0641 - 0.0013 \cdot A - 0.0651(1 + \mathbf{1}_{x^{2L}(t)>0} + \mathbf{1}_{x^{3L}(t)>0}) - 0.03503 \cdot \mathbf{1}_{x^{3H}(t)>0}]. \quad (18)$$

This reward is monotonic in \mathbf{x} and therefore satisfies Assumption 2.

Furthermore, we can express the evolution of the states as a result of monitoring such a patient. Specifically, suppose the patient is monitored at time t , resulting in an observation $\mathbf{x}(t)$. When the patient is again monitored after an additional Δt time units, the possible values for $x^{2L}(t + \Delta t)$ are as follows:

1. If $x^{2L}(t) > 0$, the patient was already in stage 2L or above at time t . Therefore, $x^{2L}(t + \Delta t) = x^{2L}(t) + \Delta t$ without any uncertainty.
2. If $x^{2L}(t) = 0$, the patient must have been in stage 1L at time t . In this case, $x^{2L}(t) = (x^{1L}(t) - L_{1L2L}(t, A))^+$, where $L_{1L2L}(t, A)$ denotes the time spent in $(1, L)$ before transitioning to $(2, L)$. Recall that a confidence interval at level ρ for $L_{1L2L}(t, A)$ is given by $[-\frac{\ln(\rho)}{\lambda_{1L2L}(t, A)}, \infty)$. Since the reward function is decreasing in x^{2L} by (18), only the upper bound on x^{2L} is relevant for the worst-case. Thus, the observation of $\mathbf{x}(t)$ at time t imposes an upper bound for $x^{2L}(t + \Delta t)$ given by $\bar{x}^{2L}(t + \Delta t) | \mathbf{x}(t) = \left(\Delta t + \frac{\ln(\rho)}{\lambda_{1L2L}(t, A)} \right)^+$.

Applying this procedure, we can also express upper bounds for x^{2L} , x^{3L} and x^{3H} as:

$$\bar{x}^{2L}(t + \Delta t) | \mathbf{x}(t) = \begin{cases} \left(\Delta t + \frac{\ln(\rho)}{\lambda_{1L2L}(t,A)}\right)^+, & \text{if } x^{2L}(t) = 0 \\ x^{2L}(t) + \Delta t, & \text{otherwise;} \end{cases} \quad (19a)$$

$$\bar{x}^{3L}(t + \Delta t) | \mathbf{x}(t) = \begin{cases} \left(\Delta t + \frac{\ln(\rho)}{\lambda_{1L3L}(t,A)}\right)^+, & \text{if } x^{2L}(t) = 0, x^{3L}(t) = 0 \\ \left(\Delta t + \frac{\ln(\rho)}{\lambda_{2L3L}(t,A)}\right)^+, & \text{if } x^{2L}(t) > 0, x^{3L}(t) = 0 \\ x^{3L}(t) + \Delta t, & \text{otherwise;} \end{cases} \quad (19b)$$

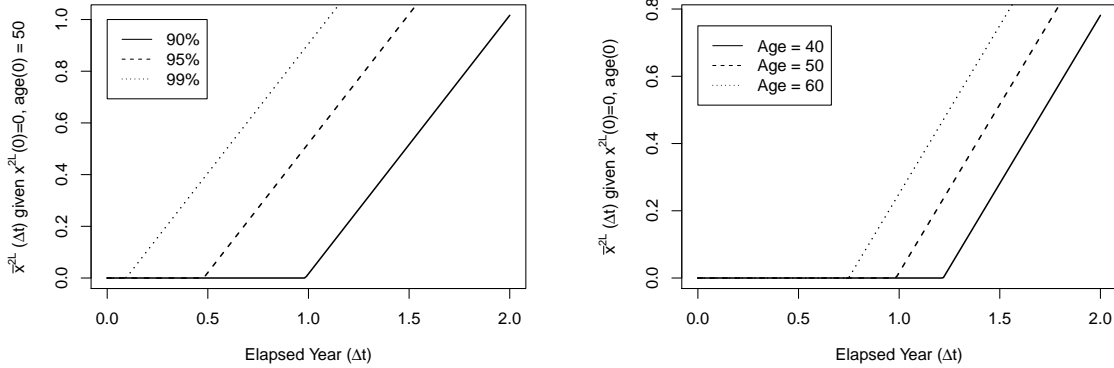
$$\bar{x}^{3H}(t + \Delta t) | \mathbf{x}(t) = \begin{cases} \left(\Delta t + \frac{\ln(\rho)}{\lambda_{1L3H}(t,A)}\right)^+, & \text{if } x^{2L}(t) = 0, x^{3L}(t) = 0, x^{3H}(t) = 0 \\ \left(\Delta t + \frac{\ln(\rho)}{\lambda_{2L3H}(t,A)}\right)^+, & \text{if } x^{2L}(t) > 0, x^{3L}(t) = 0, x^{3H}(t) = 0 \\ \left(\Delta t + \frac{\ln(\rho)}{\lambda_{3L3H}(t,A)}\right)^+, & \text{if } x^{2L}(t) > 0, x^{3L}(t) > 0, x^{3H}(t) = 0 \\ x^{3H}(t) + \Delta t, & \text{otherwise.} \end{cases} \quad (19c)$$

This implies that, for a fixed confidence level ρ , a state observation $\mathbf{x}(t)$ at time t imposes restrictions on the future state values at time $t + \Delta t$ specified by $x^{2L}(t + \Delta t) \leq \bar{x}^{2L}(t + \Delta t)$, $x^{3L}(t + \Delta t) \leq \bar{x}^{3L}(t + \Delta t)$ and $x^{3H}(t + \Delta t) \leq \bar{x}^{3H}(t + \Delta t)$. To complete our uncertainty sets, we also set the lower bounds on $(x^{2L}(t + \Delta t), x^{3L}(t + \Delta t), x^{3H}(t + \Delta t))$ to zero, without loss of generality. Since both the upper and the lower bounds are increasing in x^{2L} , x^{3L} , x^{3H} and Δt , and are concave and convex in x^{2L} , x^{3L} and x^{3H} , respectively, the resulting uncertainty sets satisfy Assumption 1.¹⁶

Naturally, the uncertainty sets we construct here depend on the confidence level ρ and the patient's initial age A . Higher confidence level produces more conservative upper bounds, resulting in higher worst-case values for $x^{2L}(t + \Delta t)$, $x^{3L}(t + \Delta t)$ and $x^{3H}(t + \Delta t)$. For instance, Figure 3a illustrates how the upper bounds for $x^{2L}(\Delta t)$ imposed by the initial observation $(x^{2L}(0), x^{3L}(0), x^{3H}(0)) = (0, 0, 0)$ depend on the confidence level ρ . Second, the negative values of β_1^{1L2L} , β_1^{2L3L} and β_1^{3L3H}

¹⁶Several technical remarks may be in order. First, since the reward g is decreasing in x^{2L} , x^{3L} and x^{3H} by (18), the lower bounds are irrelevant for determining the worst-case rewards. Our choice of zero thus only serves to ensure that the resulting uncertainty sets satisfy Assumption 1. Second, note that according to (19a)–(19c), the upper bounds are discontinuous in (x^{2L}, x^{3L}, x^{3H}) as we approach the boundary $\{x^{2L}(t) = 0\} \cup \{x^{3L}(t) = 0\} \cup \{x^{3H}(t) = 0\}$. However, since $\log \rho < 0$, the functions are upper-semicontinuous at the boundary, and are thus concave in (x^{2L}, x^{3L}, x^{3H}) . To rigorously prove that our results in Theorem 1 would hold, we could introduce an intermediate region in (19a) for $0 < x^{2L}(t) < \epsilon$ (for $\epsilon > 0$), where $\bar{x}^{2L}(t + \Delta t) | x^{2L}(t)$ is defined by linearly interpolating between the value at $x^{2L}(t) = 0$ and the value at $x^{2L}(t) + \Delta t$ (and similarly for $\bar{x}^{3L}(t + \Delta t)$ and $\bar{x}^{3H}(t + \Delta t)$). For any $\epsilon > 0$, one can check that Assumption 1 is satisfied, and Theorem 1 thus holds. Since the worst-case optimal values under both static and dynamic monitoring would be continuous functions of ϵ and have equal values for any $\epsilon > 0$, their limiting values as $\epsilon \rightarrow 0$ must be equal, so Theorem 1 holds for the uncertainty set given by the upper bounds in (19a)–(19c).

(per Table 4) imply that older patients are more likely to have shorter transition times; thus, the worst-case values for $x^{2L}(t + \Delta t)$, $x^{3L}(t + \Delta t)$ and $x^{3H}(t + \Delta t)$ are increasing in the initial age A , as shown in Figure 3b.



(a) Worst-case values for $x^{2L}(\Delta t)$ as a function of the confidence level ρ (starting age $A = 50$). (b) Worst-case values for $x^{2L}(\Delta t)$ as a function of the patient's starting age A ($\rho = 90\%$).

Figure 3: The worst-case values for $x^{2L}(\Delta t)$ given an initial observation at $t = 0$.

4.3 Monitoring Policy

To solve the dynamic monitoring problem using the algorithm described in Corollary 1, it suffices to solve a series of static monitoring problems. Based on our analysis, for patients who transplanted at age 33-62 and were diagnosed as stage 1L, 2L, 3L or 3H, the static monitoring problem with n monitoring times t_1, \dots, t_n chosen at time 0 can be written as

$$\max_{t_{\{n+1\}}} \max \left(g(t_n, \bar{x}^{2L}(t_n), \bar{x}^{3L}(t_n), \bar{x}^{3H}(t_n)), g(t_{n+1}, \bar{x}^{2L}(t_{n+1}), \bar{x}^{3L}(t_{n+1}), \bar{x}^{3H}(t_{n+1})) \right) \quad (20)$$

where $\bar{x}^{2L}(t_k)$, $\bar{x}^{3L}(t_k)$ and $\bar{x}^{3H}(t_k)$ satisfy (19a) – (19c) with $t := t_k$, $\Delta t := t_{k+1} - t_k$.

Since the reward function and the bound functions defining the uncertainty sets are all piecewise linear, we can formulate Problem (20) as a mixed-integer linear optimization problem using a classical big-M formulation. Further details can be found in Appendix E.

4.4 Results

We consider CAV management for a patient who received heart transplantation at an age of 33 to 62. According to the ISHLT, the patient would be recommended to undergo annual or bi-annual

examinations of coronary arteries through coronary angiography and intravascular ultrasonography to monitor progression in the first 3-5 years after transplant (Labarrere et al. 2017, Costanzo et al. 2010), and perform re-transplantation if CAV reaches stage 3 (Johnson 2007). We compare this with the policies prescribed by our framework.

Static Monitoring. Given that ISHLT essentially prescribes a static monitoring policy, we begin by determining the optimal static monitoring policy in our framework. We consider a time horizon of $T = 10$ years. To compare with ISHLT prescription for yearly examinations, we derive a static monitoring policy that allows for $n = 9$ intermediate monitoring times. Figure 4 illustrates our policy, depending on the patient’s age at the first transplant (A) and the level of conservativeness in the formulation (measured by the confidence level ρ).

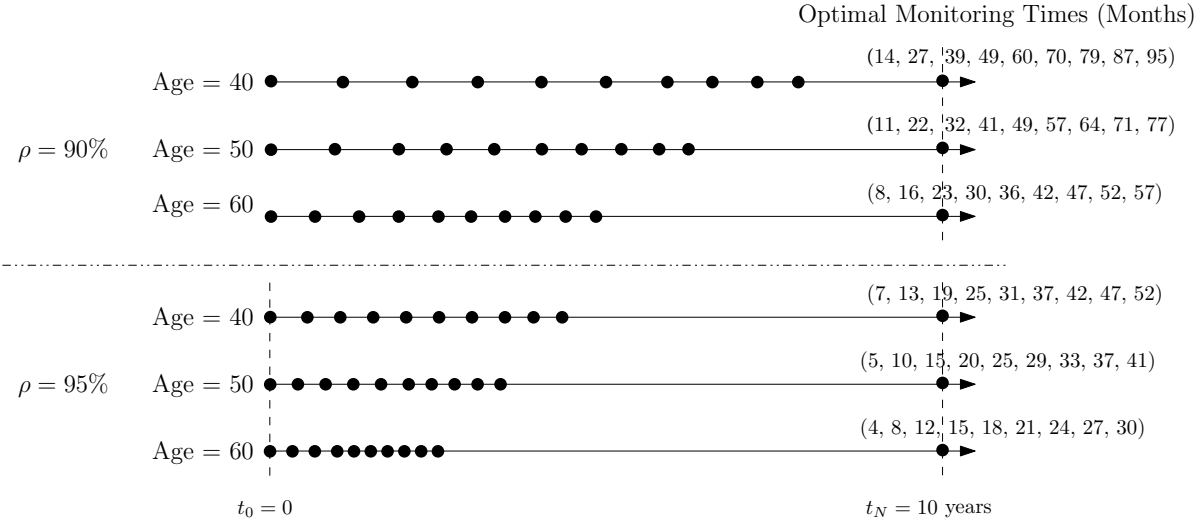


Figure 4: Monitoring times (in months) for static monitoring prescribed by our framework.

Notably, this policy departs from equi-sized monitoring and suggests a frequency that increases as time goes by, at a rate that also increases with a patient’s starting age and conservativeness. Intuitively, this arises since an increased age and a sufficient time following the primary transplant lead to a significant worsening in all the patient’s prospects, in terms of disease progression (per equation (17) and Table 4), life-years following re-transplantation (per (16)), and also survival rate (per Table 3). Thus, monitoring more aggressively could more quickly identify that the disease is following a worst-case progression, which would warrant an earlier transplant.

Dynamic Monitoring. It is worth recalling that the monitoring times in Figure 4 correspond to the optimal solution for the static problem; thus, all monitoring times beyond the first one are optimal only along the *worst-case* progression path, i.e., when the most unfortunate disease progression

occurs over time. In practice, such an unfortunate sequence of outcomes may not materialize, so a physician relying on our framework may prefer to only implement the *first* monitoring time, and to always re-evaluate all the remaining monitoring time(s) after observing each monitoring outcome. This is essentially equivalent to finding the optimal *dynamic* monitoring policy in our framework, which is done by repeatedly re-solving static monitoring problems, as explained in Corollary 1.¹⁷

Our framework can also be used to analyze how the recommendation for when to monitor next changes as a function of observed characteristics. Table 1 reports the next monitoring time depending on a patient’s age, CAV stage, and conservativeness (we again assume $n = 9$ for a fair comparison). For example, when a 50-year-old patient is diagnosed with CAV stage 1, the next monitoring time would be 11 months later with a low number of acute rejections, and 7 months later with a high number of acute rejections (for $\rho = 90\%$). But when he is diagnosed with stage 2, he should be monitored earlier: in 5 months for a low number of acute rejections, and in 4 months with a high number of acute rejections.

Number of acute rejections	Low		High	
CAV progression	Stage 1	Stage 2	Stage 1	Stage 2
40 years old	14	5	8	4
50 years old	11	5	7	4
60 years old	8	3	6	3

Table 1: The worst-case optimal next monitoring time in months ($\rho = 90\%$).

Simulation. To assess the performance of our policies, we simulate CAV progression using the transition probabilities we estimated when calibrating the Cox models. In particular, transition times are randomly generated according to an exponential distribution ($T = 10$ years, initial age = 50, number of iterations = 10^3).

We simulate both our static and dynamic monitoring policies ($\rho = 90\%$), and compare these with two benchmark policies. First, a policy that monitors annually and proceeds to re-transplant when CAV deteriorates to stage 3. This policy is in line with the ISHLT guidelines, and we therefore refer to it with that name. Second, a policy that also monitors uniformly every year, but stops

¹⁷Note that in implementing this, one should also consider non-worst-case paths, whereby a patient could be diagnosed in stage 1H, 2H or 3H. In such cases, the disease must have progressed from 1L to 1H, to 2H and to 3H, and we can follow a similar modeling approach to formulate the corresponding uncertainty sets and objective, and to solve the static monitoring problem as an IP. Details are presented in Appendix E.

when the current stopping reward exceeds the worst-case reward-to-go. This robust policy, which we refer to as Uniform RO, is motivated by Nohadani and Roy (2017), given that it relies on uniform monitoring and on a judicious (in fact, an optimal) robust criterion for when to stop.

The results are summarized in Figure 5 (Table 5 of the Appendix provides additional summary statistics). Both the static and dynamic RO policies outperform the ISHLT guidelines, by yielding noticeable improvements for lower percentiles and slightly better average rewards than the guidelines. We can also observe that the RO policies sacrifice from the higher percentiles due to following a more conservative stopping/re-transplantation policy. The maximum reward from ISHLT guidelines is higher than that from Uniform RO, static RO and dynamic RO, although dynamic RO can partly overcome the problem by resolving. This shows that a dynamic policy can achieve the best of both worlds, by taking both a conservative approach that guards against worst-case outcomes, while also allowing for sufficient flexibility to deal with improved conditions.

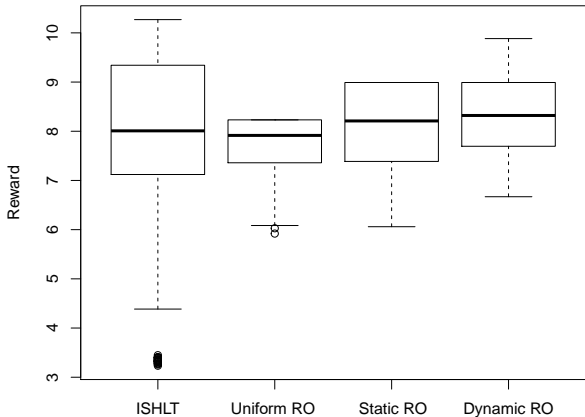


Figure 5: Boxplot for simulated rewards under the benchmark, static and dynamic RO policies ($\rho = 90\%$, Number of Simulations = 10^3 , Number of Monitoring Times = 10, $T = 10$ years).

4.5 Limitations

Despite the encouraging results, it should be noted that our numerical study suffers from a number of limitations that should be addressed before turning it into a policy recommendation. We include a brief discussion here, in the hope that this would spur subsequent research on these important topics.

First, to more accurately capture the progression for CAV, one should consider all the available

information concerning the state of the patient, including, e.g., the presence of other risk factors such as hypertension, diabetes or other vascular conditions that could precipitate the development of CAV. Additionally, the patient’s state should also include biometrics that monitor these conditions, as well as information about any medication prescribed to address these. Making this meaningful in our models also requires extensive randomized controlled trials to quantify the impact of these biometrics on CAV progression.

Second, our study could benefit by considering other alternative treatments for advanced CAV, beyond re-transplantation. One such candidate is re-vascularization (Cheng and Kobashigawa 2016), which has proven effective for certain patients. Assuming enough information exists to relate the performance of this procedure to measurable biometrics, one could extend our framework to allow for a more flexible “stopping” decision, which could be either re-vascularization or re-transplantation.

Third, our framework only allowed addressing re-transplantation decisions for patients who received a transplant at age 33-62. Although this covers a significant fraction of heart transplants, important cases (such as pediatric transplants) are left out. To extend our framework to these may also require the development of new theory that deals with cases where the disease evolution can follow multiple discrete paths, with distinct uncertainty sets corresponding to each path.

Lastly, it would be valuable to make the choice of planning horizon T more explicit, and to also allow an explicit choice of the number of monitoring chances n (instead of fixing it, as we currently do). This could be achieved by first quantifying the cost of a single monitoring episode more explicitly, and relying on the extension in §B.1 to calculate both the optimal n and the optimal monitoring policy.

5 Conclusions and Future Research

This paper considered a framework for determining optimal monitoring and stopping policies for a system evolving over time. Such stopping problems arise in a variety of applications from healthcare to finance, and often require excessive amounts of data for calibration purposes, and prohibitive computational resources. To overcome these challenges, our paper proposed a robust optimization approach, whereby adaptive uncertainty sets capture the information acquired through monitoring. We considered two versions of the problem – static and dynamic – depending on how the monitoring times are chosen, and showed that under certain conditions, the same worst-case reward is achievable under either static or dynamic monitoring. This allowed recovering the optimal dynamic

monitoring policy by resolving static versions of the problem. The paper also discussed cases when the static problem becomes tractable, and highlighted conditions when monitoring at equi-distant times is optimal. Lastly, the paper showcased the framework in the context of a healthcare problem (monitoring heart-transplant patients for Cardiac Allograft Vasculopathy), where we designed policies that showed promise relative to the status-quo recommendations.

The paper also opens several directions for future research. On a theoretical front, understanding how learning can be done under more general conditions seems very relevant. This could include cases where more information is available (e.g., beyond the support of the distributions), dynamical systems that do not abide by the paper's assumptions (such as mean-reverting processes or rewards that are not monotonic in the states), or processes that can exhibit jumps (which would warrant models where the adversary has access to discrete decisions). Furthermore, the case study could be expanded by including additional risk factors and also multiple treatment options, as discussed in 4.5.

References

- Ayer, Turgay, Oguzhan Alagoz, Natasha K Stout. 2012. OR forum – a POMDP approach to personalize mammography screening decisions. *Operations Research* **60**(5) 1019–1034.
- Bandi, Chaithanya, Dimitris Bertsimas. 2012. Tractable stochastic analysis in high dimensions via robust optimization. *Mathematical Programming* **134** 23–70.
- Bayraktar, E., I. Karatzas, S. Yao. 2010. Optimal stopping for dynamic convex risk measures. *Illinois Journal of Mathematics* **54**(3) 1025–1067.
- Bayraktar, Erhan, Song Yao. 2014. On the robust optimal stopping problem. *SIAM Journal on Control and Optimization* **52**(5) 3135–3175.
- Ben-Tal, Aharon, Laurent El Ghaoui, Arkadi Nemirovski. 2009. *Robust optimization*. Princeton University Press.
- Ben-Tal, Aharon, A. Goryashko, E. Guslitzer, Arkadi Nemirovski. 2004. Adjustable robust solutions of uncertain linear programs. *Mathematical Programming* **99**(2) 351–376.
- Ben-Tal, Aharon, Arkadi Nemirovski. 2002. Robust optimization – methodology and applications. *Mathematical Programming* **92**(3) 453–480.
- Bertsimas, D., V. Goyal, B. Lu. 2014. A tight characterization of the performance of static solutions in two-stage adjustable robust linear optimization. *Mathematical Programming* **150**(2) 281–319.
- Bertsimas, D., P. Vayanos. 2017. Data-driven learning in dynamic pricing using adaptive optimization. Working paper.
- Bertsimas, Dimitris, David B Brown, Constantine Caramanis. 2011. Theory and applications of robust optimization. *SIAM review* **53**(3) 464–501.
- CH. 2014. *Accounts Receivable and Inventory Financing*. Comptroller of the currency, administrator of national banks.
- Cheng, Richard, Jon Kobashigawa. 2016. Transplant CAD: The Role of PCI in the Transplanted Heart. *American College of Cardiology, Expert Analysis* .
- Copeland, Hannah, Romana Coelho-Anderson, Nicole Mineburg, Michael McCarthy, Jack G Copeland. 2011. Elective cardiac retransplantation: a viable option that can be repeated. *The Journal of thoracic and cardiovascular surgery* **141**(3) 822–827.
- Costanzo, Maria Rosa, Anne Dipchand, Randall Starling, Allen Anderson, Michael Chan, Shashank Desai, Savitri Fedson, Patrick Fisher, Gonzalo Gonzales-Stawinski, Luigi Martinelli, et al. 2010. The International Society of Heart and Lung Transplantation Guidelines for the care of heart transplant recipients.
- Cutler, David M, James M Poterba, Lawrence H Summers. 1991. Speculative dynamics. *The Review of Economic Studies* **58**(3) 529–546.

- Delage, E., D. A. Iancu. 2015. Robust multistage decision making. *Tutorials in Operations Research* 20–46.
- Denton, Brian T, Murat Kurt, Nilay D Shah, Sandra C Bryant, Steven A Smith. 2009. Optimizing the start time of statin therapy for patients with diabetes. *Medical Decision Making* **29**(3) 351–367.
- Dullerud, G. E., F. Paganini. 2005. *A Course in Robust Control Theory*. Springer.
- Feingold, B, SA Webber, CL Bryce, SY Park, HE Tomko, SC West, SA Hart, WT Mahle, KJ Smith. 2015. Cost-effectiveness of pediatric heart transplantation across a positive crossmatch for high waitlist urgency candidates. *American Journal of Transplantation* **15**(11) 2978–2985.
- Foley, C.F., A. Raman, N.C. Craig. 2013. Inventory-based lending industry note.
- Fujishige, Satoru. 2005. *Submodular Functions and Optimization, Annals of Discrete Mathematics*, vol. 58. 2nd ed. Elsevier.
- Goldraich, LA, Josef Stehlik, AY Kucheryavaya, LB Edwards, HJ Ross. 2016. Retransplant and medical therapy for cardiac allograft vasculopathy: International society for Heart and Lung Transplantation Registry analysis. *American Journal of Transplantation* **16**(1) 301–309.
- Gorton, Gary, Andrew Winton. 2003. Financial intermediation. *Handbook of the Economics of Finance* **1** 431–552.
- Grauhan, O, H Huang Chang, W Albert, N Hiemann, R Meyer, H Lehmkuhl, M Dandel, C Knosalla, M Hübler, M Pasic, et al. 2010. Impact of donor-transmitted coronary atherosclerosis on quality of life (QOL) and quality-adjusted life years (QALY) after heart transplantation. *Applied Cardiopulmonary Pathophysiology* **14** 58–65.
- Hanany, Eran, Peter Klibanoff. 2009. Updating ambiguity averse preferences. *The B.E. Journal of Theoretical Economics* **9**(1).
- Healthcare Bluebook. 2019. Coronary angiogram. Available online at <https://www.mdsave.com/procedures/cardiac-catheterization-with-coronary-angiogram/d785fec8>.
- Iancu, Dan A, Mayank Sharma, Maxim Sviridenko. 2013. Supermodularity and affine policies in dynamic robust optimization. *Operations Research* **61**(4) 941–956.
- Ibanez, Alfredo, Fernando Zapatero. 2004. Monte Carlo valuation of American options through computation of the optimal exercise frontier. *Journal of Financial and Quantitative Analysis* **39**(2) 253–275.
- Jaillet, P., S.D. Jena, T.S. Ng, M. Sim. 2016. Satisficing awakens: Models to mitigate uncertainty. Working paper.
- John, Ranjit, Jonathan M Chen, Alan Weinberg, Mehmet C Oz, Donna Mancini, Silviu Itescu, Mark E Galantowicz, Craig R Smith, Eric A Rose, Niloo M Edwards. 1999. Long-term survival after cardiac retransplantation: a twenty-year single-center experience. *The Journal of thoracic and cardiovascular surgery* **117**(3) 543–555.
- Johnson, Maryl R. 2007. When is retransplantation a viable option? *Heart failure clinics* **3**(1) 97–105.

- Johnson, MR, Keith D Aaronson, CE Canter, JK Kirklin, DM Mancini, MR Mehra, B Radovancevic, DO Taylor, SA Webber. 2007. Heart retransplantation. *American Journal of Transplantation* **7**(9) 2075–2081.
- Karatzas, Ioannis, Steven Shreve. 2012. *Brownian motion and stochastic calculus*, vol. 113. Springer Science & Business Media.
- Karatzas, Ioannis, Steven E Shreve, I Karatzas, Steven E Shreve. 1998. *Methods of mathematical finance*, vol. 39. Springer.
- Kobashigawa, Jon. 2000. What is the optimal prophylaxis for treatment of cardiac allograft vasculopathy? *Trials* **1**(3) 166–171.
- Kobashigawa, Jon, Daniel Pauly, Randall C. Starling, Howard Eisen, Heather Ross, Shoei-Shen Wang, Bernard Cantin, James Hill, Patricia Lopez, Gaohong Dong, Stephen J. Nicholls. 2013. Cardiac allograft vasculopathy by intravascular ultrasound in heart transplant patients. *JACC: Heart Failure* **1**(5) 389–399.
- Kolodko, Anastasia, John Schoenmakers. 2006. Iterative construction of the optimal Bermudan stopping time. *Finance and Stochastics* **10**(1) 27–49.
- Labarrere, CA, BR Jaeger, GS Kassab. 2017. Cardiac allograft vasculopathy: Microvascular arteriolar capillaries (capioles) and survival. *Frontiers in bioscience (Elite edition)* **9** 110–128.
- Lavieri, Mariel, Martin Puterman, Scott Tyldesley, William Morris. 2012. When to treat prostate cancer patients based on their PSA dynamics. *IIE Transactions on Healthcare Systems Engineering* **2**(1) 62–77.
- Leland, Hayne E. 1994. Corporate debt value, bond covenants, and optimal capital structure. *The journal of finance* **49**(4) 1213–1252.
- Leland, Hayne E, Klaus Bjerre Toft. 1996. Optimal capital structure, endogenous bankruptcy, and the term structure of credit spreads. *The Journal of Finance* **51**(3) 987–1019.
- Long, Elisa F, Gary W Swain, Abeel A Mangi. 2014. Comparative survival and cost effectiveness of advanced therapies for end-stage heart failure. *Circulation: Heart Failure* .
- Machina, Mark. 1989. Dynamic consistency and non-expected utility models of choice under uncertainty. *Journal of Economic Literature* **27** 1622–1668.
- Maillart, Lisa M, Julie Simmons Ivy, Scott Ransom, Kathleen Diehl. 2008. Assessing dynamic breast cancer screening policies. *Operations Research* **56**(6) 1411–1427.
- Marandi, A., Dick den Hertog. 2017. When are static and adjustable robust optimization problems with constraint-wise uncertainty equivalent? *Mathematical Programming, Series A* .
- Mayo Clinic. 2019. Coronary angiogram. Available at <https://www.mayoclinic.org/tests-procedures/coronary-angiogram/about/pac-20384904>.

- Montazeri, Ali. 2009. Quality of life data as prognostic indicators of survival in cancer patients: an overview of the literature from 1982 to 2008. *Health and quality of life outcomes* **7**(1) 102.
- Nohadani, O., A. Roy. 2017. Robust optimization with time-dependent uncertainty in radiation therapy. *IIESE Transactions on Healthcare Systems Engineering* **7**(2) 81–92.
- Nohadani, O., K. Sharma. 2016. Optimization under decision-dependent uncertainty. Working paper.
- Novitzky, D, DKC Cooper, RP Lanza, CN Barnard. 1985. Further cardiac transplant procedures in patients with heterotopic heart transplants. *The Annals of thoracic surgery* **39**(2) 149–154.
- Poss, M. 2013. Robust combinatorial optimization with variable budgeted uncertainty. *4OR: A Quarterly Journal of Operations Research* **11**(1) 75–92.
- Queyranne, M., F. Tardella. 2006. Bimonotone linear inequalities and sublattices of r^n . *Linear Algebra and Its Applications* **413** 100–120.
- Rajan, Raghuram, Andrew Winton. 1995. Covenants and collateral as incentives to monitor. *The Journal of Finance* **50**(4) 1113–1146.
- Ramzy, Danny, Vivek Rao, Julie Brahm, Santiago Miriuka, Diego Delgado, Heather J Ross. 2005. Cardiac allograft vasculopathy: a review. *Canadian journal of surgery* **48**(4) 319.
- Riedel, Frank. 2009. Optimal stopping with multiple priors. *Econometrica* **77**(3) 857–908.
- Saito, Aya, Richard J Novick, Bob Kiaii, F Neil McKenzie, Mackenzie Quantz, Peter Pflugfelder, Grant Fisher, Michael WA Chu. 2013. Early and late outcomes after cardiac retransplantation. *Canadian Journal of Surgery* **56**(1) 21.
- Schnetzler, Bruno, Alain Pavie, Richard Dorent, Anne-Claude Camproux, Philippe Leger, Annick Delcourt, Iradj Gandjbakhch. 1998. Heart retransplantation: a 23-year single-center clinical experience. *The Annals of thoracic surgery* **65**(4) 978–983.
- Schrijver, Alexander. 2003. *Combinatorial Optimization: Polyhedra and Efficiency*. 1st ed. Springer.
- Shechter, Steven M, Matthew D Bailey, Andrew J Schaefer, Mark S Roberts. 2008. The optimal time to initiate HIV therapy under ordered health states. *Operations Research* **56**(1) 20–33.
- Snell, James Laurie. 1952. Applications of martingale system theorems. *Transactions of the American Mathematical Society* **73**(2) 293–312.
- Taylor, Howard M. 1968. Optimal stopping in a Markov process. *The Annals of Mathematical Statistics* 1333–1344.
- Tjang, Yanto Sandy. 2007. Cardiac retransplantation. Ph.D. Thesis.
- Topkis, Donald M. 1998. *Supermodularity and Complementarity*. Princeton University Press.
- Vistarini, Nicola, Carlo Pellegrini, Marco Aiello, Alessia Alloni, Cristian Monterosso, Barbara Cattadori, Carmine Tinelli, Andrea M D’Armini, Mario Vigano. 2010. Should we perform heart retransplantation in early graft failure? *Transplant International* **23**(1) 47–53.

- Wilhelm, Markus J. 2015. Long-term outcome following heart transplantation: current perspective. *Journal of thoracic disease* **7**(3) 549.
- Wilmott, P., J. Dewynne, S. Howison. 1994. *Option Pricing: Mathematical Models and Computation*. Oxford Financial Press.
- Zhang, X., M. Kamgarpour, A. Georghiou, P. Goulart, J. Lygeros. 2016. Robust optimal control with adjustable uncertainty sets. *Automatica* **75** 249–259.
- Zhou, K., J. C. Doyle. 1998. *Essentials of Robust Control*. Prentice Hall.

A Proofs

Proof. Proof of Proposition 1. It suffices to show that for any $\mathbf{x}^{\{k\}}$ and monitorings $\mathbf{t}^{\{r\}}, \mathbf{t}^{\{r'\}}$ ($k < r < r' \leq n$),

$$\Pi_{:,i:j}\mathcal{U}(\mathbf{t}^{\{r\}}, \mathbf{x}^{\{k\}}) = \Pi_{:,i:j}\mathcal{U}(\mathbf{t}^{\{r'\}}, \mathbf{x}^{\{k\}}), \forall i \leq j < r - k + 1.$$

Note that because introducing more monitoring times imposes more constraints and thereby shrinks the uncertainty sets and their projections, we get that $\Pi_{:,i:j}\mathcal{U}(\mathbf{t}^{\{r\}}, \mathbf{x}^{\{k\}}) \supseteq \Pi_{:,i:j}\mathcal{U}(\mathbf{t}^{\{r'\}}, \mathbf{x}^{\{k\}})$, for any $r < r'$.

To prove the opposite direction, we need to show that if $\mathbf{B} \in \Pi_{:,i:j}\mathcal{U}(\mathbf{t}^{\{r\}}, \mathbf{x}^{\{k\}})$, then $\mathbf{B} \in \Pi_{:,i:j}\mathcal{U}(\mathbf{t}^{\{r'\}}, \mathbf{x}^{\{k\}})$, i.e., $\exists \mathbf{A} \in \mathbb{R}^{d \times (i-1)}, \mathbf{C} \in \mathbb{R}^{d \times (r'-k+1-j)}$ such that $[\mathbf{A}, \mathbf{B}, \mathbf{C}] \in \mathcal{U}(\mathbf{t}^{\{r'\}}, \mathbf{x}^{\{k\}})$. But, $\mathbf{B} \in \Pi_{:,i:j}\mathcal{U}(\mathbf{t}^{\{r\}}, \mathbf{x}^{\{k\}})$ implies that $\exists \overline{\mathbf{A}} \in \mathbb{R}^{d \times (i-1)}, \overline{\mathbf{C}} \in \mathbb{R}^{d \times (r-k+1-j)}$ such that $[\overline{\mathbf{A}}, \mathbf{B}, \overline{\mathbf{C}}] \in \mathcal{U}(\mathbf{t}^{\{r\}}, \mathbf{x}^{\{k\}})$. Let us denote the columns of these matrices as $\overline{\mathbf{A}} = [\mathbf{x}_{k+1}, \dots, \mathbf{x}_{k+i-1}]$, $\mathbf{B} = [\mathbf{x}_{k+i}, \dots, \mathbf{x}_{k+j}]$ and $\overline{\mathbf{C}} = [\mathbf{x}_{k+j+1}, \dots, \mathbf{x}_r, \mathbf{x}_{n+1}]$.

Let $\mathbf{A} = \overline{\mathbf{A}}$, and construct \mathbf{C} by starting off from $\overline{\mathbf{C}}$, replacing its last column with \mathbf{x}_r , and then augmenting it with another $r' - r$ copies of \mathbf{x}_r , that is

$$\mathbf{C} = [\mathbf{x}_{k+j+1}, \dots, \mathbf{x}_r, \mathbf{x}_r, \underbrace{\mathbf{x}_r, \dots, \mathbf{x}_r}_{r'-r \text{ times}}].$$

To complete the proof we shall show that $[\mathbf{A}, \mathbf{B}, \mathbf{C}] \in \mathcal{U}(\mathbf{t}^{\{r'\}}, \mathbf{x}^{\{k\}})$. To that end, we need to verify that $f(t_p, t_q, \mathbf{x}_p, \mathbf{x}_q) \leq 0$, for all $p, q \in \{0, 1, \dots, r', n+1\}$ and $p < q$, where by the construction of \mathbf{C} we have that $\mathbf{x}_v = \mathbf{x}_r$, for all $v > r$. First, for $q \leq r$, $f(t_p, t_q, \mathbf{x}_p, \mathbf{x}_q) \leq 0$ is trivially satisfied since $[\overline{\mathbf{A}}, \mathbf{B}, \overline{\mathbf{C}}] \in \mathcal{U}(\mathbf{t}^{\{r\}}, \mathbf{x}^{\{k\}})$. Second, for $p \leq r < q$, we need to check that $f(t_p, t_q, \mathbf{x}_p, \mathbf{x}_r) \leq 0$, which is true given that $f(t_p, t_r, \mathbf{x}_p, \mathbf{x}_r) \leq 0$ holds by $[\overline{\mathbf{A}}, \mathbf{B}, \overline{\mathbf{C}}] \in \mathcal{U}(\mathbf{t}^{\{r\}}, \mathbf{x}^{\{k\}})$, in conjunction with t_q being a later monitoring time than t_r and assumption (2). Therefore, we also get that $f(t_q, t_q, \mathbf{x}_r, \mathbf{x}_r) \leq 0$ for $q > r$. Finally, for $r < p$, we need to check that $f(t_p, t_q, \mathbf{x}_r, \mathbf{x}_r) \leq 0$, which follows from the previous point and (2). \blacksquare \square

Proof. Proof of Theorem 1. It is trivial that $J \geq V$, since any optimal solution for the static model is also feasible for the dynamic model. To show the opposite direction, define

$$\begin{aligned} J_k(\mathbf{t}^{\{k\}}, \mathbf{x}^{\{k\}}) &= \max\left(g(t_k, \mathbf{x}_k), C_k(\mathbf{t}^{\{k\}}, \mathbf{x}^{\{k\}})\right) \\ C_k(\mathbf{t}^{\{k\}}, \mathbf{x}^{\{k\}}) &\stackrel{\text{def}}{=} \max_{t_{k+1} \in [t_k, T]} G_{k+1}(\mathbf{t}^{\{k+1\}}, \mathbf{x}^{\{k\}}) \\ G_{k+1}(\mathbf{t}^{\{k+1\}}, \mathbf{x}^{\{k\}}) &\stackrel{\text{def}}{=} \min_{\mathbf{x}_{k+1} \in \mathcal{U}_{k+1}(\mathbf{t}^{\{k+1\}}, \mathbf{x}^{\{k\}})} J_{k+1}(\mathbf{t}^{\{k+1\}}, \mathbf{x}^{\{k+1\}}). \end{aligned}$$

The proof relies on the following auxiliary result, which we prove in Lemma 1: for any $\mathbf{t}^{\{k+1\}}$ and $\mathbf{x}^{\{k-1\}}$, with $\underline{\mathbf{x}}_k \stackrel{\text{def}}{=} \min\{\mathbf{x} : \mathbf{x} \in \mathcal{U}_k(\mathbf{t}^{\{k\}}, \mathbf{x}^{\{k-1\}})\}$, we have

$$\underline{\mathbf{x}}_k \in \arg \min_{\mathbf{x}_k \in \mathcal{U}_k(\mathbf{t}^{\{k\}}, \mathbf{x}^{\{k-1\}})} g(t_k, \mathbf{x}_k) \cap \arg \min_{\mathbf{x}_k \in \mathcal{U}_k(\mathbf{t}^{\{k\}}, \mathbf{x}^{\{k-1\}})} C_k(\mathbf{t}^{\{k\}}, \mathbf{x}^{\{k\}}) \cap \arg \min_{\mathbf{x}_k \in \mathcal{U}_k(\mathbf{t}^{\{k+1\}}, \mathbf{x}^{\{k-1\}})} G_{k+1}(\mathbf{t}^{\{k+1\}}, \mathbf{x}^{\{k\}}). \quad (21)$$

We then have:

$$\begin{aligned} G_k(\mathbf{t}^{\{k\}}, \mathbf{x}^{\{k-1\}}) &= \min_{\mathbf{x}_k \in \mathcal{U}_k(\mathbf{t}^{\{k\}}, \mathbf{x}^{\{k-1\}})} J_k(\mathbf{t}^{\{k\}}, \mathbf{x}^{\{k\}}) \\ &\stackrel{(21)}{=} \max\left(g(t_k, \underline{\mathbf{x}}_k), C_k(\mathbf{t}^{\{k\}}, [\mathbf{x}^{\{k-1\}}, \underline{\mathbf{x}}_k])\right) \\ &= \max\left(g(t_k, \underline{\mathbf{x}}_k), \max_{t_{k+1} \in [t_k, T]} G_{k+1}(\mathbf{t}^{\{k+1\}}, [\mathbf{x}^{\{k-1\}}, \underline{\mathbf{x}}_k])\right) \\ (\text{since } \underline{\mathbf{x}}_k \text{ independent of } t_{k+1}) &= \max_{t_{k+1} \in [t_k, T]} \max\left(g(t_k, \underline{\mathbf{x}}_k), G_{k+1}(\mathbf{t}^{\{k+1\}}, [\mathbf{x}^{\{k-1\}}, \underline{\mathbf{x}}_k])\right) \\ &\stackrel{(21)}{\leq} \max_{t_{k+1} \in [t_k, T]} \min_{\mathbf{x}_k \in \mathcal{U}_k(\mathbf{t}^{\{k+1\}}, \mathbf{x}^{\{k-1\}})} \max\left(g(t_k, \mathbf{x}_k), G_{k+1}(\mathbf{t}^{\{k+1\}}, \mathbf{x}^{\{k\}})\right). \end{aligned}$$

The argument then follows by induction. ■

□

Lemma 1. Consider any $1 \leq k \leq n$, any $\mathbf{t}^{\{k+1\}}$ and $\mathbf{x}^{\{k-1\}}$, and let

$$\underline{\mathbf{x}}_k \stackrel{\text{def}}{=} \min\{\mathbf{x} : \mathbf{x} \in \mathcal{U}_k(\mathbf{t}^{\{k\}}, \mathbf{x}^{\{k-1\}})\}.$$

Then,

$$\underline{\mathbf{x}}_k \in \arg \min_{\mathbf{x}_k \in \mathcal{U}_k(\mathbf{t}^{\{k\}}, \mathbf{x}^{\{k-1\}})} g(t_k, \mathbf{x}_k) \cap \arg \min_{\mathbf{x}_k \in \mathcal{U}_k(\mathbf{t}^{\{k\}}, \mathbf{x}^{\{k-1\}})} C_k(\mathbf{t}^{\{k\}}, \mathbf{x}^{\{k\}}) \cap \arg \min_{\mathbf{x}_k \in \mathcal{U}_k(\mathbf{t}^{\{k+1\}}, \mathbf{x}^{\{k-1\}})} G_{k+1}(\mathbf{t}^{\{k+1\}}, \mathbf{x}^{\{k\}}).$$

Proof. Proof of Lemma 1. Note that $\underline{\mathbf{x}}_k$ is well defined since $\mathcal{U}_k(\mathbf{t}^{\{k\}}, \mathbf{x}^{\{k-1\}})$ is a lattice, by Assumption 1(i). Since $g(t_k, \mathbf{x}_k)$ is increasing in \mathbf{x}_k for any t_k , it suffices to prove that $C_k(\mathbf{t}^{\{k\}}, \mathbf{x}^{\{k\}})$ and $G_{k+1}(\mathbf{t}^{\{k+1\}}, \mathbf{x}^{\{k\}})$ are increasing functions, for any k and any $\mathbf{t}^{\{k+1\}}$. To the latter point, we claim that it suffices to show that $J_{k+1}(\mathbf{t}^{\{k+1\}}, \mathbf{x}^{\{k+1\}})$ is increasing in $\mathbf{x}^{\{k+1\}}$; when this holds, we readily have that:

$$G_{k+1}(\mathbf{t}^{\{k+1\}}, \mathbf{x}^{\{k\}}) \stackrel{\text{def}}{=} \min_{\mathbf{x}_{k+1} \in \mathcal{U}(\mathbf{t}^{\{k+1\}}, \mathbf{x}^{\{k\}})} J_{k+1}(\mathbf{t}^{\{k+1\}}, \mathbf{x}^{\{k+1\}}) = J_{k+1}(\mathbf{t}^{\{k+1\}}, [\mathbf{x}^{\{k\}}, \underline{\mathbf{x}}_{k+1}])$$

is increasing in $\mathbf{x}^{\{k\}}$, since $\underline{\mathbf{x}}_{k+1}$ is increasing in $\mathbf{x}^{\{k\}}$, by Assumption 1(ii). And thus

$$C_k(\mathbf{t}^{\{k\}}, \mathbf{x}^{\{k\}}) \stackrel{\text{def}}{=} \max_{t_{k+1} \in [t_k, T]} G_{k+1}(\mathbf{t}^{\{k+1\}}, \mathbf{x}^{\{k\}})$$

is also increasing in $\mathbf{x}^{\{k\}}$, as a maximum of increasing functions.

To complete our proof, it thus suffices to show that $J_k(\mathbf{t}^{\{k\}}, \cdot)$ is increasing, for any k and any $\mathbf{t}^{\{k+1\}}$. We prove this by induction. For $k = n + 1$, we have that $J_{n+1}(\mathbf{t}^{\{n+1\}}, \mathbf{x}^{\{n+1\}}) \stackrel{\text{def}}{=} g(t_{n+1}, \mathbf{x}_{n+1})$, so $J_{n+1}(\mathbf{t}^{\{n+1\}}, \cdot)$ is increasing. Assume the property holds for J_{k+1} . Then, $G_{k+1}(\mathbf{t}^{\{k+1\}}, \mathbf{x}^{\{k\}})$ and $C_k(\mathbf{t}^{\{k\}}, \mathbf{x}^{\{k\}})$ are both increasing in $\mathbf{x}^{\{k\}}$, by the argument above. And since $g(t_k, \mathbf{x}_k)$ is also increasing in \mathbf{x}_k by Assumption 2, we have that

$$J_k(\mathbf{t}^{\{k\}}, \mathbf{x}^{\{k\}}) = \max\left(g(t_k, \mathbf{x}_k), C_k(\mathbf{t}^{\{k\}}, \mathbf{x}^{\{k\}})\right)$$

is also increasing in $\mathbf{x}^{\{k\}}$, completing our inductive step. ■

□

Proof. Proof of Proposition 2. According to Lemma 1, the worst-case process value at time t_k , which we denote by $\underline{\mathbf{x}}_k(\mathbf{t}^{\{n+1\}}, \underline{\mathbf{x}}^{\{k-1\}})$, can be obtained by choosing the smallest element of the corresponding uncertainty set, i.e.,

$$\underline{\mathbf{x}}_k(\mathbf{t}^{\{n+1\}}, \underline{\mathbf{x}}^{\{k-1\}}) \stackrel{\text{def}}{=} \min\left\{\boldsymbol{\xi} \in \mathbb{R}^d : \boldsymbol{\xi} \in \mathcal{U}_k(\mathbf{t}^{\{n+1\}}, \underline{\mathbf{x}}^{\{k-1\}})\right\}, \forall k \in \{1, \dots, n+1\}.$$

The result follows by induction. ■

□

Proof. Proof of Theorem 2. We claim that there always exists an optimal choice of $\mathbf{t}^{\{n+1\}}$ such that the inner maximum in (7) is reached at t_n or $t_{n+1} = T$ (i.e., for $k = n$ or $n + 1$, respectively). To see this, assume that the maximum occurs at $\bar{k} < n$, and introduce a new set of monitoring times \mathbf{z} where all times are identical except for the \bar{k} -th and the $(\bar{k} + 1)$ -th, which now equal $t_{\bar{k}-1}$ and $t_{\bar{k}}$, respectively, i.e., $\mathbf{z} = [t_0, t_1, \dots, t_{\bar{k}-1}, t_{\bar{k}-1}, t_{\bar{k}}, t_{\bar{k}+2}, \dots, t_n, t_{n+1}]$. Then,

$$V \geq \max_{k \in \{1, \dots, n+1\}} g(z_k, \underline{\mathbf{x}}_k(\mathbf{z})) \stackrel{(7)}{\geq} g(z_{\bar{k}+1}, \underline{\mathbf{x}}_{\bar{k}+1}(\mathbf{z})) = g(t_{\bar{k}}, \underline{\mathbf{x}}_{\bar{k}}(\mathbf{t}^{\{n+1\}})) = V.$$

The penultimate equality holds because $\mathcal{U}_{\bar{k}}(\mathbf{t}^{\{n+1\}}, \underline{\mathbf{x}}^{\{\bar{k}-1\}}) = \mathcal{U}_{\bar{k}+1}(\mathbf{z}, \underline{\mathbf{x}}^{\{\bar{k}\}})$, by the construction of \mathbf{z} and by the definition of \mathcal{U} in (1). Thus, under the monitoring times \mathbf{z} , the same optimum is reached at $\bar{k} + 1$. Repeating the argument inductively yields the result. ■

□

Proof. Proof of Theorem 3 For the convex case in (i), please refer to the proof of Theorem 4 for a more general result. For the concave case in (ii), consider problem (9), $\max_{\mathbf{t}^{\{n+1\}}} g(t_n, \underline{\mathbf{x}}_n(\mathbf{t}^{\{n+1\}}))$. By Theorem 4, $\underline{\mathbf{x}}_n(\mathbf{t}^{\{n+1\}}) = \underline{\mathbf{x}}_0 + \sum_{i=1}^n \ell(t_i - t_{i-1})\mathbf{1}$ when ℓ is concave. Thus, for a fixed t_n , maximizing $g(t_n, \underline{\mathbf{x}}_n(\mathbf{t}^{\{n+1\}}))$ is equivalent to maximizing $\sum_{i=1}^n \ell(t_i - t_{i-1})$, since $g(t, \mathbf{x})$ is increasing in \mathbf{x} . Due to the concavity of $\ell(\delta)$, the maximum is achieved when $t_i - t_{i-1} = t_n/n$, $i = 1, \dots, n$ (by

Jensen's inequality). Therefore, problem (9) becomes equivalent to $\max_{t_n} g(t_n, \phi_n(t_n))$. Similarly, $\max_{\mathbf{t}^{\{n+1\}}} g(t_{n+1}, \mathbf{x}_{n+1}(\mathbf{t}^{\{n+1\}})) = g(T, \phi_{n+1}(T))$. ■ \square

Proof. Proof of Theorem 4 For case (i), we first show that $\mathbf{x}_k(\mathbf{t}^{\{n+1\}}) = \mathbf{x}_0 + \ell(t_0, t_k - t_0)\mathbf{1}$. To that end, note that for any $j \in \{1, \dots, d\}$ and $1 \leq p < q \leq n$:

$$x_0^j + \ell(t_0, t_p - t_0) + \ell(t_p, t_q - t_p) \leq x_0^j + \ell(t_0, t_p - t_0) + \ell(t_0, t_q - t_p) \leq x_0^j + \ell(t_0, t_q - t_0), \quad (22)$$

where the first inequality follows since $\ell(t, \delta)$ is decreasing in t , and the second inequality follows because ℓ is superadditive in δ (since ℓ is convex in δ and $\ell(\cdot, 0) = 0$). Therefore,

$$\begin{aligned} \mathbf{x}_k(\mathbf{t}^{\{n+1\}}) &\stackrel{\text{def}}{=} \min \left\{ \boldsymbol{\xi} : \boldsymbol{\xi} \in \mathcal{U}_k(\mathbf{t}^{\{n+1\}}, [\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_{k-1}]) \right\} \\ &= \mathbf{x}_0 + \left[\max_{\{0=k_1 \leq \dots \leq k_r=k\} \in \{1, \dots, k\}} \sum_{i=1}^r \ell(t_{k_{i-1}}, t_{k_i} - t_{k_{i-1}}) \right] \mathbf{1} = \mathbf{x}_0 + \ell(t_0, t_k - t_0)\mathbf{1}. \end{aligned}$$

For case (ii), note that when $\ell(t, \delta)$ is increasing in t and concave in δ , the reverse inequalities hold in (22). Thus, nature's optimal (worst-case) response is given by:

$$\mathbf{x}_k(\mathbf{t}^{\{n+1\}}) = \mathbf{x}_0 + \left[\sum_{i=1}^k \ell(t_{i-1}, t_i - t_{i-1}) \right] \mathbf{1}.$$

Since $g(t, \mathbf{x})$ is jointly concave and increasing in \mathbf{x} , and $\ell(t, \delta)$ is jointly concave, the composition $g(t_n, \mathbf{x}_n(\mathbf{t}^{\{n+1\}}))$ is concave in $\mathbf{t}^{\{n+1\}}$, so that problem (9) requires maximizing a concave function over the convex set $0 \leq t_1 \leq \dots \leq t_n \leq T$. ■ \square

Proof. Proof of Proposition 3. Our proof relies on the following known result from lattice programming.

Lemma 2 (Topkis 1998). *If h is a convex, increasing (i.e., isotone) function, and f is supermodular and either isotone or antitone, then $h \circ f$ is supermodular.*

Note that every component of \mathbf{x}_n can be written as:

$$\begin{aligned} \mathbf{x}_n^k &= \mathbf{x}_0^k + \max_{s \in \mathcal{S}} f_s(\mathbf{t}^{\{n\}}) \\ f_s(\mathbf{t}^{\{n\}}) &\stackrel{\text{def}}{=} \sum_{i=0}^{p-1} \ell(t_{s(i)}, t_{s(i+1)} - t_{s(i)}), \end{aligned}$$

where \mathcal{S} denotes the set of all ordered subsets of $\{0, \dots, n\}$ that include 0 and n . By assumption (ii), for every $s \in \mathcal{S}$, f_s is supermodular in $\mathbf{t}^{\{n\}}$ and decreasing (i.e., antitone) in $\mathbf{t}^{\{n\}}$. Since the max function is convex and increasing (i.e., isotone), we can invoke Proposition 2 to conclude that

$\underline{\mathbf{x}}_n^k$ is supermodular in $\mathbf{t}^{\{n\}}$. Since $\underline{\mathbf{x}}_n^k$ is also decreasing in $\mathbf{t}^{\{n\}}$, and $g(t, \mathbf{x})$ is increasing and component-wise convex in \mathbf{x} , we can again invoke Proposition 2 to conclude that $g(t_n, \underline{\mathbf{x}}_n(\mathbf{t}^{\{n\}}))$ is supermodular in $\mathbf{t}^{\{n-1\}}$ for any fixed t_n . ■ □

Proof. Proof of Proposition 4. Consider any patient with starting age $A \in [33, 62]$. We first make two simple observations: (i) any transition from disease progression constitutes a worsening of the patient’s condition; and (ii) for any disease stage transition, a worst-case reward is obtained by transitioning as quickly as possible. Both observations follow in view of expression (14) for the total QALYs: transitioning always results in the same or worse QALY factors (by Table 2), and a worsening in the survival rate (by Table 3). Thus, to minimize the reward from re-transplantation (given by the product of the total QALYs from (14) and the survival rate from), it is best to transition as quickly as possible from states and thus spending more time in later states, which are “worse” in terms of both QALY factors and survival rate.

If the patient is in stage 2L, there are three possible transitions, to 3L, 2H and 3H. We argue that for any future time, a transition to 2H can always be dominated by a transition to 3L, in the sense that it leaves the patient worse off, i.e., it results in lower worst-case rewards for any possible future re-transplantation time. First, consider future times small enough so that only a transition 2L to 2H is possible (and not 2L to 2H to 3H). We argue that 2L to 3L is worse. In view of the two observations above, since the survival rate is lower in state 3L than in state 2H (by Remark 1), it suffices to show that the transition 2L-3L results in spending less time in state 2L than the transition 2L-2H. To that end, assume the patient is in state 2L at time t , and recall that the worst-case (i.e., lowest-possible) transition time $L_{ij'j'}(t, A)$ in our robust model is $\frac{-\ln(\rho)}{\lambda_{ij'j'}(t, A)}$. By using Table 4, we can check that $1/\lambda_{2L3L}(t, A) \leq 1/\lambda_{2L2H}(t, A)$ holds for all $A \leq 62.77$, which gives the desired result.

For longer future times so that 2L to 3L to 3H is also possible, note that 2L to 3L to 3H would still provide worse outcomes. To see this, we can again check that $1/\lambda_{2L3L}(t, A) + 1/\lambda_{3L3H}(t, A) \leq 1/\lambda_{2L2H}(t, A) + 1/\lambda_{2L3H}(t, A)$ for the age groups we are considering.

Similarly, we can rule out transitions from 1L to 1H and from 1L to 2H.

Result (ii) follows by replacing the variables along the worst-case progression path. ■ □

B Extensions

We explore two extensions of our model that may be relevant in practice.

B.1 Costly Monitoring & Optimal Number of Monitoring Times

Consider our base model, but assume that each monitoring incurs a fixed cost $c \geq 0$. In particular, the reward from stopping at the k -th monitoring time t_k with a state of \mathbf{x}_k becomes $g(t_k, \mathbf{x}_k) - kc$, for all $k = 1, \dots, n$. When monitoring is costly, the DM need not opt for monitoring n times; instead, the number of monitoring times becomes an implicit decision.

In this extension, it can be readily checked that Theorem 1 continues to hold, so that the worst-case optimal rewards under dynamic and static monitoring are the same. Thus, it suffices again to focus on the static monitoring problem. Let $\tilde{V}_k^c(\mathbf{t}^{\{n+1\}}, \mathbf{x}^{\{k\}})$ be the worst-case value-to-go function for the static monitoring problem at time t_k , with the first k observations made. The Bellman equations become:

$$\begin{aligned} \tilde{V}_k^c(\mathbf{t}^{\{n+1\}}, \mathbf{x}^{\{k\}}) &= \max \left(g(t_k, \mathbf{x}_k) - ck, \min_{\mathbf{x}_{k+1} \in \tilde{U}_{k+1}(\mathbf{t}^{\{n+1\}}, \mathbf{x}^{\{k\}})} \tilde{V}_{k+1}^c(\mathbf{t}^{\{n+1\}}, \mathbf{x}^{\{k+1\}}) \right), \\ \tilde{V}_{n+1}^c(\mathbf{t}^{\{n+1\}}, \mathbf{x}^{\{n+1\}}) &= g(t_{n+1}, \mathbf{x}_{n+1}) - c(n+1). \end{aligned}$$

It can be readily checked that, under a given set of monitoring times $\mathbf{t}^{\{n+1\}}$, nature's optimal (worst-case) response $\underline{\mathbf{x}}_k(\mathbf{t}^{\{n+1\}})$ in period t_k is still given by Proposition 2. Thus, the DM's problem can again be reformulated as:

$$\tilde{V}_0^c(t_0, \mathbf{x}_0) = \max_{\mathbf{t}^{\{n+1\}}} \max_{k \in \{1, \dots, n+1\}} \left[g(t_k, \underline{\mathbf{x}}_k(\mathbf{t}^{\{n+1\}})) - ck \right]. \quad (23)$$

A key difference from our earlier results lies in the DM's optimal stopping strategy. Recall that in our base model, it was worst-case optimal to either stop at the last monitoring time t_n or continue until the end of the horizon (see Theorem 2). This is no longer the case here, as an optimal policy may require stopping at an earlier time due to the monitoring cost c . Thus, the optimal k^* in (23) may be strictly smaller than n . In other words, in the above problem (23), by choosing the worst-case stopping strategy, the DM also implicitly chooses the optimal number of monitoring times, k^* . Therefore, n should only be interpreted here as an upper bound on monitoring opportunities, and k as the number of monitorings in the monitoring schedule that the DM needs to decide.

To solve (23), one can switch the order of the maximization operators. Since finding the optimal $\mathbf{t}^{\{n+1\}}$ for a fixed k requires solving problems that are structurally identical to Problem (9), our results in §3.3.1 and §3.3.2 can be directly leveraged. By iterating over k , one can then recover the optimal number of monitoring times.

Moreover, when the bounds are stationary, the problem of finding the optimal number of monitoring times is also tractable under mild conditions, as summarized next.

Proposition 5. *Under Assumption 2 and for the uncertainty set in (10) with stationary lower bounds $\ell(t_q - t_p)$,*

- (i) *if $\ell(\cdot)$ is convex, then a single monitoring time is sufficient for achieving the worst-case optimal reward;*
- (ii) *if $\ell(\cdot)$ is concave, and $g(t, \mathbf{x})$ is jointly concave, then the optimal number of monitoring times can be done by solving convex optimization problems.*

Proof. Proof. Part (i) follows directly from Theorem 3(i). For part (ii), recall from Theorem 3(ii) that finding the optimal stopping time under a fixed number of monitoring times n requires solving the problem $\max_{t \in [0, T]} g(t, \mathbf{x}_0 + n\ell(t/n)\mathbf{1})$. The function to be maximized is jointly concave in (t, n) , since $g(t, \mathbf{x})$ is jointly concave and increasing in \mathbf{x} , and the functions $n\ell(t/n)$ are jointly concave in (t, n) since ℓ is concave. Thus, one can find an optimal t and n by first maximizing a concave function over a convex feasible set (considering n continuous), and then checking the nearest integers (possibly solving two additional one dimensional convex optimization problems to determine the corresponding t). ■ □

B.2 More General Decision Process

Some of our results concerning the monitoring policy also extend to a more general decision problem, where the DM, instead of simply stopping, can modify the processes by increasing the state values (an action we refer to as “injection”) or decreasing them (“extraction”). This allows capturing several applications of interest. In chronic disease monitoring, “injections” could denote interventions that are costly or have immediate side-effects in the short run, but carry long-term benefits, while “extractions” could capture relaxing a strict treatment, leading to immediate relief but carrying potential long-term consequences. In collateralized lending, “injections” could denote the costly addition of new collateral, which improves the borrowing base, and “extractions” could denote immediate collateral liquidations, which generate cash but reduce the borrowing base.

To formalize this, consider our setup in §2, but assume that at the k -th monitoring time t_k , upon observing the state value $\mathbf{x}_k \stackrel{\text{def}}{=} \mathbf{x}(t_k)$, the DM decides an action $\mathbf{y}_k \in A(\mathbf{x}_k) \subseteq \mathbb{R}^d$, which results in an immediately updated state $\mathbf{z}_k \stackrel{\text{def}}{=} \mathbf{x}_k - \mathbf{y}_k$, and a net reward $r(t_k, \mathbf{x}_k, \mathbf{y}_k)$ accruing to the DM. When $\mathbf{y}_k \geq 0$ (< 0), the action can be thought of as extracting value from (injecting value into) the processes, in which case the corresponding net reward would typically be positive (respectively, negative). Not all actions are possible, and $A(\mathbf{x}_k)$ captures the feasible set when the initial state is \mathbf{x}_k .

Following the DM's action, the system subsequently evolves from time t_k to the next monitoring time t_{k+1} , where it takes a value of \mathbf{x}_{k+1} , chosen by nature from an uncertainty set. More precisely, for any $0 \leq k \leq r \leq n + 1$, and given a fixed choice of monitoring times $\mathbf{t}^{\{r\}}$, observations $\mathbf{x}^{\{k\}}$ and post-action states $\mathbf{z}^{\{k\}}$ up to time t_k , the set of possible future values for $[\mathbf{x}_{k+1}, \dots, \mathbf{x}_r, \mathbf{x}_{n+1}]$ is given by:

$$\tilde{\mathcal{U}}(\mathbf{t}^{\{r\}}, \mathbf{x}^{\{k\}}, \mathbf{z}^{\{k\}}) \stackrel{\text{def}}{=} \left\{ [\mathbf{x}_{k+1}, \dots, \mathbf{x}_r] \in \mathbb{R}^{d \times (r-k+1)} : \tilde{f}(t_p, t_q, \mathbf{x}_p, \mathbf{z}_p, \mathbf{x}_q), \right. \\ \left. \forall p, q \in \{0, \dots, r, n+1\}, p < q \right\}. \quad (24)$$

As before, we consider two versions of the DM's problem—*static* and *dynamic*—depending on whether the monitoring times are chosen at inception or throughout the problem horizon. The DM's objective is to determine the monitoring times $\mathbf{t}^{\{n+1\}}$ and the optimal actions $\mathbf{y}^{\{n+1\}}$ that maximize his cumulative reward up to time t_{n+1} .

Assumptions. We assume that rewards and action sets are monotonic in states.

Assumption 3. *The net reward $r(t, \mathbf{x}, \mathbf{y})$ is increasing in \mathbf{x} , and the action set $A(\mathbf{x})$ is increasing in \mathbf{x} with respect to set inclusion, i.e., $\mathbf{x}_1 \leq \mathbf{x}_2 \Rightarrow A(\mathbf{x}_1) \subseteq A(\mathbf{x}_2)$.*

Several feasible sets satisfy our requirement; for instance, $A(\mathbf{x}) = \{\mathbf{y} : 0 \leq \mathbf{y} \leq \mathbf{x}\}$. Paralleling Assumption 1, we also require the uncertainty sets to be lattices, with suitable monotonicity and dynamic consistency properties.

Assumption 4. *For any $0 \leq k < r \leq n$, and given $\mathbf{t}^{\{r\}}, \mathbf{x}^{\{k\}}$ and $\mathbf{z}^{\{k\}}$,*

(i) *(Lattice) $\tilde{\mathcal{U}}(\mathbf{t}^{\{r\}}, \mathbf{x}^{\{k\}}, \mathbf{z}^{\{k\}})$ is a lattice;*

(ii) *(Monotonicity) $\tilde{\mathcal{U}}(\mathbf{t}^{\{r\}}, \mathbf{x}^{\{k\}}, \mathbf{z}^{\{k\}})$ is increasing in $\mathbf{x}^{\{k\}}$ and $\mathbf{z}^{\{k\}}$;*

(iii) *(Dynamic Consistency) $\Pi_{:,i:j} \tilde{\mathcal{U}}(\mathbf{t}^{\{r\}}, \mathbf{x}^{\{k\}}, \mathbf{z}^{\{k\}}) = \Pi_{:,i:j} \tilde{\mathcal{U}}(\mathbf{t}^{\{r'\}}, \mathbf{x}^{\{k\}}, \mathbf{z}^{\{k\}})$, $\forall i \leq j < r - k + 1$, $\forall r \leq r' \leq n$, and $\forall t_{r'} \geq t_{r'-1} \geq \dots \geq t_r$.*

These generalized sets allow future states to depend on historical state values both immediately before and immediately after the DM's actions. As before, we can prove that dynamic consistency is guaranteed when \tilde{f} is monotonic in its second argument (details are omitted.)

Analysis. We first consider the dynamic problem. With $\tilde{J}_k(\mathbf{t}^{\{k\}}, \mathbf{x}^{\{k\}}, \mathbf{z}^{\{k-1\}})$ denoting¹⁸ the DM's

¹⁸To simplify notation, we define $\mathbf{z}^{\{-1\}} \stackrel{\text{def}}{=} \emptyset$.

value-to-go function at time t_k , the Bellman recursions become:

$$\begin{aligned}\tilde{J}_k(\mathbf{t}^{\{k\}}, \mathbf{x}^{\{k\}}, \mathbf{z}^{\{k-1\}}) &= \max_{\mathbf{y}_k \in A(\mathbf{x}_k)} \left[r(t_k, \mathbf{x}_k, \mathbf{y}_k) + \right. \\ &\quad \left. \max_{t_{k+1} \in [t_k, T]} \min_{\mathbf{x}_{k+1} \in \tilde{U}(\mathbf{t}^{\{k+1\}}, \mathbf{x}^{\{k\}}, \mathbf{z}^{\{k\}})} \tilde{J}_{k+1}(\mathbf{t}^{\{k+1\}}, \mathbf{x}^{\{k+1\}}, \mathbf{z}^{\{k\}}) \right], \\ \tilde{J}_{n+1}(\mathbf{t}^{\{n+1\}}, \mathbf{x}^{\{n+1\}}, \mathbf{z}^{\{n\}}) &= \max_{\mathbf{y}_{n+1} \in A(\mathbf{x}_{n+1})} r(t_{n+1}, \mathbf{x}_{n+1}, \mathbf{y}_{n+1}).\end{aligned}$$

Let $\tilde{J}_0 \stackrel{\text{def}}{=} \tilde{J}_0(t_0, \mathbf{x}_0)$.

In the static problem, the DM chooses $\mathbf{t}^{\{n+1\}}$ at time t_0 . With $\tilde{V}_k(\mathbf{t}^{\{k\}}, \mathbf{x}^{\{k\}}, \mathbf{z}^{\{k-1\}})$ denoting the value-to-go function at time t_k , the Bellman recursions become:

$$\begin{aligned}\tilde{V}_k(\mathbf{t}^{\{n+1\}}, \mathbf{x}^{\{k\}}, \mathbf{z}^{\{k-1\}}) &= \max_{\mathbf{y}_k \in A(\mathbf{x}_k)} \left[r(t_k, \mathbf{x}_k, \mathbf{y}_k) + \min_{\mathbf{x}_{k+1} \in \tilde{U}(\mathbf{t}^{\{k+1\}}, \mathbf{x}^{\{k\}}, \mathbf{z}^{\{k\}})} \tilde{V}_{k+1}(\mathbf{t}^{\{n+1\}}, \mathbf{x}^{\{k+1\}}, \mathbf{z}^{\{k\}}) \right], \\ \tilde{V}_{n+1}(\mathbf{t}^{\{n+1\}}, \mathbf{x}^{\{n+1\}}, \mathbf{z}^{\{n\}}) &= \max_{\mathbf{y}_{n+1} \in A(\mathbf{x}_{n+1})} r(t_{n+1}, \mathbf{x}_{n+1}, \mathbf{y}_{n+1}),\end{aligned}$$

and the optimal choice of monitoring times yields a value of $\tilde{V}_0 \stackrel{\text{def}}{=} \max_{\mathbf{t}^{\{n+1\}}} \tilde{V}_0(\mathbf{t}^{\{n+1\}}, \mathbf{x}^{\{0\}})$.

In this context, we can confirm that an analogous result to Theorem 1 holds, and the dynamic problem yields the same worst-case optimal reward as the static problem.

Theorem 5. *Under Assumption 3 and Assumption 4, $\tilde{J}_0 = \tilde{V}_0$.*

Proof. Proof. The Bellman recursion under dynamic monitoring can be written:

$$\begin{aligned}\tilde{J}_k(\mathbf{t}^{\{k\}}, \mathbf{x}^{\{k\}}, \mathbf{z}^{\{k-1\}}) &= \max_{\mathbf{y}_k \in A(\mathbf{x}_k)} \max_{t_{k+1} \in [t_k, T]} \left[r(t_k, \mathbf{x}_k, \mathbf{y}_k) + \tilde{G}_k(\mathbf{t}^{\{k+1\}}, \mathbf{x}^{\{k\}}, \mathbf{z}^{\{k\}}) \right], \quad \text{where} \\ \tilde{G}_k(\mathbf{t}^{\{k+1\}}, \mathbf{x}^{\{k\}}, \mathbf{z}^{\{k\}}) &\stackrel{\text{def}}{=} \min_{\mathbf{x}_{k+1} \in \tilde{U}(\mathbf{t}^{\{k+1\}}, \mathbf{x}^{\{k\}}, \mathbf{z}^{\{k\}})} \tilde{J}_{k+1}(\mathbf{t}^{\{k+1\}}, \mathbf{x}^{\{k+1\}}, \mathbf{z}^{\{k\}}), \quad \forall k \in \{1, \dots, n\}.\end{aligned}$$

First, using induction, we prove that \tilde{J}_k and \tilde{G}_k are increasing in all arguments except time. By Assumption 3, this is true¹⁹ for $\tilde{J}_{n+1}(\mathbf{t}^{\{n+1\}}, \mathbf{x}^{\{n+1\}}, \mathbf{z}^{\{n\}})$. Assuming this is true at $k+1$, and using Assumption 4(i,ii), note that:

$$\begin{aligned}\operatorname{argmin}_{\mathbf{x}_{k+1} \in \tilde{U}(\mathbf{t}^{\{k+1\}}, \mathbf{x}^{\{k\}}, \mathbf{z}^{\{k\}})} \tilde{J}_{k+1}(\mathbf{t}^{\{k+1\}}, \mathbf{x}^{\{k+1\}}, \mathbf{z}^{\{k\}}) &= \min_{\mathbf{x}_{k+1} \in \tilde{U}(\mathbf{t}^{\{k+1\}}, \mathbf{x}^{\{k\}}, \mathbf{z}^{\{k\}})} \mathbf{x}_{k+1} \\ &\stackrel{\text{def}}{=} \underline{\mathbf{x}}_{k+1}(\mathbf{t}^{\{k+1\}}, \mathbf{x}^{\{k\}}, \mathbf{z}^{\{k\}}),\end{aligned}$$

¹⁹To see why this follows, consider $f(x) \stackrel{\text{def}}{=} \max_{y \in A(x)} g(x, y)$ where $g(\cdot, y)$ is increasing for any y , and let $y^*(x)$ denote a maximizer in the problem. Then, for $x_1 \leq x_2$, we have $f(x_1) = g(x_1, y^*(x_1)) \leq g(x_2, y^*(x_1)) \leq \max_{y \in A(x_2)} g(x_2, y) = f(x_2)$, where the inequality in the second step relies on $y^*(x_1) \in A(x_2)$, which is guaranteed by Assumption 3.

and $\underline{\mathbf{x}}_{k+1}(\mathbf{t}^{\{k+1\}}, \mathbf{x}^{\{k\}}, \mathbf{z}^{\{k\}})$ is increasing in $\mathbf{x}^{\{k\}}$ and $\mathbf{z}^{\{k\}}$. Therefore,

$$\tilde{G}_k(\mathbf{t}^{\{k+1\}}, \mathbf{x}^{\{k\}}, \mathbf{z}^{\{k\}}) = \tilde{J}_{k+1}\left(\mathbf{t}^{\{k+1\}}, [\mathbf{x}^{\{k\}}, \underline{\mathbf{x}}_{k+1}(\mathbf{t}^{\{k+1\}}, \mathbf{x}^{\{k\}}, \mathbf{z}^{\{k\}})], \mathbf{z}^{\{k\}}\right)$$

is increasing in $(\mathbf{x}^{\{k\}}, \mathbf{z}^{\{k\}})$. But then, note that the maximand in the problem:

$$\tilde{J}_k(\mathbf{t}^{\{k\}}, \mathbf{x}^{\{k\}}, \mathbf{z}^{\{k-1\}}) = \max_{\mathbf{y}_k \in A(\mathbf{x}_k)} \max_{t_{k+1} \in [t_k, T]} \left[r(t_k, \mathbf{x}_k, \mathbf{y}_k) + \tilde{G}_k(\mathbf{t}^{\{k+1\}}, \mathbf{x}^{\{k\}}, [\mathbf{z}^{\{k-1\}}, \mathbf{x}_k - \mathbf{y}_k]) \right].$$

is increasing in $(\mathbf{x}^{\{k\}}, \mathbf{z}^{\{k-1\}})$, for any fixed value of \mathbf{y}_k and t_{k+1} . And since the action set $A(\mathbf{x}_k)$ is increasing in \mathbf{x}_k with respect to set inclusion by Assumption 3, this implies that \tilde{J}_k is increasing in $(\mathbf{x}^{\{k\}}, \mathbf{z}^{\{k-1\}})$, which completes our induction.

Using these monotonicity properties, we then obtain:

$$\tilde{G}_{k-1}(\mathbf{t}^{\{k\}}, \mathbf{x}^{\{k-1\}}, \mathbf{z}^{\{k-1\}}) \tag{25}$$

$$\begin{aligned} &= \min_{\mathbf{x}_k \in \tilde{\mathcal{U}}(\mathbf{t}^{\{k\}}, \mathbf{x}^{\{k-1\}}, \mathbf{z}^{\{k-1\}})} \max_{\mathbf{y}_k \in A(\mathbf{x}_k)} \max_{t_{k+1} \in [t_k, T]} \left[r(t_k, \mathbf{x}_k, \mathbf{y}_k) + \tilde{G}_k(\mathbf{t}^{\{k+1\}}, [\mathbf{x}^{\{k-1\}}, \mathbf{x}_k], \mathbf{z}^{\{k\}}) \right] \\ &= \max_{t_{k+1} \in [t_k, T]} \max_{\mathbf{y}_k \in A(\mathbf{x}_k)} \left[r(t_k, \underline{\mathbf{x}}_k, \mathbf{y}_k) + \tilde{G}_k(\mathbf{t}^{\{k+1\}}, [\mathbf{x}^{\{k-1\}}, \underline{\mathbf{x}}_k], \mathbf{z}^{\{k\}}) \right] \\ &= \max_{t_{k+1} \in [t_k, T]} \min_{\mathbf{x}_k \in \tilde{\mathcal{U}}(\mathbf{t}^{\{k+1\}}, \mathbf{x}^{\{k-1\}}, \mathbf{z}^{\{k-1\}})} \max_{\mathbf{y}_k \in A(\mathbf{x}_k)} \left[r(t_k, \mathbf{x}_k, \mathbf{y}_k) + \tilde{G}_k(\mathbf{t}^{\{k+1\}}, [\mathbf{x}^{\{k-1\}}, \mathbf{x}_k], \mathbf{z}^{\{k\}}) \right]. \end{aligned} \tag{26}$$

The second equality follows from the monotonicity of r and \tilde{G}_k in \mathbf{x}_k ; the last equality follows from the same monotonicity and the dynamic consistency Assumption 4(iii), which ensures that $\tilde{\mathcal{U}}(\mathbf{t}^{\{k+1\}}, \mathbf{x}^{\{k-1\}}, \mathbf{z}^{\{k-1\}}) = \tilde{\mathcal{U}}(\mathbf{t}^{\{k\}}, \mathbf{x}^{\{k-1\}}, \mathbf{z}^{\{k-1\}})$, so that the nature's worst-case response $\underline{\mathbf{x}}_k$ is independent of the choice t_{k+1} . Therefore, we can interchange the order of $\max_{t_{k+1}}$ and $\min_{\mathbf{x}_k \in \tilde{\mathcal{U}}_k}$. Repeating the argument inductively, we obtain $\tilde{J}_0 = \tilde{V}_0$. ■ □

This result again allows reconstructing the DM's optimal dynamic monitoring policy by (re-solving) static versions of the monitoring problem. In fact, a further simplification is also possible here, as summarized in our next result.

Proposition 6. *Consider the static monitoring problem. The DM can make all the injection decisions at time t_0 and recover the same worst-case reward, i.e.,*

$$\tilde{V}_0 = \max_{\mathbf{t}^{\{n+1\}}} \max_{\mathbf{y}^{\{n+1\}} \in \mathbb{R}^{d \times (n+1)}} \min_{\substack{\mathbf{x}^{\{n+1\}} : \forall k \in \{1, \dots, n+1\}, \\ \mathbf{x}_k \in \tilde{\mathcal{U}}(\mathbf{t}^{\{n+1\}}, \mathbf{x}^{\{k-1\}}, \mathbf{z}^{\{k-1\}})}} \sum_{k=0}^N r(t_k, \mathbf{x}_k, \mathbf{y}_k).$$

Proof. Proof. Running through the same arguments as in the proof of Theorem 5, let $\mathbf{y}_k^*(\mathbf{t}^{\{k+1\}}, \mathbf{x}^{\{k\}}, \mathbf{z}^{\{k\}})$ denote an optimal policy for the DM in (26). It can be checked that the operators $\min_{\mathbf{x}_k \in \tilde{\mathcal{U}}(\mathbf{t}^{\{k+1\}}, \mathbf{x}^{\{k-1\}}, \mathbf{z}^{\{k-1\}})} \max_{\mathbf{y}_k}$

in (26) can be interchanged under a choice $\mathbf{y}_k^*(\mathbf{t}^{\{k+1\}}, \underline{\mathbf{x}}^{\{k\}}, \mathbf{z}^{\{k-1\}})$, since this action remains feasible for any \mathbf{x}_k by Assumption 3, and nature’s worst-case response under knowledge of this action remains $\underline{\mathbf{x}}^{\{k\}}$. Repeating the argument by induction then yields the result. ■ □

In view of Proposition 6, for purposes of recovering the worst-case reward, the DM can restrict attention to static policies for *both* monitoring *and* extraction; this simplifies the problem, and allows reconstructing a dynamic policy by repeatedly finding static policies.

C High acute rejections for CAV

In our data, the distribution of number of acute rejections is highly skewed. It ranges from 0 to 12, while 75% were less than or equal to 3. So in our analysis, we define a binary variable, indicating high or low number of acute rejections. The threshold for high acute rejections are chosen so as to maximize the statistical power of our analysis. In particular, we look at the statistical significance of the binary variable in the regression model for reward function (Table 3).

D Tables

QOL_1	QOL_2	QOL_3	QOL_{re}
0.8583	0.7138	0.5774	0.6456

Table 2: Quality of Life Factors

	Estimate	Standard Error	Significance
Intercept	1.06412	0.02503	***
Starting age (A)	-0.00134	0.00048	**
CAV stage ($i \in \{1, 2, 3\}$)	-0.06510	0.00835	***
Acute rejections ($j \in \{L, H\}$)	-0.03503	0.01412	*

Table 3: Regression Model for Survival Probability

$(i, j) \rightarrow (i', j')$	$\beta_0^{ijj'j'}$	$\beta_1^{ijj'j'}$	$\beta_2^{ijj'j'}$
1L \rightarrow 2L	20.48 (1.10)	-0.22 (0.02)	-0.69 (0.05)
1H \rightarrow 3H	17.57 (3.98)	-0.086 (0.07)	-0.42 (0.46)
1L \rightarrow 3L	46.02 (3.60)	-0.53 (0.07)	-1.15 (0.22)
2L \rightarrow 3L	9.57 (1.53)	-0.10 (0.03)	-0.13 (0.1)
1L \rightarrow 1H	7.39 (2.48)	0.299 (0.07)	29.109 (3.62)
2L \rightarrow 2H	2126.97 (2377.73)	-33.84 (38.99)	54.16 (175.57)
1L \rightarrow 2H	39.451 (14.53)	0.31 (0.34)	107.34 (22.45)
2L \rightarrow 3H	3428.443 (5392.32)	-54.245 (87.69)	17.031 (186.05)
1L \rightarrow 3H	73.70 (33.43)	0.38 (0.78)	168.36 (45.09)
2H \rightarrow 3H	2.02 (0.95)	0.033 (0.02)	-0.219 (0.15)
1H \rightarrow 2H	9.59 (1.38)	-0.073 (0.02)	-0.25 (1.15)
3L \rightarrow 3H	9.96 (1.04 $\cdot 10^{-11}$)	-0.05 (2.41 $\cdot 10^{-11}$)	-1.07 (2.03 $\cdot 10^{-11}$)

Table 4: Coefficients for mean time spent in each state. (Standard errors are reported in parenthesis.)

E IP Formulation for CAV Dynamic Monitoring Policy

Problem (20) in §4.3 can be reformulated as an IP as follows. We start by defining binary variables \mathbf{y} as follows:

$$y_k^{ij} \stackrel{\text{def}}{=} \begin{cases} 1, & \text{if } \bar{x}^{ij}(t_k) > 0, \\ 0, & \text{otherwise} \end{cases}, \quad \forall i \in \{1, 2, 3\}, j \in \{L, H\}, k \in \{0, \dots, N\}. \quad (27)$$

Similarly, consider the binary variable u_n such that $u_n = 1$ if and only if $t_n > b = 5.060$. Then, we define multiplicative variables \mathbf{z} , \mathbf{w} and \mathbf{r} as

$$z_n^{ijj'j'} \stackrel{\text{def}}{=} y_n^{ij} \bar{x}^{i'j'}(t_n), \quad w_n^{ij} \stackrel{\text{def}}{=} y_n^{ij} t, \quad r_n^{ij} \stackrel{\text{def}}{=} y_n^{ij} (t_n - b)^+, \quad (28)$$

Policy	Min	25% Quantile	Median	75% Quantile	Max	Mean
ISHLT	3.2345	7.1207	8.0081	9.3416	10.2690	7.9976
Uniform RO	5.9200	7.3592	7.9155	8.2316	8.2316	7.7749
Static RO	6.0596	7.3868	8.2096	8.9909	8.9909	8.1411
Dynamic RO	6.6691	7.6948	8.3207	8.9909	9.8833	8.3409

Table 5: Rewards under different monitoring policies

($T = 10$ years, initial age = 50, number of monitoring = 9, $\rho = 90\%$, number of iterations = 10^3).

for all $i, i' \in \{1, 2, 3\}, j, j' \in \{L, H\}$. Then, for sufficiently large $M > 0$ and sufficiently small $\epsilon > 0$, maximizing $g(t_n, \bar{x}^{2L}(t_n), \bar{x}^{3L}(t_n), \bar{x}^{3H}(t_n))$ in Problem (20) can be reformulated as follows:

$$\max_{\substack{\bar{x}^{2L}, \bar{x}^{3L}, \bar{x}^{3H}, t, \\ s_n, u_n, \mathbf{y}, \mathbf{z}, \mathbf{w}, \mathbf{r}}} (1.5269 t_n - 0.1445 \bar{x}_n^{2L} - 0.1364 \bar{x}_n^{3L} + 1.3968 - 1.1445 s_n)(0.9990 - 0.0013 \text{ age})$$

$$- 0.0994 w_n^{2L} + 0.0094 z_n^{2L2L} + 0.0089 z_n^{2L3L} - 0.0909 + 0.0745 r_n^{2L}$$

$$- 0.0994 w_n^{3L} + 0.0094 z_n^{3L2L} + 0.0089 z_n^{3L3L} - 0.0909 + 0.0745 r_n^{3L}$$

$$- 0.0534 w_n^{3H} + 0.0051 z_n^{3H2L} + 0.0048 z_n^{3H3L} - 0.0489 + 0.0401 r_n^{3H}$$

$$\text{subject to } \bar{x}_0^{ij} = 0, \quad \bar{x}_N^{ij} \leq T, \quad (i, j) \in \{2L, 3L, 3H\}, \quad (29a)$$

$$t_k \leq t_{k+1}, \quad \bar{x}_k^{ij} \leq \bar{x}_{k+1}^{ij}, \quad (i, j) \in \{2L, 3L, 3H\}, \quad k \in \{0, 1, \dots, n+1\} \quad (29b)$$

$$\bar{x}_k^{ij} \leq M y_k^{ij}, \quad \bar{x}_k^{ij} \geq \epsilon y_k^{ij}, \quad (i, j) \in \{2L, 3L, 3H\}, \quad k \in \{0, 1, \dots, n+1\} \quad (29c)$$

$$t_n \geq b u_n, \quad t_n \leq b + M u_n \quad (29d)$$

$$s_n \geq -M u_n, \quad s_n \geq t_n - b - M(1 - u_n) \quad (29e)$$

$$s_n \leq M u_n, \quad s_n \leq t_n - b + M(1 - u_n) \quad (29f)$$

$$z_n^{ij i' j'} \leq \bar{x}_n^{i' j'} + M(1 - y_n^{ij}), \quad z_n^{ij i' j'} \geq \bar{x}_n^{i' j'} - M(1 - y_n^{ij}), \quad (29g)$$

$$z_n^{ij i' j'} \leq M y_n^{ij}, \quad z_n^{ij i' j'} \geq -M y_n^{ij}, \quad (i, j), (i' j') \in \{2L, 3L, 3H\} \quad (29h)$$

$$w_n^{ij} \leq t_n + M(1 - y_n^{ij}), \quad w_n^{ij} \geq t_n - M(1 - y_n^{ij}), \quad (29i)$$

$$w_n^{ij} \leq M y_k^{ij}, \quad w_n^{ij} \geq -M y_n^{ij}, \quad (29j)$$

$$r_n^{ij} \leq s_n + M(1 - y_n^{ij}), \quad r_n^{ij} \geq s_n - M(1 - y_n^{ij}), \quad (29k)$$

$$r_n^{ij} \leq M y_k^{ij}, \quad r_n^{ij} \geq -M y_n^{ij}, \quad (i, j) \in \{2L, 3L, 3H\}, \quad (29l)$$

$$\bar{x}_{k+1}^{2L} \geq t_{k+1} - t_k + \ln(\rho)(c^{1L2L} + \beta_2^{1L2L} t_k) - M y_k^{2L}, \quad (29m)$$

$$\bar{x}_{k+1}^{2L} \geq \bar{x}_k^{2L} + t_{k+1} - t_k - M + M y_k^{2L}, \quad (29n)$$

$$\bar{x}_{k+1}^{3L} \geq t_{k+1} - t_k + \ln(\rho)(c^{1L3L} + \beta_2^{1L3L} t_k) - M y_k^{2L} - M y_k^{3L}, \quad (29o)$$

$$\bar{x}_{k+1}^{3L} \geq t_{k+1} - t_k + \ln(\rho)(c^{2L3L} + \beta_2^{2L3L} t_k) - M + M y_k^{2L} - M y_k^{3L}, \quad (29p)$$

$$\bar{x}_{k+1}^{3L} \geq \bar{x}_k^{3L} + t_{k+1} - t_k - 2M + M y_k^{2L} + M y_k^{3L}, \quad (29q)$$

$$\bar{x}_{k+1}^{3H} \geq t_{k+1} - t_k + \ln(\rho)(c^{1L3H} + \beta_2^{1L3H} t_k) - M y_k^{2L} - M y_k^{3L} - M y_k^{3H}, \quad (29r)$$

where $c^{ijj'j'} = \beta_0^{ijj'j'} + \beta_1^{ijj'j'} age$. The inequalities in (29c) define the binary variables \mathbf{y} , the inequalities in (30d) define u_n , and the inequalities in (29e)-(29f) define s_n as $(t_n - b)^+$. Then, (29g)-(29h) and (29i)-(29l) define multiplicative variables \mathbf{w} and \mathbf{r} respectively. The inequalities (29m)-(30n), (29o)-(30q) and (29r)-(29u) correspond to (19a), (19b) and (19c), respectively. The problem of maximizing the term $g(t_{n+1}, \bar{x}^{2L}(t_{n+1}), \bar{x}^{3L}(t_{n+1}), \bar{x}^{3H}(t_{n+1}))$ in (20) can similarly be formulated as an IP.

When dynamically re-solving static problems, note that a patient could also be diagnosed as 1H, 2H or 3H. In such cases, the disease would progress from 1L to 1H, to 2H and to 3H. Then, we can follow a similar modeling approach and formulate the static monitoring problem as an IP

in terms of $\bar{x}^{1H}, \bar{x}^{2H}, \bar{x}^{3H}$ as follows.

$$\max_{\substack{\bar{x}^{1H}, \bar{x}^{2H}, \bar{x}^{3H}, t, \\ s_n, u_n, \mathbf{y}, \mathbf{z}, \mathbf{w}, \mathbf{r}}} (1.5269 t_n - 0.1445 \bar{x}_n^{2H} - 0.1364 \bar{x}_n^{3H} + 1.3968 - 1.1445 s_n)(0.9990 - 0.0013 age)$$

$$\begin{aligned} & - 0.0994 w_n^{2H} + 0.0094 z_n^{2H2H} + 0.0089 z_n^{2H3H} - 0.0909 + 0.0745 r_n^{2H} \\ & - 0.0994 w_n^{3H} + 0.0094 z_n^{3H2H} + 0.0089 z_n^{3H3H} - 0.0909 + 0.0745 r_n^{3H} \\ & - 0.0534 w_n^{1H} + 0.0051 z_n^{1H2H} + 0.0048 z_n^{1H3H} - 0.0489 + 0.0401 r_n^{1H} \end{aligned}$$

$$\text{subject to } \bar{x}_0^{ij} = 0, \quad \bar{x}_N^{ij} \leq T, \quad (i, j) \in \{1H, 2H, 3H\}, \quad (30a)$$

$$t_k \leq t_{k+1}, \quad \bar{x}_k^{ij} \leq \bar{x}_{k+1}^{ij}, \quad (i, j) \in \{1H, 2H, 3H\}, \quad k \in \{0, 1, \dots, n+1\} \quad (30b)$$

$$\bar{x}_k^{ij} \leq M y_k^{ij}, \quad \bar{x}_k^{ij} \geq \epsilon y_k^{ij}, \quad (i, j) \in \{1H, 2H, 3H\}, \quad k \in \{0, 1, \dots, n+1\} \quad (30c)$$

$$t_n \geq b u_n, \quad t_n \leq b + M u_n \quad (30d)$$

$$s_n \geq -M u_n, \quad s_n \geq t_n - b - M(1 - u_n) \quad (30e)$$

$$s_n \leq M u_n, \quad s_n \leq t_n - b + M(1 - u_n) \quad (30f)$$

$$z_n^{ij i' j'} \leq \bar{x}_n^{i' j'} + M(1 - y_n^{ij}), \quad z_n^{ij i' j'} \geq \bar{x}_n^{i' j'} - M(1 - y_n^{ij}), \quad (30g)$$

$$z_n^{ij i' j'} \leq M y_n^{ij}, \quad z_n^{ij i' j'} \geq -M y_n^{ij}, \quad (i, j), (i' j') \in \{1H, 2H, 3H\} \quad (30h)$$

$$w_n^{ij} \leq t_n + M(1 - y_n^{ij}), \quad w_n^{ij} \geq t_n - M(1 - y_n^{ij}), \quad (30i)$$

$$w_n^{ij} \leq M y_k^{ij}, \quad w_n^{ij} \geq -M y_k^{ij}, \quad (30j)$$

$$r_n^{ij} \leq s_n + M(1 - y_n^{ij}), \quad r_n^{ij} \geq s_n - M(1 - y_n^{ij}), \quad (30k)$$

$$r_n^{ij} \leq M y_k^{ij}, \quad r_n^{ij} \geq -M y_k^{ij}, \quad (i, j) \in \{1H, 2H, 3H\}, \quad (30l)$$

$$\bar{x}_{k+1}^{1H} \geq t_{k+1} - t_k + \ln(\rho)(c^{1L1H} + \beta_2^{1L1H} t_k) - M y_k^{1H}, \quad (30m)$$

$$\bar{x}_{k+1}^{1H} \geq \bar{x}_k^{1H} + t_{k+1} - t_k - M + M y_k^{1H}, \quad (30n)$$

$$\bar{x}_{k+1}^{2H} \geq t_{k+1} - t_k + \ln(\rho)(c^{1L2H} + \beta_2^{1L2H} t_k) - M y_k^{1H} - M y_k^{2H}, \quad (30o)$$

$$\bar{x}_{k+1}^{2H} \geq t_{k+1} - t_k + \ln(\rho)(c^{1H2H} + \beta_2^{1H2H} t_k) - M + M y_k^{1H} - M y_k^{2H}, \quad (30p)$$

$$\bar{x}_{k+1}^{2H} \geq \bar{x}_k^{2H} + t_{k+1} - t_k - 2M + M y_k^{1H} + M y_k^{2H}, \quad (30q)$$

$$\bar{x}_{k+1}^{3H} \geq t_{k+1} - t_k + \ln(\rho)(c^{1L3H} + \beta_2^{1L3H} t_k) - M y_k^{1H} - M y_k^{2H} - M y_k^{3H}, \quad (30r)$$

$$\bar{x}_{k+1}^{3H} \geq t_{k+1} - t_k + \ln(\rho)(c^{1H3H} + \beta_2^{1H3H} t_k) - M + M y_k^{1H} - M y_k^{2H} - M y_k^{3H}, \quad (30s)$$

$$\bar{x}_{k+1}^{3H} \geq t_{k+1} - t_k + \ln(\rho)(c^{2H3H} + \beta_2^{2H3H} t_k) - 2M + M y_k^{1H} + M y_k^{2H} - M y_k^{3H}, \quad (30t)$$

$$\bar{x}_{k+1}^{3H} \geq \bar{x}_k^{3H} + t_{k+1} - t_k - 3M + M y_k^{1H} + M y_k^{2H} + M y_k^{3H}, \quad k \in \{0, 1, \dots, n\}. \quad (30u)$$