Management Science

Monitoring with Limited Information

Dan Andrei Iancu, Nikolaos Trichakis, Do Young Yoon

Monitoring with Limited Information

Dan Andrei Iancu,[a,b] **Nikolaos Trichakis,**[c] **Do Young Yoon**[d]

[a] Stanford Graduate School of Business, Stanford University, Stanford, California 94305; [b] INSEAD, Fontainebleau 77300, France; [c] Operations Research Center and Sloan School of Management, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139; [d] UBER Freight, San Francisco, California 94103

Contact: daniancu@stanford.edu, https://orcid.org/0000-0001-9890-0928 (DAI); ntrichakis@mit.edu, https://orcid.org/0000-0002-8324-9148 (NT); doyoung@stanford.edu (DYY)

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 these types 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 equidistant 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 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 healthcare 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, whereas performing it too late comes with the risk of CAV progression and health deterioration. Monitoring requires invasive and potentially expensive procedures but 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 are 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 Section 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 asset 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 (U.S. Department of the Treasury 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 multidimensional 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 or her 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 resolving 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 (CLT) 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 Zhou and Doyle 1998, Dullerud and Paganini 2005, 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 nonconvex 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.

Also relevant are Nohadani and Roy (2017) (N&R), who deal with a control problem and a monitoring problem in isolation. For the former, N&R formulate a two-stage robust linear optimization problem under a single, exogenously given observation time; within this model, they also use time-dependent uncertainty sets that capture increased uncertainty as information ages with time. For the latter, N&R consider a monitoring problem that is devoid of any control problem and without a model for the underlying state dynamics that are monitored. Rather, N&R present a model of aggregate information degradation in which monitoring at a given time yields deterministically the same information in every instance it occurs. Consequently, in the monitoring problem of N&R, there is no value in dynamically updating monitoring schedules, and all monitoring decisions are static. In contrast, we consider a joint problem of controlling a system (deciding when to stop, and a more general decision problem in Online Appendix B) and monitoring it, with multiple monitoring opportunities that can be dynamically adjusted. Notably, in our model, monitoring policies are determined endogenously, jointly with the optimal control policies, which gives rise to a different and significantly more complex problem. Furthermore, compared with the monitoring problem of N&R, we explicitly model a system whose state fluctuates. Importantly, our model captures explicitly how (i) monitoring reveals the state of the system, (ii) the decision maker updates the information depending on observations, and (iii) subsequent control and monitoring decisions are adapted to the information revealed. That is, we consider dynamic monitoring policies and dynamic information updating.

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; that is, the uncertainty is "constraint-wise." Marandi and den Hertog (2018) 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 multiperiod 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 et al. 2010, Bayraktar and Yao 2014, 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 sharper 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 (see also Iancu et al. 2013 and Delage and Iancu 2015 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 predetermined 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 and Shreve 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, Rajan and Winton 1995, Leland and Toft 1996, Gorton and Winton 2003, 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, for example, $x \in \mathbb{R}^n$. For a set of vectors $x_0, x_1, \ldots, x_k$ in $\mathbb{R}^d$, we use $x^{\{k\}} \overset{\text{def}}{=} [x_0, x_1, \ldots, x_k] \in \mathbb{R}^{d \times (k+1)}$ 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

$$\Pi_{i_1:i_2,:}\mathcal{U} = \Big\{ Y \in \mathbb{R}^{(i_2-i_1+1)\times k} \ : \ \exists X \in \mathbb{R}^{(i_1-1)\times k},$$
$$Z \in \mathbb{R}^{(d-i_2)\times k} \ : \ [X; Y; Z] \in \mathcal{U} \Big\},$$

$$\Pi_{:,j_1:j_2}\mathcal{U} = \Big\{ Y \in \mathbb{R}^{d\times(j_2-j_1+1)} \ : \ \exists X \in \mathbb{R}^{d\times(j_1-1)},$$
$$Z \in \mathbb{R}^{d\times(k-j_2)} \ : \ [X, Y, Z] \in \mathcal{U} \Big\},$$

as the projections along a subset of rows and a subset of columns, respectively. To simplify notation, we also use $\mathcal{U}_j \overset{\text{def}}{=} \Pi_{:,j:j}\mathcal{U}$ 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, that is, $x \geq y$ for $x, y \in \mathbb{R}^d$ 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 then applying the component-wise order. For a set of vectors in $\mathbb{R}^d$, *increasing* and *decreasing* are understood in the usual induced set order (Topkis 1998),[1] 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 observes the state of the system, updates his or her 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, then he or she collects a reward that depends on the stopping time and the state. If the DM decides to continue until the end, then he or she 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 $x(t) \in \mathbb{R}^d$ be a vector whose components are the system's processes, for all t in $[0, T]$. We refer to $x(t)$ as the state of the system at time t.

The DM has knowledge of the initial state, that is, $x(0)$, 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 $t_1, \ldots, t_n$ denote these n monitoring times with $0 \leq t_1 \leq \cdots \leq t_n \leq T$. To ease notation, we also let $t_0 \overset{\text{def}}{=} 0$ and $t_{n+1} \overset{\text{def}}{=} T$.

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

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 kth monitoring time t_k. At that time, the DM has made k observations at monitoring times $t_1, \ldots, t_k$, in addition to the initial observation at t_0, resulting in an observation matrix $x^{\{k\}} \overset{\text{def}}{=} [x_0, x_1, \ldots, x_k]$. Suppose that the DM is considering some future monitoring times $t_{k+1}, \ldots, t_r$ satisfying $t_k \leq t_{k+1} \leq \cdots \leq t_r \leq t_{n+1}$, for $r > k$. Then, in view

Figure 1. Illustration of How Observed State Values Constrain Future Values for $d = 1$

(a)

$\{x(t) \in \mathbf{R} : f(t_p, t, x_p, x(t)) \leq 0\}$
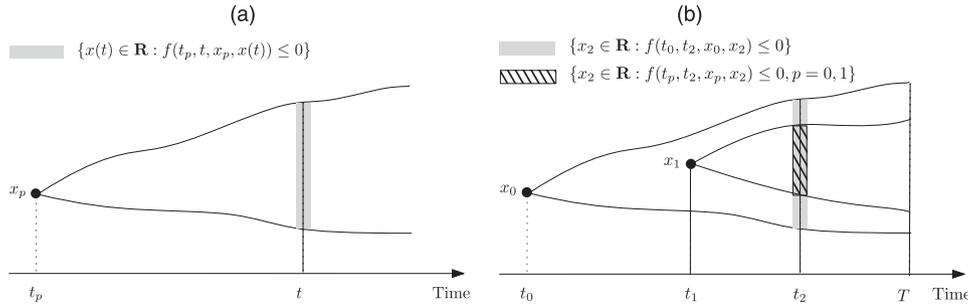
Consider the case $d = 1$. Each observation x_p imposes a restriction on the future process values $x(t)$ for all $t > t_p$, which is identified by $f(t_p, t, x_p, x(t))$.

(b)

$\{x_2 \in \mathbf{R} : f(t_0, t_2, x_0, x_2) \leq 0\}$
$\{x_2 \in \mathbf{R} : f(t_p, t_2, x_p, x_2) \leq 0, p = 0, 1\}$

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

of observations $\boldsymbol{x}^{\{k\}}$, the process values at $t_{k+1}, \ldots, t_r$, together with the terminal process value at t_{n+1}, lie in the following uncertainty set:

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

where $\boldsymbol{t}^{\{r\}} \overset{\text{def}}{=} [t_0, t_1, \ldots, t_r]$ 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 1(b), 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 initial information on the future values $x_1, x_2,$ and $x_3 = x_{n+1}$, represented by $\mathcal{U}(\boldsymbol{t}^{\{2\}}, \boldsymbol{x}^{\{0\}}) = \{[x_1, x_2, x_3] \in \mathbb{R}^3 : f(t_0, t_q, x_0, x_q) \leq 0, q = 1, 2, 3\}$. Then, with an observation x_1 at t_1, the updated set $\mathcal{U}(\boldsymbol{t}^{\{2\}}, \boldsymbol{x}^{\{1\}})$ becomes the intersection of $\{[x_2, x_3] \in \mathbb{R}^2 : f(t_0, t_q, x_0, x_q) \leq 0, q = 2, 3\}$ 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 1(b) depicts the set of feasible values for x_2, which is the projection of $\mathcal{U}(\boldsymbol{t}^{\{2\}}, \boldsymbol{x}^{\{1\}})$ on the first dimension.

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 $\boldsymbol{t}^{\{r\}}$ and $\boldsymbol{x}^{\{k\}}$,*
1. *(lattice) $\mathcal{U}(\boldsymbol{t}^{\{r\}}, \boldsymbol{x}^{\{k\}})$ is a lattice;*
2. *(monotonicity) $\mathcal{U}(\boldsymbol{t}^{\{r\}}, \boldsymbol{x}^{\{k\}})$ is increasing in $\boldsymbol{x}^{\{k\}}$;*

3. *(dynamic consistency) $\Pi_{:,i:j} \mathcal{U}(\boldsymbol{t}^{\{r\}}, \boldsymbol{x}^{\{k\}}) = \Pi_{:,i:j} \mathcal{U}(\boldsymbol{t}^{\{r'\}}, \boldsymbol{x}^{\{k\}}), \forall i \leq j < r - k + 1, \forall r \leq r' \leq n, \text{ and } \forall t_{r'} \geq t_{r'-1} \geq \cdots \geq t_r.$*

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 $\boldsymbol{x}^{\{k\}}$, which leads to larger future state values, that is, good news for the future. Conversely, "worse past performance" means lower $\boldsymbol{x}^{\{k\}}$, which leads to lower future state values, that is, 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, $t_{r+1}, \ldots, t_{r'}$, 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.

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

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

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 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 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 $\beta \geq \alpha \geq 0$, $\mathcal{M} \subseteq \{1, \ldots, d\}^2$, and functions $\ell : \mathbb{R}^2 \to \mathbb{R}_-$ and $u : \mathbb{R}^2 \to \mathbb{R}_+$ such that ℓ decreases in its second argument and u increases in its second argument, consider the following set:

$$
\mathcal{U}\left(t^{\{r\}}, x^{\{k\}}\right) = \Big\{ [x_{k+1}, \ldots, x_r, x_{n+1}] \in \mathbb{R}^{d \times (r-k+1)} :
$$
$$
\alpha \cdot x_p^m + \ell(t_p, t_q - t_p) \leq x_q^{m'}
$$
$$
\leq \beta \cdot x_p^m + u(t_p, t_q - t_p),
$$
$$
\forall (m, m') \in \mathcal{M}, \ \forall p \in \{0, 1, \ldots, k\},
$$
$$
\forall q \in \{k+1, \ldots, r\} \Big\}. \tag{3}
$$

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 mth 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 depend 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 and is therefore a lattice (Queyranne and Tardella 2006).[4] 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. The time dependency in (3) resembles that of Nohadani and Roy (2017), who model uncertainty progression using a stationary conic description based on a general norm. Our specification allows the bounding functions ℓ and u

to depend on both elapsed time and monitoring time, which enables us to capture important nonstationarities in the process evolution, an issue that we revisit in Section 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, that is, $x_q - x_p = \sum_{i=p+1}^{q} (x_i - x_{i-1})$, one can then rely on the ideas introduced by Bandi and Bertsimas (2012) to formulate the following CLT-type uncertainty set:

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

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 $t_1, \ldots, t_n$ at time t_0. The choice of monitoring times remains fixed throughout the planning horizon, whereas the stopping decision remains dynamic, adapted to available information; more formally, a stopping policy at the kth monitoring time, π_k^S, is a mapping from the σ-algebra induced by the choice of times $t_1, \ldots, t_{n+1}$ and the observation matrix $x^{\{k\}}$ to the set {Stop, Continue}. The DM's *static monitoring problem* thus entails choosing the monitoring times $t^{\{n\}}$ and a set of stopping policies, $\pi^S \overset{\text{def}}{=} \{\pi_0^S, \pi_1^S, \ldots \pi_n^S\}$.

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 $\pi^D = \{\pi_0^D, \pi_1^D, \ldots, \pi_n^D\}$, where each π_k^D is a mapping from the σ-algebra induced by $(t^{\{k\}}, x^{\{k\}})$ to $\{\text{Stop}, \text{Continue}\} \times [t_k, T]$.

In both models, the DM collects a reward $g(t, x)$ when stopping at time t with state 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. *The reward function $g(t, x)$ is monotonic in each x_j, $\forall j \in \{1, \ldots, d\}$.*

As we shall see in the next section, reward monotonicity is consistent with our motivating applications and several practical considerations. Without loss of generality, we henceforth assume that $g(t, x)$ 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 Section 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 agents in imaging studies) or even microsurgical procedures (e.g., collecting tissues for biopsy). In addition, it can also be expensive in monetary terms, generating significant costs for 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 $x(t)$ 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, $x(t)$ could consist of the degree of angiographic lesion, the number of acute rejections, the CAV stage, and the age. The stopping reward $g(t, x)$ 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 Section 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 retransplantation decreases with age, the number of acute rejections, and CAV stage (see the model of rewards in Section 4).

Lending. Consider a lender issuing a loan with collateral consisting of several working assets, for example, accounts receivable, inventory, equipment, and so on. 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. 2012 and U.S. Department of the Treasury 2014, p. 16). Based on these appraisals and on certain advance rates (i.e., haircuts) applied to each collateral type,[6] the lender then calculates a borrowing base, which essentially corresponds to the liquidity-adjusted collateral value (U.S. Department of the Treasury 2014, pp. 15–20). This borrowing base is critical in determining whether the borrower is "over-advanced," that is, whether the outstanding loan is not supported by the collateral value.

Relating to our model, the state $x(t)$ 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 ith collateral type, each monitoring opportunity then allows the lender to observe the current value of the borrowing base, $x^T x(t)$. The reward $g(t, x)$ 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 function of t and x that increases in 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 (U.S. Department of the Treasury 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 $J_k(t^{\{k\}}, x^{\{k\}})$ denote the worst-case value-to-go at time t_k, with the first k observations $x^{\{k\}}$ made at times $t^{\{k\}}$. The Bellman recursions become

$$J_k\left(t^{\{k\}}, x^{\{k\}}\right) = \max\left(g(t_k, x_k), \max_{t_{k+1} \in (t_k, T]} \min_{x_{k+1} \in \mathcal{U}_{k+1}\left(t^{\{k+1\}}, x^{\{k\}}\right)} J_{k+1} \right.$$
$$\left. \times \left(t^{\{k+1\}}, x^{\{k+1\}}\right)\right),$$

(4)

where $J_{n+1}(t^{\{n+1\}}, x^{\{n+1\}}) = g(t_{n+1}, x_{n+1})$. For an intermediate monitoring time t_k, $J_k(t^{\{k\}}, x^{\{k\}})$ 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 $\tau_k^D(t^{\{k\}}, x^{\{k\}})$ for choosing the next monitoring time t_{k+1}. More precisely,

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

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

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

$$V_k\left(t^{\{n+1\}}, x^{\{k\}}\right) = \max\left(g(t_k, x_k), \min_{x_{k+1} \in \mathcal{U}_{k+1}\left(t^{\{n+1\}}, x^{\{k\}}\right)} \right.$$
$$\left. \times V_{k+1}\left(t^{\{n+1\}}, x^{\{k+1\}}\right)\right),$$

(5)

where $V_{n+1}(t^{\{n+1\}}, x^{\{n+1\}}) = g(t_{n+1}, x_{n+1})$. 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 $t^S \in \arg\max_{t^{\{n+1\}}} V_0(t^{\{n+1\}}, x^{\{0\}})$. Let $V = V_0(t^S, x^{\{0\}})$.

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*

$$J = V, \qquad t_1^S \in \tau_1^D(t_0, x_0).$$

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 $t_2^S, \ldots, t_n^S$ 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, then 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 noncommittal fashion, that is, by implementing the first monitoring time and resolving 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 $k = 0, 1, \ldots, n$:*

1. At the kth monitoring time t_k, having observed 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 x_k, and observation matrix $x^{\{k\}}$. Let $V(t^{\{k\}}, x^{\{k\}})$ denote the worst-case optimal reward, and let $t_1^S(t^{\{k\}}, x^{\{k\}})$ denote the first monitoring time under this policy.

2. Stop if $g(t_k, x_k) \geq V(t^{\{k\}}, x^{\{k\}})$. Otherwise, continue and $\tau_k^D(t^{\{k\}}, x^{\{k\}}) \leftarrow t_1^S(t^{\{k\}}, x^{\{k\}})$.

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 $t^{\{n+1\}}$. Nature's optimal (worst-case) response when the DM continues in period t_k can be recovered recursively as*

$$\underline{x}_k\left(t^{\{n+1\}}\right) \stackrel{\text{def}}{=} \min\Big\{\xi \in \mathbb{R}^d : \xi \in \mathcal{U}_k\Big(t^{\{n+1\}}, \\ \Big[x_0, \underline{x}_1\Big(t^{\{n+1\}}\Big), \ldots, \underline{x}_{k-1}\Big(t^{\{n+1\}}\Big)\Big]\Big)\Big\},$$
$$1 \leq k \leq n + 1. \quad (6)$$

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 reformulated as

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

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*

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

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 that we can find a solution for the following problem:

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

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 Section 2.2). For simplicity, we take $\alpha = \beta = 1$, so that

$$\mathcal{U}\left(t^{\{r\}}, x^{\{k\}}\right) = \Big\{ [x_{k+1}, \dots, x_r, x_{n+1}] \in \mathbb{R}^{d \times (r-k+1)} :$$
$$x_p^m + \ell(t_p, t_q - t_p) \le x_q^{m'} \le x_p^m$$
$$+ u(t_p, t_q - t_p), \forall (m, m') \in \mathcal{M},$$
$$\forall p \in \{0, 1, \dots, k\}, \forall q \in \{k+1, \dots, r\} \Big\}. \qquad (10)$$

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, that is, 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), *we have the following:*
  1. *If* $\ell(\cdot)$ *is convex, then*

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

and $t_2^S, \dots, t_n^S$ *can be chosen arbitrarily from* $[t_1^S, T]$.
  2. *If* $\ell(\cdot)$ *is concave, then*

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

where $\phi_k(t) \stackrel{\text{def}}{=} \underline{x}_0 + k\ell(t/k)\mathbf{1}, \forall k \in \{n, n+1\}$. *Moreover, if* $g(t, x)$ *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, 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.

3.3.2. Nonstationary Bound Functions. We now treat the more general case in which the bounds depend on both monitoring time and elapsed time, that is, $\ell(t_p, t_q - t_p)$ and $u(t_p, t_q - t_p)$. Allowing the bounds to depend on the monitoring time enables us to capture seasonality or nonstationary patterns for the uncertainty evolution, for instance, nonstationary 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 microscale 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), *we have the following:*

1. *If* $\ell(t, \delta)$ *is decreasing in* t *and convex in* δ, *then*

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

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

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

2. *If* $\ell(t, \delta)$ *is increasing in* t *and concave in* δ, *then*

$$V = \max\left(\max_{t_n \in [0,T]} g(t_n, x_0 + \xi_n(t_n)\mathbf{1}),\right.$$
$$\left. \times g(T, x_0 + \xi_{n+1}(T)\mathbf{1})\right), \quad \text{where} \tag{13a}$$

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

Furthermore, if ℓ *is jointly concave, then* (13b) *is a convex optimization problem; and if* $g(t, x)$ *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, the 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 $g(t_n, \underline{x}_n(t^{\{n+1\}}))$ is a supermodular function of $t^{\{n+1\}}$, then problem (9) would involve maximizing a supermodular function over a lattice—a problem that can be tackled through tractable combinatorial algorithms (see, e.g., Schrijver 2003, Fujishige 2005, and references therein). Although it is difficult to exactly characterize when $g(t_n, \underline{x}_n(t^{\{n+1\}}))$ 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*

1. $g(t, x)$ *is supermodular in* x, *convex in every component* x_k, *and*
2. ℓ *is supermodular and decreasing in* (t_p, t_q) *on the lattice* $[0, T]^2$,

then $g(t_k, \underline{x}_k(t^{\{k\}}))$ *is a supermodular function of* $t^{\{k-1\}}$ *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 $g(t_n, \underline{x}_n(t^{\{n\}}))$ over the lattice $0 \le t_1 \le t_2 \le \cdots \le t_{n-1} \le t_n$. 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, and so on. Some of the premises of Proposition 3 are reasonable. Having rewards g that are supermodular is natural in 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 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 piecewise 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 retransplantation 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-à-vis established guidelines in the medical community. Simulation suggests that our policies provide a QALY distribution that dominates the one 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 postoperative outlooks have improved over the last decades, with five-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 posttransplant 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 United States (Healthcare Bluebook 2019).

Second, after CAV detection, retransplantation is the main viable treatment option.[9] This itself introduces a difficult trade-off. On the one hand, retransplantation 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, John et al. 1999, Johnson 2007, Vistarini et al. 2010, Saito et al. 2013). However, as the disease progresses, the quality of life before and

after retransplantation and the survivability prognosis significantly worsen.

In order to manage CAV in an effective way, patients and their physicians must make joint decisions concerning monitoring and retransplantation 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 retransplantation 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 retransplantation 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 toward a quantitative process for personalizing such decisions; we discuss the limitations and future research directions in Section 4.5.

4.1. Data Description

Very limited data exist that can be used to guide CAV disease management decisions. To calibrate CAV development, we use a data set 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 data set comprises 2,846 monitoring observations.

To calibrate survival after retransplantation, we obtained data from Novitzky et al. (1985), Schnetzler et al. (1998), Tjang (2007), and Copeland et al. (2011) 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 who had a heart transplantation at $t = 0$, at an age of A. The patient and his or her physician need to decide when to monitor for CAV progression and when to perform a retransplantation throughout a horizon T. To match our robust monitoring framework to this problem, we must characterize

the rewards collected upon stopping (i.e., at retransplantation) 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, that is, from a state (i, j) to a state (i', j') with $i' \geq i$ and with $j' = H$ whenever $j = H$.

Rewards. We calculate the rewards from retransplanting (i.e., "stopping") at time t as the total quality-adjusted life years (QALY), provided that the patient is alive at t:

reward at t = (total QALY when retransplanted at t)

$\qquad \times$ (Survival rate at t)

$\qquad$ = ((QALY until t) + (QALY after

$\qquad$ retransplant at t)) $\times$ (Survival rate at t).

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

Quality of life prior to retransplantation 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 we let $LY(t)$ denote the expected life-years after retransplantation at t, then the total QALY when retransplanting at time t can be expressed as

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

Figure 2. Disease Progression Model
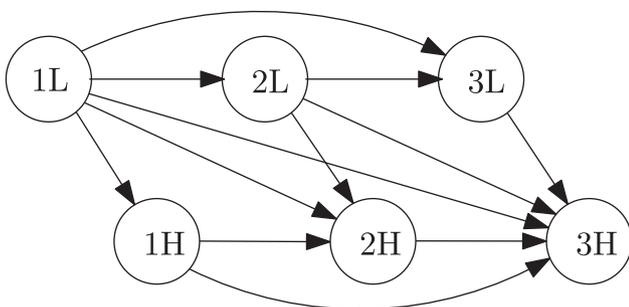
Note. The number $i \in \{1, 2, 3\}$ captures the CAV stage, and the letter $j \in \{L, H\}$ indicates whether the patient had a "low" (< 4) or "high" (≥ 4) number of acute rejection episodes.

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

To estimate the life-years following retransplantation, 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 retransplantions done too soon after the primary transplant tend to result in worse survival rates compared with those performed later (Johnson et al. 2007, Saito et al. 2013, Goldraich et al. 2016). In view of this, we consider the following piecewise-linear model for the life-years after retransplantation:

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

which we fit using regression based on our retransplantation data, to obtain the following:[13]

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

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 posttransplant year t (with outcomes 1 = "alive" and 0 = "death") on the patient's starting age A and disease stage at the time of retransplant (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 D.3 of Online Appendix D.

Consistent with medical data, as well as intuition, a patient who is older, has a more advanced form of CAV, or suffered from four or more acute rejections has a worse survival rate. More importantly, Table D.3 also shows that the survival rate decreases 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 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 (posttransplant) time t, the time until he or she progresses to stage (i', j'), denoted by $L_{iji'j'}(t, A)$, is exponentially distributed with rate $\lambda_{iji'j'}(t, A)$, where

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

Table E.4 in Online Appendix E 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 nonprobabilistic 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 $L_{iji'j'}(t, A)$ at confidence level ρ of the form $[\frac{-\ln(\rho)}{\lambda_{iji'j'}(t,A)}, \infty)$.

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*, that is, sequences of nodes connected by arcs in the graph of Figure 2. The following important observation will allow us to simplify the formulation considerably.

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

1. *The worst-case disease progression path does not involve nodes 1H and 2H;*

2. *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*
$= 1 + 1_{x^{2L}(t)>0} + 1_{x^{3L}(t)>0}, 1_{j=H} = 1_{x^{3H}(t)>0}.$

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 retransplant at time t as

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

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 $x^{2L}(t) > 0$, then 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 $x^{2L}(t) = 0$, then 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 $x^2(t + \Delta t)$ given by $\bar{x}^{2L}(t + \Delta t)|\mathbf{x}(t) = (\Delta t + \frac{\ln(\rho)}{\lambda_{1L2L}(t,A)})^+$.

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

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

$$(19a)$$

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

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

This implies that, for a fixed confidence level ρ, a state observation $x(t)$ 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 3(a) 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} (per Table E.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 3(b).

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 $t_1, \ldots, t_n$ chosen at time 0 can be written as

$$\max_{t^{\{n+1\}}} \max\big(g\big(t_n, \bar{x}^{2L}(t_n), \bar{x}^{3L}(t_n), \bar{x}^{3H}(t_n)\big),$$
$$g\big(t_{n+1}, \bar{x}^{2L}(t_{n+1}), \bar{x}^{3L}(t_{n+1}), \bar{x}^{3H}(t_{n+1})\big)$$
$$\text{where } \bar{x}^{2L}(t_k), \bar{x}^{3L}(t_k) \text{ and } \bar{x}^{3H}(t_k) \text{ satisfy}$$
$$(19a) - (19c) \text{ with } t := t_k, \Delta t := t_{k+1} - t_k.$$
$$(20)$$

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 Online Appendix E.

4.4. Results

We consider CAV management for a patient who received heart transplantation at an age of 33–62. According to the ISHLT, the patient would be recommended to undergo annual or biannual examinations of coronary arteries through coronary angiography and
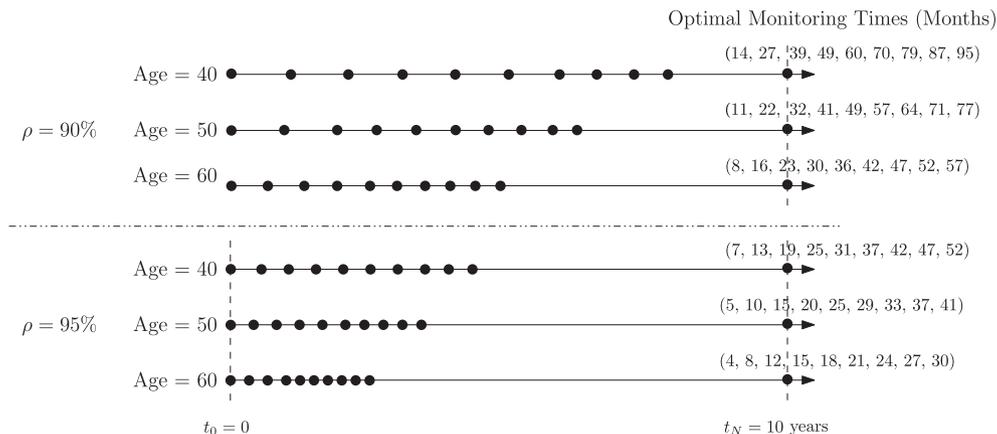
Figure 3. The Worst-Case Values for $x^{2L}(\Delta t)$ Given an Initial Observation at $t = 0$

(a)
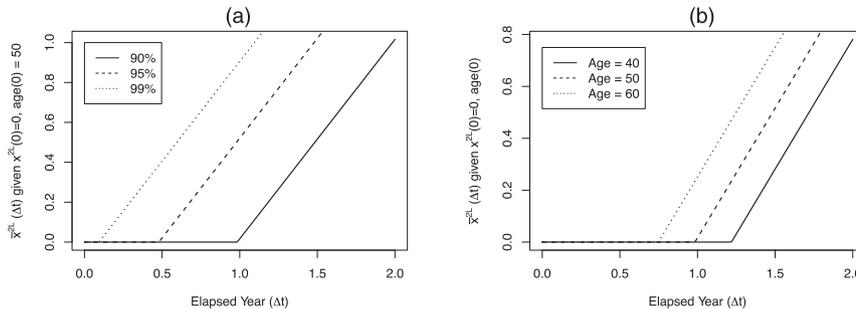
Worst-case values for $x^{2L}(\Delta t)$ as a function of the confidence level ρ (starting age $A = 50$).

(b)

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

Figure 4. Monitoring Times (in Months) for Static Monitoring Prescribed by Our Framework

intravascular ultrasonography to monitor progression in the first 3–5 years after transplant (Costanzo et al. 2010, Labarrere et al. 2017) and perform retransplantation 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).

Notably, this policy departs from equisized 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 E.4], life-years following retransplantation [per (16)], and also survival rate (per Table D.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, that is, 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 reevaluate 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 resolving 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 seven months later with a high number of acute rejections (for $\rho = 90\%$). But when the patient is diagnosed with stage 2, he or she should be monitored earlier: in five months for a low number of acute rejections and in four months with a high number of acute rejections.

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).

Table 1. The Worst-Case Optimal Next-Monitoring Time in Months ($\rho = 90\%$)

| 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

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 retransplant 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 when the current stopping reward exceeds the worst-case reward-to-go, we refer to as *uniform RO*.

The results are summarized in Figure 5 (Table E.5 of Online Appendix E 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/retransplantation 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.

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.

Figure 5. Boxplot for Simulated Rewards Under the Benchmark, Static and Dynamic RO Policies
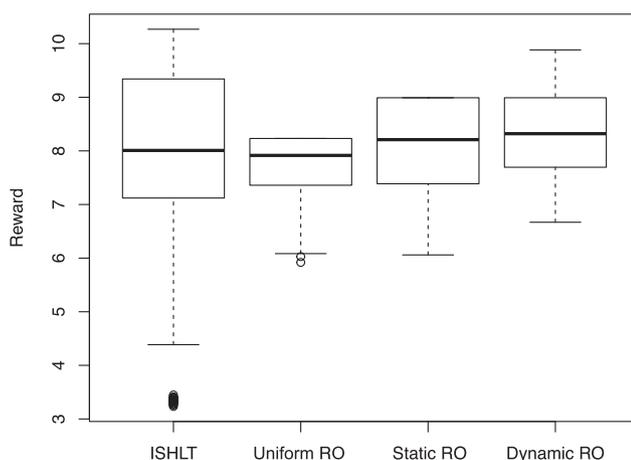
Note. $\rho = 90\%$; number of simulations $= 10^3$; number of monitoring times $= 10$; $T = 10$ years.

First, to more accurately capture the progression for CAV, one should consider all the available information concerning the state of the patient, including, for example, 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 retransplantation. One such candidate is revascularization (Cheng and Kobashigawa 2016), which has proven effective for certain patients. Assuming that 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 revascularization or retransplantation.

Third, our framework only allowed addressing retransplantation 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 a 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 Section B.1 of Online Appendix B 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 equidistant 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 Section 4.5.

Endnotes

[1] We first define the operator "⊑" as follows: for any nonempty subsets $A, B \subseteq \mathbb{R}^d$, we have $A \sqsubseteq B$ if $\min(x, y) \in A$ and $\max(x, y) \in B$ holds for any $x \in A, x \in B$. We can then define the notion of "increasing" subsets of $\mathbb{R}^d$. Suppose that T is a nonempty subset of $\mathbb{R}^p$ and that A_t is a sequence of nonempty subsets of $\mathbb{R}^d$ parameterized by $t \in T$. Then A_t is "increasing" ("decreasing") in t if $A_{t_1} \sqsubseteq A_{t_2}$ ($A_{t_2} \sqsubseteq A_{t_1}$), for any $t_1, t_2 \in T$ with $t_2 \geq t_1$.

[2] Taking 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.

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

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

[5] That 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.

[6] In 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 U.S. Department of the Treasury 2014, pp. 17–19).

[7] Note that this remains in line with Theorem 2—simply pick $t_1^S = t_2^S = \ldots = t_n^S$ 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.

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

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

[10] Data are publicly available in an R package called "msm."

[11] We 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 Online Appendix C).

[12] To 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.

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

[14] Technically, survival at t could depend also on how long the patient spent at each stage, before arriving at the stage he or she 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.

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

[16] Several 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 $\epsilon \to 0$ must be equal, so Theorem 1 holds for the uncertainty set given by the upper bounds in (19a)–(19c).

[17] Note 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 solve the static monitoring problem as an integer programming. Details are presented in Online Appendix E.

References

Ayer T, Alagoz O, Stout NK (2012) OR forum—a POMDP approach to personalize mammography screening decisions. *Oper. Res.* 60(5):1019–1034.

Bandi C, Bertsimas D (2012) Tractable stochastic analysis in high dimensions via robust optimization. *Math. Programming* 134(1):23–70.

Bayraktar E, Yao S (2014) On the robust optimal stopping problem. *SIAM J. Control Optim.* 52(5):3135–3175.

Bayraktar E, Karatzas I, Yao S (2010) Optimal stopping for dynamic convex risk measures. *Illinois J. Math.* 54(3):1025–1067.

Ben-Tal A, Nemirovski A (2002) Robust optimization—methodology and applications. *Math. Programming* 92(3):453–480.

Ben-Tal A, El Ghaoui L, Nemirovski A (2009) *Robust Optimization* (Princeton University Press, Princeton, NJ).

Ben-Tal A, Goryashko A, Guslitzer E, Nemirovski A (2004) Adjustable robust solutions of uncertain linear programs. *Math. Programming* 99(2):351–376.

Bertsimas D, Vayanos P (2017) Data-driven learning in dynamic pricing using adaptive optimization. Working paper, Sloan School of Management at the Massachusetts Institute of Technology, Cambridge, MA.

Bertsimas D, Brown DB, Caramanis C (2011) Theory and applications of robust optimization. *SIAM Rev.* 53(3):464–501.

Bertsimas D, Goyal V, Lu B (2014) A tight characterization of the performance of static solutions in two-stage adjustable robust linear optimization. *Math. Programming* 150(2):281–319.

Cheng R, Kobashigawa J (2016) Transplant CAD: The role of PCI in the transplanted heart. Report, American College of Cardiology, Washington, DC.

Copeland H, Coelho-Anderson R, Mineburg N, McCarthy M, Copeland JG (2011) Elective cardiac retransplantation: A viable option that can be repeated. *J. Thoracic Cardiovascular Surgery* 141(3):822–827.

Costanzo MR, Dipchand A, Starling R, Anderson A, Chan M, Desai S, Fedson S, et al. (2010) The International Society of Heart and Lung Transplantation guidelines for the care of heart transplant recipients. *J. Heart Lung Transplantation* 29(8):914–956.

Cutler DM, Poterba JM, Summers LH (1991) Speculative dynamics. *Rev. Econom. Stud.* 58(3):529–546.

Delage E, Iancu DA (2015) Robust multistage decision making. Aleman DM, Thiele AC, eds. *The Operations Research Revolution*, TutORials in Operations Research (INFORMS, Catonsville, MD), 20–26.

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

Dullerud GE, Paganini F (2005) *A Course in Robust Control Theory* (Springer, New York).

Feingold B, Webber SA, Bryce CL, Park SY, Tomko HE, West SC, Hart SA, Mahle WT, Smith KJ (2015) Cost-effectiveness of pediatric heart transplantation across a positive crossmatch for high waitlist urgency candidates. *Amer. J. Transplantation* 15(11):2978–2985.

Foley CF, Raman A, Craig NC (2012) Inventory-based lending industry note. Background Note 612-057, Harvard Business School, Boston.

Fujishige S (2005) *Submodular Functions and Optimization*, 2nd ed. (Elsevier, Amsterdam).

Goldraich LA, Stehlik J, Kucheryavaya AY, Edwards LB, Ross HJ (2016) Retransplant and medical therapy for cardiac allograft vasculopathy: International Society for Heart and Lung Transplantation Registry analysis. *Amer. J. Transplantation* 16(1):301–309.

Gorton G, Winton A (2003) Financial intermediation. Constantinides GM, Harris M, Stulz RM, eds. *Handbook of the Economics of Finance* (North-Holland, Amsterdam), 431–552.

Grauhan O, Huang Chang H, Albert W, Hiemann N, Meyer R, Lehmkuhl H, Dandel M, et al. (2010) Impact of donor-transmitted coronary atherosclerosis on quality of life (QOL) and quality-adjusted life years (QALY) after heart transplantation. *Appl. Cardiopulmonary Pathophysiology* 14(1):58–65.

Hanany E, Klibanoff P (2009) Updating ambiguity averse preferences. *B. E. J. Theoret. Econom.* 9(1):1–53.

Healthcare Bluebook (2019) Cardiac catheterization with and without coronary angiogram. Accessed October 5, 2020, https://www.mdsave.com/procedures/cardiac-catheterization-with-coronary-angiogram/d785fec8.

Iancu DA, Sharma M, Sviridenko M (2013) Supermodularity and affine policies in dynamic robust optimization. *Oper. Res.* 61(4):941–956.

Ibanez A, Zapatero F (2004) Monte Carlo valuation of American options through computation of the optimal exercise frontier. *J. Financial Quant. Anal.* 39(2):253–275.

Jaillet P, Jena SD, Ng TS, Sim M (2016) Satisficing awakens: Models to mitigate uncertainty. Working paper.

John R, Chen JM, Weinberg A, Oz MC, Mancini D, Itescu S, Galantowicz ME, Smith CR, Rose EA, Edwards NM (1999) Long-term survival after cardiac retransplantation: A twenty-year single-center experience. *J. Thoracic Cardiovascular Surgery* 117(3):543–555.

Johnson MR (2007) When is retransplantation a viable option? *Heart Failure Clinics* 3(1):97–105.

Johnson MR, Aaronson KD, Canter CE, Kirklin JK, Mancini DM, Mehra MR, Radovancevic B, Taylor DO, Webber SA (2007) Heart retransplantation. *Amer. J. Transplantation* 7(9):2075–2081.

Karatzas I, Shreve SE (1998) *Methods of Mathematical Finance* (Springer, New York).

Karatzas I, Shreve SE (2012) *Brownian Motion and Stochastic Calculus* (Springer, New York).

Kobashigawa J (2000) What is the optimal prophylaxis for treatment of cardiac allograft vasculopathy? *Trials* 1(3):166–171.

Kobashigawa J, Pauly D, Starling RC, Eisen H, Ross H, Wang S-S, Cantin B, et al. (2013) Cardiac allograft vasculopathy by intravascular ultrasound in heart transplant patients. *JACC Heart Failure* 1(5):389–399.

Kolodko A, Schoenmakers J (2006) Iterative construction of the optimal Bermudan stopping time. *Finance Stochastics* 10(1):27–49.

Labarrere CA, Jaeger BR, Kassab GS (2017) Cardiac allograft vasculopathy: Microvascular arteriolar capillaries (capioles) and survival. *Frontiers Bioscience (Elite Ed.)* 9(1):110–128.

Lavieri M, Puterman M, Tyldesley S, Morris W (2012) When to treat prostate cancer patients based on their PSA dynamics. *IIE Trans. Healthcare Systems Engrg.* 2(1):62–77.

Leland HE (1994) Corporate debt value, bond covenants, and optimal capital structure. *J. Finance* 49(4):1213–1252.

Leland HE, Toft KB (1996) Optimal capital structure, endogenous bankruptcy, and the term structure of credit spreads. *J. Finance* 51(3):987–1019.

Long EF, Swain GW, Mangi AA (2014) Comparative survival and cost effectiveness of advanced therapies for end-stage heart failure. *Circulation Heart Failure* 7(3):470–478.

Machina M (1989) Dynamic consistency and non-expected utility models of choice under uncertainty. *J. Econom. Literature* 27(4):1622–1668.

Maillart LM, Ivy JS, Ransom S, Diehl K (2008) Assessing dynamic breast cancer screening policies. *Oper. Res.* 56(6):1411–1427.

Marandi A, den Hertog D (2018) When are static and adjustable robust optimization problems with constraint-wise uncertainty equivalent? *Math. Programming* 170:555–568.

Mayo Clinic (2019) Coronary angiogram. Accessed October 5, 2020, https://www.mayoclinic.org/tests-procedures/coronary-angiogram/about/pac-20384904.

Montazeri A (2009) Quality of life data as prognostic indicators of survival in cancer patients: An overview of the literature from 1982 to 2008. *Health Quality Life Outcomes* 7(1):102.

Nohadani O, Roy A (2017) Robust optimization with time-dependent uncertainty in radiation therapy. *IISE Trans. Healthcare Systems Engrg.* 7(2):81–92.

Nohadani O, Sharma K (2016) Optimization under decision-dependent uncertainty. *SIAM J. Optim.* 28(2):1773–1795.

Novitzky D, Cooper DKC, Lanza RP, Barnard CN (1985) Further cardiac transplant procedures in patients with heterotopic heart transplants. *Ann. Thoracic Surgery* 39(2):149–154.

Poss M (2013) Robust combinatorial optimization with variable budgeted uncertainty. *4OR* 11(1):75–92.

Queyranne M, Tardella F (2006) Bimonotone linear inequalities and sublattices of r^n. *Linear Algebra Appl.* 413:100–120.

Rajan R, Winton A (1995) Covenants and collateral as incentives to monitor. *J. Finance.* 50(4):1113–1146.

Ramzy D, Rao V, Brahm J, Miriuka S, Delgado D, Ross HJ (2005) Cardiac allograft vasculopathy: A review. *Canadian J. Surgery* 48(4):319.

Riedel F (2009) Optimal stopping with multiple priors. *Econometrica* 77(3):857–908.

Saito A, Novick RJ, Kiaii B, McKenzie FN, Quantz M, Pflugfelder P, Fisher G, Chu MWA (2013) Early and late outcomes after cardiac retransplantation. *Canadian J. Surgery* 56(1):21.

Schnetzler B, Pavie A, Dorent R, Camproux A-C, Leger P, Delcourt A, Gandjbakhch I (1998) Heart retransplantation: A 23-year single-center clinical experience. *Ann. Thoracic Surgery* 65(4):978–983.

Schrijver A (2003) *Combinatorial Optimization: Polyhedra and Efficiency* (Springer, New York).

Shechter SM, Bailey MD, Schaefer AJ, Roberts MS (2008) The optimal time to initiate HIV therapy under ordered health states. *Oper. Res.* 56(1):20–33.

Snell JL (1952) Applications of martingale system theorems. *Trans. Amer. Math. Soc.* 73(2):293–312.

Taylor HM (1968) Optimal stopping in a Markov process. *Ann. Math. Statist.* 39(4):1333–1344.

Tjang YS (2007) Cardiac retransplantation. Unpublished doctoral dissertation, Universitätsklinik der Ruhr-Universität Bochum, Germany.

Topkis DM (1998) *Supermodularity and Complementarity* (Princeton University Press, Princeton, NJ).

U.S. Department of Treasury (2014) Accounts receivable and inventory financing. Report, Comptroller of the Currency, Administrator of National Banks, Washington, DC.

Vistarini N, Pellegrini C, Aiello M, Alloni A, Monterosso C, Cattadori B, Tinelli C, D'Armini AM, Vigano M (2010) Should we perform heart retransplantation in early graft failure? *Transplant Internat.* 23(1):47–53.

Wilhelm MJ (2015) Long-term outcome following heart transplantation: current perspective. *J. Thoracic Disease* 7(3):549.

Wilmott P, Dewynne J, Howison S (1994) *Option Pricing: Mathematical Models and Computation* (Oxford Financial Press, Oxford, UK).

Zhang X, Kamgarpour M, Georghiou A, Goulart P, Lygeros J (2016) Robust optimal control with adjustable uncertainty sets. *Automatica* 75(C):249–259.

Zhou K, Doyle JC (1998) *Essentials of Robust Control* (Prentice Hall, Upper Saddle River, NJ).