

Robustness of proactive ICU transfer policies.

Julien Grand-Clément

IEOR Department, Columbia University jg3728@columbia.edu

Carri W. Chan

Columbia Business School, Columbia University, cwchan@gsb.columbia.edu

Vineet Goyal

IEOR Department, Columbia University vg2277@columbia.edu

Gabriel Escobar

Kaiser Permanente Division of Research, gabriel.escobar@kp.org

Patients whose transfer to the Intensive Care Unit (ICU) is unplanned are prone to higher mortality rates and longer length-of-stay than those who were admitted directly to the ICU. Recent advances in machine learning to predict patient deterioration have introduced the possibility of *proactive transfer* from the ward to the ICU. In this work, we study the problem of finding *robust* patient transfer policies which account for uncertainty in statistical estimates due to data limitations when optimizing to improve overall patient care. We propose a Markov Decision Process model to capture the evolution of patient health, where the states represent a measure of patient severity. Under fairly general assumptions, we show that an optimal transfer policy has a threshold structure, i.e., that it transfers all patients above a certain severity level to the ICU (subject to available capacity). As model parameters are typically determined based on statistical estimations from real-world data, they are inherently subject to misspecification and estimation errors. We account for this parameter uncertainty by deriving a robust policy that optimizes the worst-case reward across all plausible values of the model parameters. We show that the robust policy also has a threshold structure under fairly general assumptions. Moreover, it is more aggressive in transferring patients than the optimal nominal policy, which does not take into account parameter uncertainty. We present computational experiments using a dataset of hospitalizations at 21 Kaiser Permanente Northern California hospitals, and present empirical evidence of the sensitivity of various hospital metrics (mortality, length-of-stay, average ICU occupancy) to small changes in the parameters. While threshold policies are a simplification of the actual complex sequence of decisions leading (or not) to a transfer to the ICU, our work provides useful insights into the impact of parameter uncertainty on deriving simple policies for proactive ICU transfer that have strong empirical performance and theoretical guarantees.

Key words: Intensive Care Units, Markov Models, Robust Optimization, Threshold policies.

1. Introduction.

In a hospital, critically ill patients are treated in the Intensive Care Unit (ICU), where they require a significant amount of human and material resources (Milbrandt et al. (2008)). Effective management of ICUs has substantial implications, both for the patient outcomes and for the operational costs of the hospital. The sudden health deterioration of a patient in the general medical/surgical ward can result in an unplanned transfer to the ICU and a severe downturn in the chance of survival of the patient. Such unplanned transfers typically have worse outcomes than patients who are directly admitted to the ICU (e.g. Barnett et al. (2002), Escobar et al. (2013)). Developing strategies to effectively manage the limited ICU beds (Green 2002) is becoming even more critical as demand for ICU care is increasing (Mullins et al. 2013). The primary focus of this work is to derive and evaluate *robust proactive ICU transfer policies* in order to improve patient flow and patient outcomes.

Recent advances in machine learning have brought real-time risk scores of a patient’s likelihood of deterioration available to clinicians’ use in hospitals (Kipnis et al. 2016). Consequently, understanding the impact of intervening on patients based on such scores (which is currently occurring based primarily on informed clinical judgment and empirical observation) needs to be better understood from a theoretical perspective. In practice, alerts based on such scores are known to trigger multiple types of response, which can range from simple maneuvers (e.g., increased monitoring, one time fluid boluses) to immediate transfer to the ICU. In some cases, alerts trigger discontinuation of life support. In this work, as a first step towards better theoretical understanding of the pathways involved in early warning systems and in order to focus on the impact of parameter uncertainty, we focus on proactive ICU admissions, while recognizing this is just one of many potential interventions that could take place. Using the data of nearly 300,000 hospitalizations at

KPNC, the authors in Hu et al. (2018) provide empirical evidence that proactively transferring patients to the ICU can significantly reduce the average mortality risk and the Length-Of-Stay (LOS). However, using simulation to consider the system-wide effect of various proactive transfer policies, the authors provide a cautionary tale that overly aggressive transfers can have a significant impact on increasing ICU occupancy, which is associated with worse outcomes (e.g. Kc and Terwiesch (2012)).

While Hu et al. (2018) demonstrates that there is promising potential in the use of proactive transfers, there are some limitations with respect to the insights developed at the system-wide level. First, in practice, the actual ICU admission decision relies on a complex sequence of events that are activated when a patient’s severity score reaches the alert threshold. Therefore, the model of ICU transfer based only on the severity scores is a simplification. Second, there is limited theoretical basis for the class of policies (threshold and random) which are considered. But, perhaps more critically, the core parameters of the simulation model are calibrated from real data and are subject to uncertainty. In particular, the transition rates of a Markov chain are estimated from a finite dataset of patient hospitalizations and are an approximation of the true parameters. This is concerning because the performance of a policy can significantly deteriorate, even under small variations from the true parameters (e.g. Section 5 of this paper). Consequently, an optimal transfer policy for the estimated parameters might perform very poorly in practice even if the true parameters are close but different. This limitation is widely acknowledged in the healthcare community and is typically addressed by conducting sensitivity analysis. This is the approach taken in Hu et al. (2018). However, when models have many parameters – as ours does – the comprehensiveness of these types of sensitivity analysis can be limited due to computational reasons. This paper proposes to look at the problem by optimizing the worst-case performance over an uncertainty set by using tools from robust optimization.

Our goal in this paper is to develop *robust* transfer policies, i.e., transfer policies with guarantees of good performance over a given set of plausible hospital parameters, which are consistent with

available data. This is in contrast to *nominal* transfer policies, which are only guaranteed to have good performances for known fixed values of the parameters and could have very bad performance for close, but different, parameters. In doing so, we will leverage results from the Robust MDP literature (e.g. Iyengar (2005), Mannor et al. (2007), Wiesemann et al. (2013), Goh et al. (2018), Goyal and Grand-Clement (2018)) to develop a theoretical and empirical basis for our proposed transfer policies.

Our main contributions can be summarized as follows:

Markov model for a single patient. We propose an approximation of the full hospital dynamics, using the health evolution of a *single* patient. In particular, we present a Markov Decision Process (MDP) to model the patient health evolution. This MDP is able to capture the fundamental trade-off between the benefit of proactive transfer for individual patients versus suboptimal use of limited ICU resources for patients who may not ‘really need it’.

Structure of optimal nominal policies. Under fairly general and interpretable assumptions that we expect to hold in practice, we show that an optimal proactive transfer policy in our single-patient MDP is a threshold policy. In particular, there exists an optimal policy that transfers all patients above a certain severity score. This structure is particularly nice because of its interpretability and implementability.

Robustness of transfer policies. Building upon the nominal model, we incorporate the real-world limitation of parameter uncertainty. In particular, we consider parameter misspecification for the transition matrix in our single-patient MDP model. Underlying factors, such as genetics, demographics, and/or physiologic characteristics of certain diseases, could dictate the health evolution of patients in specific health states. As such, we consider a model of uncertainty where the transition probabilities of different health states are *correlated* and depend on a factor model (Goh et al. (2018) and Goyal and Grand-Clement (2018)), which is a departure from Iyengar (2005) and Wiesemann et al. (2013).

We present an algorithm to compute an *optimal robust policy*, i.e., a policy that maximizes the worst-case possible outcomes over all plausible transition matrices. Moreover, we prove structural

results for the optimal robust policy. In particular, an optimal robust policy is always deterministic and – under the same assumptions as in our nominal model – of threshold type. Additionally, the threshold of the optimal robust policy is lower than the threshold of the optimal nominal policy. Therefore, the optimal robust policy transfers more patients to the ICU than the optimal nominal policy.

Numerical experiments. We present detailed numerical experiments to compare the performances of the optimal nominal and robust transfer policies, making use of the hospitalization data of almost 300,000 patients at Kaiser Permanente. We observe that, for our single-patient MDP, the performance of the optimal nominal policy can deteriorate even for small variations of the model parameters. Moreover, there are significant differences in the recommended thresholds between the nominal and robust policies, which reflects that these policies could have substantial differences in the proportion of patients who are proactively transferred. When considering the full hospital model, we observe similar deteriorations in performance (as measured by mortality, length-of-stay, average ICU occupancy) even for small parameters deviations. Additionally, we find that correlated uncertainty in the transition matrix results in different and more useful insights than when considering uncorrelated uncertainty which can be overly conservative. We also highlight the contrast between this worst-case analysis and more standard sensitivity analysis approaches.

The rest of the paper is structured as follows. We finish this section with a brief overview of related literature. In Section 2, we present the hospital model and the Markov chain that describes the evolution of a patient’s health. In Section 3, we introduce a Markov Decision Process to approximate the full hospital model and we theoretically characterize the structure of optimal nominal policies. We address parameter uncertainty in Section 4, where we introduce our model of uncertainty and we prove some theoretical results on the structure of optimal robust policies. In Section 5, we present computational experiments based on a dataset from Kaiser Permanente Northern California and we examine the contrast between the optimal nominal and optimal robust policies.

Notations. For an integer $n \geq 0$, we denote by $[n]$ the set $\{1, \dots, n\}$. Vectors and matrices are in bold font whereas scalars are in regular font, except for policies π which are also in regular font. The vector \mathbf{e} has every component equal to one and its dimension depends on the context.

1.1. Related work.

Our work mainly involves three topics of research: (i) ICU management and proactive care in hospitals, (ii) Markov Decision Process in healthcare and (iii) robust Markov Decision Process, particularly those applied to problems in healthcare.

ICU management and proactive care in hospitals. There is a large and growing body of literature in both the operations and medical literature on the management of ICUs. For instance, the impact of congestion and demand-driven discharges has been considered both empirically (e.g. Chrusch et al. (2009) and Kc and Terwiesch (2012)) and theoretically (e.g. Chan et al. (2012)). More closely related to our work is admission into the ICU. A number of papers, including Shmueli et al. (2004), Bountourelis et al. (2012), and Kim et al. (2014) consider the impact of ICU admission decisions on patient outcomes when patients arrive to the hospital. Theoretical guarantees for the optimality of threshold policies in various admission control settings (including the ICU) have been investigated in Altman et al. (2001) and Shmueli et al. (2003). In contrast to these works, we consider a dynamic decision-making approach where patients can be admitted to the ICU at any point while they are in the general medical/surgical ward. In many hospitals, rapid response teams are utilized to make these types of dynamic decisions about which patients should be admitted to the ICU (Butcher et al. (2013), Guirgis et al. (2013)).

Given the limited number of hospital resources and the adverse impact of strained ICUs on the quality of care provided (Opgenorth et al. (2018)), there has been a growing interest in the development of predictive models for patients dynamics and outcomes (LOS, death, readmission to the ICU, etc.). For instance, the authors in Putnam et al. (2002) develop a risk-adjustment metric to predict patient hospitalization. The authors in Bilben et al. (2016) study the performances of the National Early Warning Score (NEWS) as a risk score for ED patients, while the NEWS risk score

has been specifically developed for patients in the hospital ward. Peck et al. (2012) utilize expert opinion, naive Bayes and logistic regression to predict the number of patients in the Emergency Department (ED) who will be admitted to a particular inpatient unit. Higgins et al. (1997) develop ICU admission scores aimed at predicting mortality risk after coronary artery bypass grafting, while Brunelli et al. (2008) develop a scoring system for predicting ICU admission after major lung resection; see Rapsang and Shyam (2014) for a review of the medical severity scores for ICU patients.

The operations community has studied how preventative/proactive care can be used to improve patient care. For instance, Xu and Chan (2016) design efficient proactive ED admission control policies based on predictions of potential patient arrivals, while proactive care using Markov models can be dated back to at least Özekici and Pliska (1991), where the authors introduce an MDP to compute an optimal inspection schedule in the case of post-operative periumbilical pruritus and breast cancer.

This paper focuses on the dynamic decision of whether and when to proactively transfer a patient to the ICU based on a patient’s risk of deterioration (e.g. Kipnis et al. (2016)). The particular problem we study is related to that in Hu et al. (2018) which uses simulation to investigate the impact of proactive transfers to the ICU on the patients’ flow in the hospital and on the in-hospital mortality, LOS and ICU occupancy. While a substantial focus of Hu et al. (2018) is to rigorously estimate the causal effect of proactive transfers on individual patients, we focus on utilizing MDP approaches to derive theoretically justified transfer policies. Interestingly, Hu et al. (2018) conduct a sensitivity analysis by considering random deviations in the parameters of their model over the confidence intervals of the parameter estimates as an acknowledgement of the potential impact of parameter uncertainty. However, this approach does not capture a possible *adversarial* deviation of these parameters and, as we will show in this work, such deviations can substantially impact system performances.

Markov Decision Process in healthcare. In this work, we will leverage the methodology of Markov Decision Processes (MDP). This modeling framework has been used extensively in many healthcare applications including early detection, prevention, screening and treatment of diseases. MDPs are particularly efficient to analyze chronic diseases and decisions that are made sequentially over time in a stochastic environment. In particular, MDPs have been used, among others, for kidney transplantation (Alagoz et al. (2004)), HIV treatment recommendation (Shechter et al. (2008)), breast cancer detections (Ayer et al. (2012)), cardiovascular controls for patients with Type 2 diabetes (Steimle and Denton (2017)) and determining the optimal stopping time for medical treatment (Cheng et al. (2019)). We refer the reader to Alagoz et al. (2010), Schaefer et al. (2005) for reviews of applications of MDP to medical decision making.

Robust Markov Decision Process. In most medical applications, we only have access to observational data. Consequently, we can only obtain a *noisy estimate* of the true parameters of the MDP, and the decision-maker may recommend a treatment that performs poorly with respect to the true parameters. *Robust* MDPs address the issue of parameters misspecifications in the MDP (Satia and Lave (1973), Iyengar (2005), Nilim and El Ghaoui (2005), Xu and Mannor (2006), Delage and Mannor (2010), Wiesemann et al. (2013), and Mannor et al. (2016)). The goal is to compute an optimal *robust* policy, i.e., a policy that maximizes the worst-case expected outcome over the set of all plausible parameters. More specifically, the authors in Iyengar (2005), Nilim and El Ghaoui (2005) and Wiesemann et al. (2013) present algorithms to efficiently compute an optimal robust policy, provided that the parameters related to different state-action pairs are unrelated. Such *rectangular* uncertainty sets are quite conservative and do not allow relations across transition probabilities from different states. This is potentially very conservative especially if the transition probabilities depend on a common set of underlying factors, as could be the case in healthcare applications.

Robust Markov Decision Process in healthcare. In light of the limitations of the *rectangularity* assumption for modeling parameters uncertainty, the authors in Steimle et al. (2018) develop a

multi-model MDP approach and apply it to a case study of cholesterol management. However, computing the optimal robust policy of multi-model MDP is intractable in general. We use the model of *factor matrix uncertainty*, introduced in Goh et al. (2018) and further analyzed in Goyal and Grand-Clement (2018). In particular, the authors in Goh et al. (2018) use a model of uncertainty (later referred to as *factor matrix uncertainty set*) which captures transitions that are jointly varying in the set of all plausible parameters; prove that one can compute the worst-case reward for a given policy; and, then apply these methods to a cost-effectiveness analysis of fecal immunochemical testing for detecting colorectal cancer. Goyal and Grand-Clement (2018) show that for a factor matrix uncertainty set, one can also compute the optimal robust policy, i.e., the policy with the highest worst-case reward. Our work builds upon Goh et al. (2018) and Goyal and Grand-Clement (2018) by making use of a factor matrix uncertainty set in the specific setting of a Markov chain to model the patient’s trajectory in a hospital and capture different levels of connectivity in parameter deviations. Moreover, we make use of the theoretical tools developed in Goyal and Grand-Clement (2018), and more specifically of the *robust maximum principle* (see Section 4), in order to prove optimality of threshold policies in both the nominal MDP and the robust MDP setting, as well as to prove ordering on the thresholds of the optimal nominal policy and the optimal robust policy. While this paper builds upon the same model of uncertainty as in Goh et al. (2018), we are able to (i) compute an optimal robust policy where as Goh et al. (2018) only compares the worst-case performances of various policies, and (ii) give theoretical guarantees on the *structure* of the optimal policies, i.e., we prove that the optimal nominal and the optimal robust policies both have a threshold structure. Moreover, our work differs from Goyal and Grand-Clement (2018) in that we consider the specific case of an MDP modeling patient dynamics and the ICU admission decision. While we make use of the theoretical tools developed in Goyal and Grand-Clement (2018), we go further by providing the structure of the optimal policies and detailed numerical experiments (Section 5) on the impact of proactive policies on the hospital performances, which goes beyond our MDP model for the single-patient dynamics.

2. Hospital model and proactive transfer policies.

We formally present our discrete time hospital patient flow model. This model is similar to Hu et al. (2018) and is depicted in Figure 1. We consider a hospital with two levels of care, the Intensive Care Unit (ICU) and the general medical/surgical ward (ward). In order to focus on the management of the ICU, we assume that the ward has unlimited capacity while the ICU has a limited capacity $C < +\infty$.

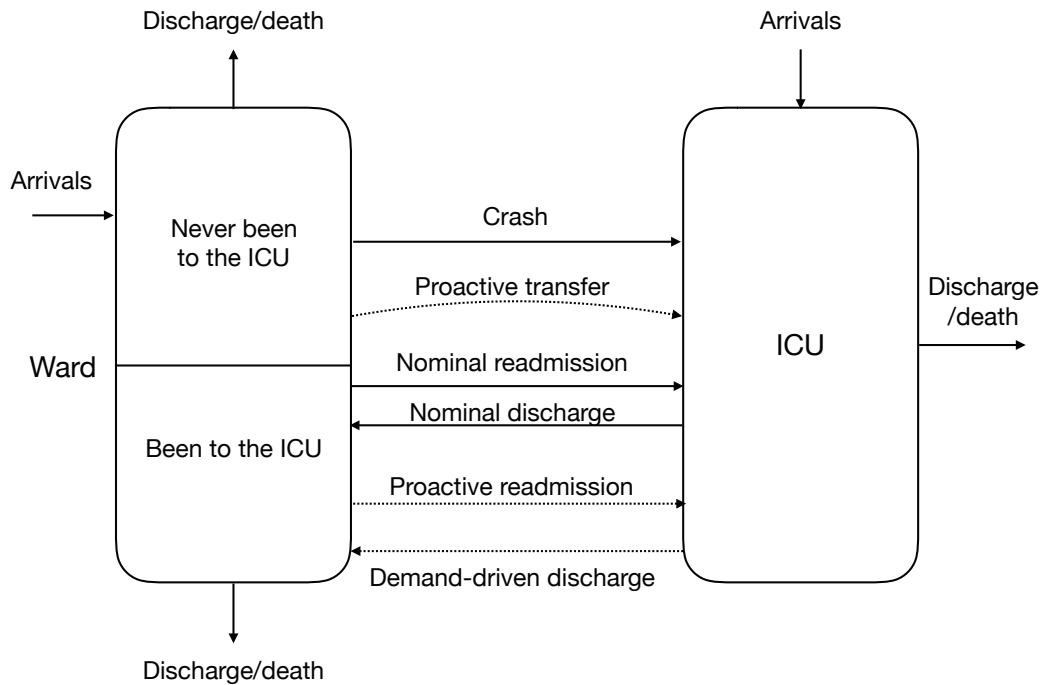


Figure 1 Simulation model for the hospital.

2.1. Model Dynamics

Ward patients: We start by describing the dynamics of the patients on the ward. These patients are divided into those who have already been to the ICU during their hospital stay and those who have not. The state of a patient who has never been to the ICU is captured by a severity score

$i \in \{1, \dots, n\}$, for a given number of severity scores $n \in \mathbb{N}$; each patient is assigned a severity score at arrival in the hospital, and this score is then updated in each time slot. We model the arrivals of patients with severity score i as a non-homogeneous Poisson process $\lambda_i(t)$. We model the evolution of the severity scores as a Markov Process with transition matrix $\mathbf{T}^0 \in \mathbb{R}^{n \times (n+3)}$. In each time slot, a patient whose current severity score is $i \in [n]$ transitions to another severity score in $j \in [n]$ with probability T_{ij}^0 . In addition, one of the three following events may happen at the end of each time slot:

- 1) With probability $T_{i,n+1}^0$, the patient may *crash* and require a *reactive* transfer to the ICU.
- 2) With probability $T_{i,n+2}^0$, the patient may fully recover and leave the hospital.
- 3) With probability $T_{i,n+3}^0$, the patient may die.

Reactive transfer to the ICU. If a patient crashes on the ward, s/he will be admitted to the ICU. Upon ICU admission, this patient's remaining hospital Length-Of-Stay (LOS) is modeled as being lognormally distributed with mean $1/\mu_C$ and standard deviation σ_C . We consider a model where a proportion p_W (following a distribution with density f_{p_W}) of this LOS is spent in the ICU, while the remaining proportion of time, $1 - p_W$, is spent in the ward. During this time, the patient may again require ICU admission, with a rate of ρ_C . Any patient still in the ward at the end of this LOS is assumed to die with probability d_C , or to fully recover and be discharged with probability $1 - d_C$. If there are no available beds in the ICU when a patient crashes, the ICU patient with the shortest remaining service time in the ICU is discharged back to the ward in order to accommodate the incoming patient; we refer to this event as a *demand-driven discharge*. A demand-driven discharged patient has an ICU readmission rate of ρ_D . The authors in Kc and Terwiesch (2012) suggest that ρ_D is higher than ρ_E , the readmission rate of patients who are naturally discharged from the ICU. Therefore, we set $\rho_D = 1.15 \cdot \rho_E$.

Proactive transfers to the ICU. If there are beds available in the ICU, a patient can be *proactively transferred* from the ward to the ICU. Such patients typically have better outcomes than those who crash and require a reactive ICU transfer (Hu et al. (2018)). If a patient with severity score $i \in [n]$

is proactively transferred, the LOS is modeled as being lognormally distributed with mean $1/\mu_{A,i}$ and standard deviation $\sigma_{A,i}$, while a proportion $p_W \sim f_{p_W}$ of this LOS is spent in the ICU. As in the case of reactive transfer, the patient will then survive to hospital discharge with probability $1 - d_{A,i}$. We assume that $1 - d_{A,i} \geq 1 - d_C$, i.e., the patient is more likely to survive if proactively transferred. If the patient is naturally discharged from the ICU, the readmission rate is $\rho_{A,i}$, otherwise it is ρ_D . We set $\rho_{A,i} = \rho_C$; note that the readmission rate for a demand-driven discharged patient is the same, regardless of whether the patient is a reactive or a proactive transfer.

Note that in practice when a patient reaches an alert threshold he may enter an evaluation state where further tests and examinations are required before an admission decision is made. In some instances, the patient may never be admitted to the ICU; e.g. if the alert is an error or the patient has a directive to not provide rescue measures. In order to focus on the impact of parameter uncertainty, our model assumes that ICU admission decisions are made right after the alert threshold is attained.

Direct admits to the ICU: In addition to reactive and proactive ICU transfers from the ward, patients can also be directly admitted to the ICU. We model the arrivals of these patients with a non-homogeneous Poisson process with rate $\lambda_E(t)$. Their LOS is lognormally distributed with mean $1/\mu_E$ and standard deviation σ_E , and a proportion p_E (following a distribution with density f_{p_E}) of this LOS is spent in the ICU, while the remaining time is spent in the ward. At the end of this LOS, the patient fully recovers and leaves the hospital with probability $1 - d_E$, or dies with probability d_E .

Details about the distribution laws of the different stochastic processes (arrivals in the ward and in the ICU, transition matrix across severity scores, distribution of LOS and mortality rate, etc.) involved in the hospital model of Figure 1 can be found in Appendix A.

2.2. Transfer policies

A transfer policy π is a decision rule that, for each patient in the ward, decides whether and when to proactively transfer the patient to the ICU (subject to bed availability). Our goal is to study the

impact of the proactive transfer policies on hospital performance as measured by the mortality rate, average LOS and average ICU occupancy. A particular class of simple and interpretable transfer policies is the class of *threshold* policies. A policy is said to be threshold if it transfers to the ICU all patients whose severity score are higher than a certain fixed threshold. Such proactive transfers are subject to bed availability in the ICU.

2.3. Challenges

The hospital model just described captures many salient features of real patient flows. Moreover, it is able to capture the core trade-off we are interested in studying between the benefits of proactive transfers for individual patients and needlessly utilizing expensive ICU resources. That said, the model also suffers from some limitations that we elaborate below.

Tractability. While our model could be described as a Markov Decision Process (MDP, Puterman (1994)), the state space is prohibitively large. For instance, with p patients in the ward and n severity scores, one needs a state space of cardinality n^p to describe the state of the ward. Thus, numerically solving this MDP is highly intractable.

Alternatively, one could take the approach in Hu et al. (2018) and use simulation. However, there are 2^n deterministic, state-independent, transfer policies (with n the number of severity scores). The number of state-dependent policies grows by the size of the ICU and/or the large ward state-space. Expanding to allow for randomized policies results in an uncountable number of potential transfer policies. Therefore, it is intractable to simply compare all the deterministic transfer policies using simulations¹.

Parameter uncertainty. The hospital model is specified by the stochastic processes detailed above, including a transition matrix $\mathbf{T}^0 \in \mathbb{R}_+^{n \times (n+3)}$. The coefficients of this matrix are estimated from historical data and consequently suffer from statistical estimation errors. The hospital performance could be highly sensitive to variations in coefficients of the transition matrix \mathbf{T}^0 and therefore,

¹For each transfer policy, computing the hospital performance (average mortality, LOS, ICU occupancy) takes a couple of hours on a laptop with 2.2 GHz Intel Core i7 with 8 GB of RAM.

optimizing the policies using the estimated transition matrix could lead to suboptimal policies. In particular, the optimal transfer policy for the hospital model with transition matrix \mathbf{T}^0 might lead to significantly suboptimal performance for the true, underlying transition matrix, even for small deviations in \mathbf{T}^0 . Additionally, the hospital model is itself an approximation of the true hospital dynamics and is therefore, subject to further misspecification errors.

Given these limitations of the model, we turn our attention to the development of insights using an approximation of the hospital dynamics.

3. A single-patient Markov model.

In light of the discussion in Section 2.3, we propose a tractable approximation of the full hospital model using an MDP that captures the health dynamics of a *single* patient.

3.1. Single-patient MDP.

State and Action spaces. We consider an MDP with $(n + 4)$ states. The set of states is

$$\mathbb{S} = \{1, \dots, n\} \cup \{n + 1 = CR, n + 2 = RL, n + 3 = D, n + 4 = PT\}.$$

The states $i \in \{1, \dots, n\}$ model the severity scores of the patient. There are 4 terminal states, CR, RL, D and PT . The state CR models the *crash* of a patient, its subsequent reactive transfer to the ICU, as well as the outcome when the crashed patient finally exits the hospital (i.e., fully recovering or dying). The state RL corresponds to *Recover and Leave*, the state D corresponds to in-hospital *Death*, and the state PT corresponds to a patient who has been *Proactively Transferred*, as well as the outcome when the patient finally exits the hospital (i.e., fully recovering or dying).

For each state $i = 1, \dots, n$, there are 2 possible actions, which model the decision of proactively transferring the patient (action 1) or not (action 0).

Figure 2 depicts the single-patient MDP.

Policies. A policy consists of a map $\pi : \mathbb{S} \rightarrow [0, 1]$, where for each severity score $i \in [n]$, $\pi(i) \in [0, 1]$ represents the probability of proactive transfer of a patient with current severity score i . For terminal states $i \in \{CR, RL, D, PT\}$, we set $\pi(i) = 0$.

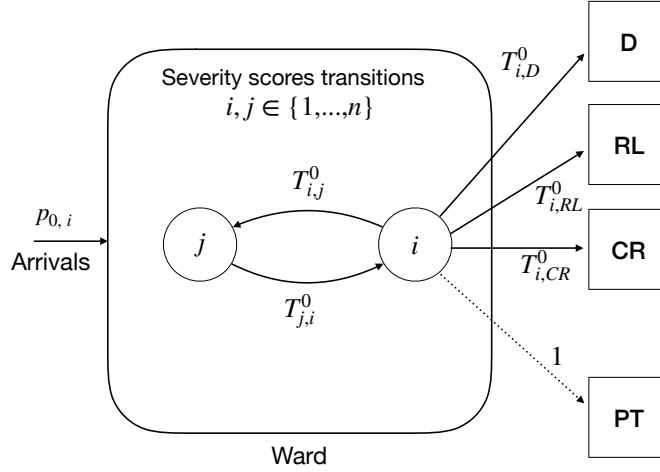


Figure 2 Single-patient MDP model. Terminal states are indicated as square. The patient arrives in the ward with a severity score of $i \in \{1, \dots, n\}$ with an initial probability $p_{0,i}$. The solid arcs correspond to transitions where no transfer decision is taken (action 0), and the patient can transition to another severity score j with probability $T_{i,j}^0$ or to the terminal states CR, RL or D . When the patient is in state i and the decision-maker takes the decision to proactively transfer the patient (action 1), the patient transitions with probability 1 to the terminal state PT (dashed arc).

A policy π is said to be of *threshold* type when $\pi(i) = 1 \Rightarrow \pi(i+1) = 1, \forall i \in [n-1]$. In other words, the policy proactively transfers patients at all severity scores above a given threshold. Therefore, the policy π can be described by an integer $\tau \in [1, n+1]$, such that all patients with severity score higher (or equal) than τ is transferred. Note that a threshold of $\tau = n+1$ means that no patient is proactively transferred, while a threshold of $\tau = 1$ means that all patients are proactively transferred. We write $\pi^{[\tau]}$ to denote the threshold policy parametrized by threshold τ . For any threshold policy π , we write its threshold $\tau(\pi)$.

Transitions. The states $i \in [n]$ model the n possible severity scores of a patient. When a patient is in state $i \in [n]$, the decision-maker can choose to proactively transfer the patient to the ICU (action 1) or not (action 0).

- If the patient is proactively transferred from $i \in [n]$, s/he transitions with probability 1 to the state PT . The state PT is a terminal state, the decision-maker receives a terminal reward.

- If the patient is not proactively transferred from state $i \in [n]$, the patient transitions to state $j \in [n]$ with probability T_{ij}^0 in the next 6 hours (where \mathbf{T}^0 is the transition matrix among severity scores). Alternatively, the patient can either transition to state CR with probability $T_{i,n+1}^0$, to state RL , with probability $T_{i,n+2}^0$ or the patient transitions to D with probability $T_{i,n+3}^0$. When the patient reaches one of the terminal states – CR, RL or D – s/he receives the associated terminal reward. Note that the patient exits the ward almost surely, assuming that $\theta = \min_{i \in [n]} \min\{T_{i,CR}^0, T_{i,RL}^0, T_{i,D}^0\} > 0$.

Rewards. The discount factor $\lambda \in (0, 1)$ captures the decreasing importance of future rewards compared to present rewards. The goal of the decision-maker is to pick a policy π that maximizes the expected discounted cumulated rewards, defined as

$$R(\pi, \mathbf{T}^0) = E^{\pi, \mathbf{T}^0} \left[\sum_{t=0}^{\infty} \lambda^t r_{i_t a_t} \right], \quad (3.1)$$

where $r_{i_t a_t}$ is the reward associated with visiting state i_t and choosing action a_t at time $t \in \mathbb{N}$.

For each policy π , we can associate a *value-vector* $\mathbf{V}^\pi \in \mathbb{R}^{n+4}$, defined as

$$V_i^\pi = E^{\pi, \mathbf{T}} \left[\sum_{t=0}^{\infty} \lambda^t r_{i_t a_t} \mid i_0 = i \right], \forall i \in \{1, \dots, n+4\}. \quad (3.2)$$

We want our MDP model to capture the trade-off between the benefits of proactive transfers for the patients' health and the costly use of resources and staff in the ICU. We achieve this by choosing the rewards in order to reflect the preference of the decision-maker who is balancing between improving patient outcomes by transferring them to the ICU proactively and the risk of such transfers resulting in a congested ICU.

Without loss of generality, we can consider that all rewards are non-negative. We consider a uniform reward across both actions for all states, i.e., $r_{i,0} = r_{i,1} = r_i, \forall i \in \mathbb{S}$. Therefore, the reward only depends of the current state, while the action dictates the likelihood of transitioning to states with different rewards. There is a reward of r_W associated with being in the ward: $r_i = r_W, \forall i \in [n]$.

If a patient is proactively transferred, s/he transitions to state PT with probability 1. In state PT , the patient either dies with probability d_A or recovers. Hence, the reward r_{PT} is

$$r_{PT} = d_A \cdot (r_{PT-D}) + (1 - d_A) \cdot (r_{PT-RL}),$$

where r_{PT-RL} (respectively, r_{PT-D}) corresponds to the rewards for a patient recovering (respectively, dying) after having been proactively transferred. The scalar d_A is the probability to die when having been proactively transferred and is calibrated to be the same as in the Markov model for the hospital in Section 2.

Similarly, there is a reward of r_{CR} associated with the state CR . A patient who crashes (and does not die immediately) will be transferred to the ICU before recovering or dying. We have that,

$$r_{CR} = d_C \cdot (r_{CR-D}) + (1 - d_C) \cdot (r_{CR-RL}),$$

where r_{CR-RL} (respectively, r_{CR-D}) corresponds to the rewards for a patient recovering (respectively, dying) after having been proactively transferred and d_C is the probability that a patient dies after crashing.

We would like to note that the rewards r_W, r_D, r_{RL}, r_{CR} and r_{PT} are a priori policy-dependent. For instance, for a policy that proactively transfers many patients, the reward r_{PT} should take into account the (a priori) detrimental increase in ICU occupancy. Moreover, estimating the exact values of the rewards can be challenging (see McClean and Millard (2006) for a solution for a two-state hospital model, and see Yauney and Shah (2018) for a Reinforcement Learning approach). Therefore, we focus on the relative *ordering* of these rewards, in order to capture the trade-off between better health outcomes by proactive transfers and increased congestion in ICU. For the same outcomes (e.g. mortality, LOS), the decision-maker favors the policy which uses the fewest ICU resources. Conditional on the patient recovering and leaving the hospital, it is natural to assume that $r_{RL} \geq r_{PT-RL} \geq r_{CR-RL}$. This is because leaving the hospital after recovering in the ward uses less ICU resources than recovering after being proactively transferred, which in turn uses less ICU resources than if the patient crashes (see Hu et al. (2018) for empirical evidence of

this relationship). For similar reasons, $r_D \geq r_{PT-D} \geq r_{CR-D}$. We assume that $r_{CR-RL} \geq r_D$, since the decision-maker wants to achieve a low in-hospital mortality rate.

Note that, as expected, this ordering of the rewards implies that $r_{RL} \geq r_{PT} \geq r_{CR}$.

In the rest of the paper, for any state $i \in [n]$ we define the *outside option* as

$$out(i) = r_{CR} \cdot T_{i,n+1}^0 + r_{RL} \cdot T_{i,n+2}^0 + r_D \cdot T_{i,n+3}^0.$$

The outside option $out(i)$ represents the expected one-step reward if a patient with severity score i is not proactively transferred and leaves the ward in the next period, i.e., if this patient transitions to one of the states in CR, RL , or D in the next period.

We make a first mild assumption, which has the following interpretation: the total cumulated reward is higher when the patient recovers and leaves after one period in the ward than if s/he stays in the ward forever.

ASSUMPTION 1. $\frac{r_W}{1-\lambda} \leq r_W + \lambda \cdot r_{RL}$.

This natural assumption implies the natural relationship that it is most desirable for a patient in the ward to recover and leave the hospital at the next period. This is stated formally in the following lemma:

LEMMA 1 (**Upper bound on the value vector**). *Let \mathbf{V}^π be the value vector of a policy π .*

Under Assumption 1, we have

$$V_i^\pi \leq r_W + \lambda \cdot r_{RL}, \forall i \in [n].$$

We present the proof in Appendix B.

We are ready to state the main result in this section, Theorem 1. Namely, under a mild assumption, the optimal nominal policy in our single-patient MDP is a threshold policy. In particular, we consider the following assumption.

ASSUMPTION 2. *We assume that*

$$out(i) \geq out(i+1), \forall i \in [n-1]. \tag{3.3}$$

and we assume that

$$\frac{r_W + \lambda \cdot r_{PT}}{r_W + \lambda \cdot r_{RL}} \geq \frac{\left(\sum_{j=1}^n T_{i+1,j}^0\right)}{\left(\sum_{j=1}^n T_{ij}^0\right)}, \forall i \in [n-1]. \quad (3.4)$$

Condition (3.3) implies that $out(i)$ is decreasing in the severity score i . This is meaningful since we expect that in practice, the severity score i captures the health condition of a patient, from $i = 1$ (as healthy as possible in the hospital) to $i = n$ (a very severe health condition). Therefore, it is reasonable to assume that the outside option of a patient is worse than the outside option of a patient with a better health condition (e.g. the healthier patient is more likely to leave the ward in a better state).

Condition (3.4) assumes that the chance of staying in the system in risk score i is non-increasing in severity score i . Additionally, the rate at which the chance of staying in the system decreases is higher than the ratio $\frac{r_W + \lambda \cdot r_{PT}}{r_W + \lambda \cdot r_{RL}}$, which captures the preference between the reward for proactively transferring a patient ($r_W + \lambda \cdot r_{PT}$) and an optimistic reward in the case that the patient is not transferred ($r_W + \lambda \cdot r_{RL}$). Similar to condition (3.3), we expect condition (3.4) to hold in practice, since patients with more severe health states are more likely to crash or die (and therefore exit the ward) than patients with better health conditions. In particular, it holds for our dataset of nearly 300,000 patients across 21 hospitals.

We would like to note that both condition (3.3) and condition (3.4) are homogeneous; they hold if we scale all rewards by the same (positive) scalar. Moreover, condition (3.4) is stable by translation, i.e., it holds if we add the same scalar to all rewards. However, this is not the case for condition (3.3) (see Lemma 2 in Appendix C).

3.2. Optimality of threshold policies.

We are now in a position to characterize structural properties of the optimal transfer policy for our single-patient MDP. Using standard arguments, without loss of generality we can restrict our attention to stationary deterministic policies. We show that there exists an optimal policy that is a threshold policy in the single-patient MDP. Formally, we have the following theorem.

THEOREM 1. *Under Assumptions 1 and 2, there exists a threshold policy that is optimal in the single-patient MDP.*

The proof relies on proving that all the policies generated by a Value Iteration algorithm (Section 6.3, Puterman (1994)) are threshold policies, provided that we initialize the algorithm with a threshold policy. Since Value Iteration is known to converge to an optimal policy of the MDP, we can conclude that there exists an optimal policy that is threshold. We present the detailed proof in Appendix D.

We give some intuition on why Assumptions 1 and 2 are sufficient to prove the existence of an optimal policy that is threshold. In order to show that a policy π is threshold, it is sufficient to show that for any state $i \in [n - 1]$,

$$\pi(i) = 1 \Rightarrow \pi(i + 1) = 1.$$

We note that the reward associated with a proactive transfer, $(r_W + \lambda \cdot r_{PT})$, is constant across the severity scores. However, for a patient in severity score $i \in [n - 1]$, the optimal policy is comparing the expected reward associated with a proactive transfer and the expected reward without proactive transfer, which decomposes into $out(i)$ (when the patient transfers to CR, D or RL in the next period) and the expected reward if the patient remains in the ward in the next period. Conditions (3.3) and (3.4) ensure that the reward for not proactively transferring a patient is non-increasing in the severity scores. Indeed, the outside option, i.e., the expected reward when exiting the ward, is non-increasing (condition (3.3)), while the probability to exit the ward is increasing (condition (3.4)). If the decision-maker chooses to proactively transfer a patient with severity score i , this means that the reward for proactive transfer, $r_W + \lambda \cdot r_{PT}$, is larger than the reward for not proactively transferring the patient. This last quantity is non-increasing, and therefore, for any severity score $j > i$, the optimal policy should also choose to proactively transfer the patient.

Several remarks are in order.

REMARK 1. We would like to note that Theorem 1 holds if we replace Assumption 2 by the following weaker, but less interpretable condition:

$$\left(\sum_{j=1}^n T_{ij}^0\right) \cdot (r_W + \lambda \cdot r_{PT}) + out(i) \geq \left(\sum_{j=1}^n T_{i+1,j}^0\right) \cdot (r_W + \lambda \cdot r_{RL}) + out(i+1), \forall i \in [n-1]. \quad (3.5)$$

Note that condition (3.5) is implied by Assumption 2, simply by summing up (3.3) and (3.4). Moreover, condition (3.5) is homogeneous and holds under translation of the rewards (see Lemma 3 in Appendix C). However, even though condition (3.5) is more general than Assumption 2, it is much less interpretable.

REMARK 2. We show in Appendix E that Assumption 2 is necessary for the existence of an optimal policy. We provide an instance of a single-patient MDP, whose parameters do not satisfy Assumption 2 where there is no optimal policy that is threshold. The instance has two severity scores $\{1, 2\}$, and $out(2) > out(1)$, which violates condition (3.3). By selecting parameters such that the outside option $out(2)$ is high, the optimal policy does not proactively transfer the patient in state 2. However, by setting a sufficiently small discount factor λ , the reward in state 1 becomes unrelated to the reward in state 2 and the optimal policy will proactively transfer a patient in severity score 1. We would like to note that condition (3.5) does not hold either for this instance.

REMARK 3. Recall that, by assumption, the rewards in the ward are uniform, i.e., $r_i = r_W, \forall i \in [n]$. Also, the reward for proactively transferring a patient from severity score $i \in [n]$ does not depend of the severity score. We can relax these assumptions and consider a model where the rewards in the ward and the rewards for proactive transfers are heterogeneous across different severity scores and still establish the optimality of a threshold policy under assumptions which are generalizations of Assumptions 1 and 2. We present the assumptions and the proof of optimality of threshold policies in Appendix F.

4. Robustness analysis.

Thus far, we have assumed the health state of a patient evolves according to a Markov chain with known transition matrix \mathbf{T}^0 . However, the parameters of the transition matrix are estimated

from a historical data and are subject to statistical estimation errors. At best, we have a noisy estimate of \mathbf{T}^0 . Therefore, an important practical consideration is to develop an understanding of the impact of small deviations in the hospital parameters. We start by focusing on the robustness of the optimal policy for the single-patient MDP.

4.1. Robust MDP and model of uncertainty set.

In a classical MDP, it is assumed that the transition kernel \mathbf{T}^0 is known so that one finds a policy that maximizes the expected nominal reward. That is, one solves the optimization problem:

$$\max_{\pi \in \Pi} R(\pi, \mathbf{T}^0).$$

In Theorem 1, we show that the optimal nominal policy for the matrix \mathbf{T}^0 is a threshold policy, under certain fairly reasonable assumptions. In order to tackle model misspecification, we consider a robust MDP framework, where the true transition matrix is unknown. We model the uncertainty as adversarial deviations from the nominal matrix in some *uncertainty set* \mathcal{U} , that can be interpreted as a safety region. The goal is to compute a transfer policy that maximizes the worst-case reward over the set of all possible transition matrices \mathcal{U} , i.e., our goal is to solve:

$$\max_{\pi \in \Pi} \min_{\mathbf{T} \in \mathcal{U}} R(\pi, \mathbf{T}). \quad (4.1)$$

A solution π^{rob} to the optimization problem (4.1) will be called an *optimal robust policy*.

The choice of uncertainty set is important and dictates the conservatism and usefulness of the model. In this paper we consider a *factor matrix uncertainty set* (Goh et al. (2018), Goyal and Grand-Clement (2018)). In this model we consider that the transition probabilities are convex combination of some factors, which can themselves be uncertain. Such a model allows us to capture correlations across transitions probabilities, unlike rectangular uncertainty sets (Iyengar (2005), Wiesemann et al. (2013)) that allow unrelated adversarial deviations. In particular, we assume that \mathcal{U} is of the following form:

$$\mathcal{U} = \left\{ \mathbf{T} \in \mathbb{R}_+^{n \times (n+3)} \mid T_{ij} = \sum_{\ell=1}^r u_i^\ell w_{\ell,j} \forall (i, j) \in [n] \times [n+3], \mathbf{w}_\ell \in \mathcal{W}^\ell, \forall \ell \in [r] \right\} \quad (4.2)$$

where $\mathbf{u}_1, \dots, \mathbf{u}_n$ are fixed vectors in \mathbb{R}_+^r and $\mathcal{W}^\ell, \ell = 1, \dots, r$ are convex, compact subsets of $\mathbb{R}_+^{(n+3) \times r}$ such that:

$$\sum_{\ell=1}^r u_i^\ell = 1, \forall i \in [n],$$

$$\sum_{j=1}^{n+3} w_{\ell,j} = 1, \forall \ell \in [r], \forall \mathbf{w}_\ell \in \mathcal{W}^\ell.$$

To understand better the implications of this uncertainty set, consider \mathbf{T} , a transition matrix in our factor matrix uncertainty set \mathcal{U} . Each of the rows of the matrix \mathbf{T} is a convex combination of the factors $\mathbf{w}_1^\top, \dots, \mathbf{w}_r^\top \in \mathbb{R}^{1 \times (n+3)}$. Therefore, for $\mathbf{U} = (u_i^\ell)_{(i,\ell)} \in \mathbb{R}^{n \times r}$ and $\mathbf{W} = (\mathbf{w}_1, \dots, \mathbf{w}_r) \in \mathbb{R}^{(n+3) \times r}$, we have $\mathbf{T} = \mathbf{U}\mathbf{W}^\top$. This model of uncertainty set is very general, and covers the case of (s, a) -rectangular uncertainty sets (Iyengar (2005)), i.e., the case where all the rows of the transition matrix \mathbf{T} are chosen independently, when $r = n$ (where n is the number of rows of \mathbf{T}). This model is able to incorporate relationships in the transitions from various severity scores which may arise since the health dynamics of a patient is likely to be influenced by some underlying common factors, such as demographics, past medical histories, current blood sugar level, etc. As different rows of a matrix $\mathbf{T} \in \mathcal{U}$ are convex combinations of the same r factor vectors, we can use our factor matrix uncertainty set to model correlations between the probability distributions related to different states when r is smaller than n .

We assume that the nominal matrix \mathbf{T}^0 satisfies Assumption 2. Our uncertainty set \mathcal{U} models small parameters variations from \mathbf{T}^0 . Therefore, it is reasonable to assume that all the matrices in \mathcal{U} will satisfy Assumption 2. In particular, we assume:

ASSUMPTION 3. *Every matrix \mathbf{T} in \mathcal{U} satisfies Assumption 2.*

Under this model of uncertainty, the authors in Goh et al. (2018) prove that one can compute the worst-case reward of a given policy. Goyal and Grand-Clement (2018) prove that an optimal robust policy can be chosen to be deterministic and provide an efficient algorithm to compute an optimal robust policy. Moreover, they show that for \mathcal{U} as in (4.2), the *robust maximum principle*

holds. As this is a key theorem that will be essential for our analysis, we state it formally for completeness.

Let $\mathbf{v}^{\pi, \mathbf{T}}$ be the value vector of the decision-maker when s/he chooses policy π and the adversary chooses factor matrix $\mathbf{T} = \mathbf{U}\mathbf{W}^\top$, defined by the Bellman Equation:

$$v_i^{\pi, \mathbf{T}} = r_{\pi, i} + \lambda \cdot (1 - \pi(i)) \cdot \sum_{\ell=1}^r u_i^\ell \mathbf{w}_\ell^\top \mathbf{v}^{\pi, \mathbf{W}} + \lambda \cdot \pi(i) \cdot r_{PT}, \forall i \in [n].$$

For any state $i \in [n]$, the scalar $v_i^{\pi, \mathbf{T}}$ represents the infinite horizon discounted expected reward, starting from state i .

THEOREM 2 (Robust Maximum Principle (Goyal and Grand-Clement 2018)). *Let \mathcal{U} be a factor matrix uncertainty set as in (4.2).*

1. Let $\hat{\mathbf{T}} \in \mathcal{U}$ and

$$\hat{\pi} \in \arg \max_{\pi \in \Pi} R(\pi, \hat{\mathbf{T}}).$$

Then

$$v_i^{\pi, \hat{\mathbf{T}}} \leq v_i^{\hat{\pi}, \hat{\mathbf{T}}}, \forall \pi \in \Pi, \forall i \in [n]. \quad (4.3)$$

2. Let $\hat{\pi}$ be a policy and

$$\hat{\mathbf{T}} \in \arg \min_{\mathbf{T} \in \mathcal{U}} R(\hat{\pi}, \mathbf{T}).$$

Then

$$v_i^{\hat{\pi}, \hat{\mathbf{T}}} \leq v_i^{\hat{\pi}, \mathbf{T}}, \forall \mathbf{T} \in \mathcal{U}, \forall i \in [n]. \quad (4.4)$$

3. Let

$$(\pi^*, \mathbf{T}^*) \in \arg \max_{\pi \in \Pi} \min_{\mathbf{T} \in \mathcal{U}} R(\pi, \mathbf{T}).$$

For all policy $\hat{\pi}$, for all transition matrix $\hat{\mathbf{T}} \in \arg \min_{\mathbf{T} \in \mathcal{U}} R(\hat{\pi}, \mathbf{T})$,

$$v_i^{\hat{\pi}, \hat{\mathbf{T}}} \leq v_i^{\pi^*, \mathbf{T}^*}, \forall i \in [n]. \quad (4.5)$$

Inequality (4.3) implies that in a classical MDP setting, the value vector of the optimal nominal policy is component-wise higher than the value vector of any other policy. Therefore, for any state,

the nominal expected reward obtained when the decision-maker follows the optimal nominal policy is higher than the nominal expected reward obtained when the decision-maker follows any other policy. Following Inequality (4.4), when a policy is fixed but the transition matrix varies in the uncertainty set \mathcal{U} , the worst-case value vector of the policy is component-wise lower than the value vector of the policy for any other transition matrix. Finally, when we consider an optimal robust policy, Inequality (4.5) implies that the worst-case value vector of the optimal robust policy is component-wise higher than the worst-case value vector of any other policy. Therefore, the optimal robust policy is maximizing the worst-case expected reward *starting from any state*.

4.2. Theoretical guarantees.

We show that under Assumption 1 and Assumption 3, the optimal robust policy is a threshold policy. Moreover, we show that the threshold of the optimal robust policy is smaller than the threshold of the optimal nominal policy. Therefore, the optimal robust policy is more aggressive in proactively transferring the patients. In particular, we start with the following theorem.

THEOREM 3. *Under Assumptions 1 and 3, there exists an optimal robust policy that is a threshold policy.*

Proof. Following Assumption 3 and Theorem 1, for any transfer policy $\tilde{\pi} \in \Pi$,

$$\exists \tilde{\mathbf{T}} \in \mathcal{U}, \tilde{\pi} \in \arg \max_{\pi \in \Pi} R(\pi, \tilde{\mathbf{T}}) \Rightarrow \tilde{\pi} \text{ is a threshold policy.}$$

Theorem 4.2 in Goyal and Grand-Clement (2018) shows that

$$(\pi^{\text{rob}}, \mathbf{T}^{\text{rob}}) \in \arg \max_{\pi \in \Pi} \min_{\mathbf{T} \in \mathcal{U}} R(\pi, \mathbf{T}) \iff \pi^{\text{rob}} \in \arg \max_{\pi \in \Pi} R(\pi, \mathbf{T}^{\text{rob}}).$$

Since the matrix \mathbf{T}^{rob} belongs to the uncertainty set \mathcal{U} , it satisfies Assumption 2 by Assumption 3 and therefore the optimal robust policy π^{rob} is threshold.

□

This result highlights the critical role of threshold policies. Not only is the optimal nominal policy a threshold policy (Theorem 1), but the optimal robust policy, i.e., the policy with the highest worst-case reward, is also a threshold policy. It is natural to compare the thresholds of the optimal nominal and the optimal robust policies. Our next result states that the threshold of the optimal robust policy π^{rob} is always *lower* than the threshold of the optimal nominal policy π^{nom} .

THEOREM 4. *Under Assumptions 1 and 3, we have $\tau(\pi^{\text{rob}}) \leq \tau(\pi^{\text{nom}})$, where π^{rob} is the optimal robust policy and π^{nom} is the optimal nominal policy.*

Proof. Let $\hat{\Pi}$ denote the set of policies that are optimal for some transition kernel in \mathcal{U} :

$$\hat{\Pi} = \{\pi \mid \exists \mathbf{T} \in \mathcal{U}, \pi \in \arg \max_{\pi \in \Pi} R(\pi, \mathbf{T})\}.$$

Note that $\pi^{\text{nom}} \in \hat{\Pi}$. We will prove that

$$\tau(\pi^{\text{rob}}) \leq \tau(\pi), \forall \pi \in \hat{\Pi}.$$

Following Theorem 3, we can pick π^{rob} to be an optimal robust policy that is a threshold policy.

We denote \mathbf{T}^{rob} a matrix in \mathcal{U} such that

$$(\pi^{\text{rob}}, \mathbf{T}^{\text{rob}}) \in \max_{\pi \in \Pi} \min_{\mathbf{T} \in \mathcal{U}} R(\pi, \mathbf{T}).$$

Let $\hat{\pi} \in \hat{\Pi}$. There exists a transition matrix $\hat{\mathbf{T}} \in \mathcal{U}$ such that

$$\hat{\pi} \in \arg \max_{\pi \in \Pi} R(\pi, \hat{\mathbf{T}}).$$

Let us assume that $\hat{\pi}(i) = 1$ for some $i \in [n]$. We will prove that $\pi^{\text{rob}}(i) = 1$. We have

$$r_W + \lambda \cdot r_{PT} > r_W + \lambda \cdot \hat{\mathbf{T}}_{i,\cdot}^\top \mathbf{V}^{\hat{\pi}, \hat{\mathbf{T}}} \tag{4.6}$$

$$\geq r_W + \lambda \cdot \hat{\mathbf{T}}_{i,\cdot}^\top \mathbf{V}^{\pi^{\text{rob}}, \hat{\mathbf{T}}}, \tag{4.7}$$

where Inequality (4.6) follows from the Bellman Equation for the MDP with transition matrix $\hat{\mathbf{T}}$ and Inequality (4.7) follows from Inequality (4.3) of Theorem 2:

$$\hat{\pi} \in \arg \max_{\pi \in \Pi} R(\pi, \hat{\mathbf{T}}) \Rightarrow V_j^{\hat{\pi}, \hat{\mathbf{T}}} \geq V_j^{\pi, \hat{\mathbf{T}}}, \forall j \in [n], \forall \pi \in \Pi.$$

Now for the sake of contradiction let us assume that $\pi^{\text{rob}}(i) = 0$. Therefore,

$$r_W + \lambda \cdot \hat{\mathbf{T}}_{i,\cdot}^\top \mathbf{V}^{\pi^{\text{rob}}, \hat{\mathbf{T}}} = V_i^{\pi^{\text{rob}}, \hat{\mathbf{T}}}. \quad (4.8)$$

Therefore, if $\pi^{\text{rob}}(i) = 0$, we can conclude that

$$r_W + \lambda \cdot r_{PT} > V_i^{\pi^{\text{rob}}, \hat{\mathbf{T}}} \quad (4.9)$$

$$\geq V_i^{\pi^{\text{rob}}, \mathbf{T}^{\text{rob}}}, \quad (4.10)$$

where the strict Inequality (4.9) follows from (4.8) and (4.6), and Inequality (4.10) follows from (4.4) in the robust maximum principle:

$$\mathbf{T}^{\text{rob}} \in \arg \min_{\mathbf{T} \in \mathcal{U}} R(\pi^{\text{rob}}, \mathbf{T}) \Rightarrow V_j^{\pi^{\text{rob}}, \hat{\mathbf{T}}} \geq V_j^{\pi^{\text{rob}}, \mathbf{T}^{\text{rob}}}, \forall j \in [n].$$

We can therefore conclude that

$$r_W + \lambda \cdot r_{PT} > V_i^{\pi^{\text{rob}}, \mathbf{T}^{\text{rob}}}. \quad (4.11)$$

But since π^{rob} is an optimal robust policy, we know following Theorem 2 that

$$\pi^{\text{rob}} \in \arg \max_{\pi \in \Pi} R(\pi, \mathbf{T}^{\text{rob}}).$$

Therefore, from the Bellman Equation we know that $\pi^{\text{rob}}(i) = 1$ if $r_W + \lambda \cdot r_{PT} > r_W + \lambda \cdot \mathbf{T}_{i,\cdot}^{\text{rob} \top} \mathbf{V}^{\pi^{\text{rob}}, \mathbf{T}^{\text{rob}}}$ and $\pi^{\text{rob}}(i) = 0$ if $r_W + \lambda \cdot r_{PT} \leq r_W + \lambda \cdot \mathbf{T}_{i,\cdot}^{\text{rob} \top} \mathbf{V}^{\pi^{\text{rob}}, \mathbf{T}^{\text{rob}}}$. This implies that

$$V_i^{\pi^{\text{rob}}, \mathbf{T}^{\text{rob}}} \geq r_W + \lambda \cdot r_{PT},$$

which contradicts Inequality (4.11), and therefore it is impossible that $\pi^{\text{rob}}(i) = 0$. Since π^{rob} is a deterministic policy,

$$\pi^{\text{rob}}(i) \neq 0 \Rightarrow \pi^{\text{rob}}(i) = 1.$$

We have proved that if $\hat{\pi}(i) = 1$ for some $\hat{\pi}$ in $\hat{\Pi}$ and some $i \in [n]$, then $\pi^{\text{rob}}(i) = 1$. Therefore, we can conclude that

$$\tau(\pi^{\text{rob}}) \leq \tau(\pi), \forall \pi \in \hat{\Pi}.$$

Since π^{nom} is the optimal policy for the nominal transition kernel \mathbf{T}^0 , we can conclude that $\pi^{\text{nom}} \in \hat{\Pi}$ and therefore in particular

$$\tau(\pi^{\text{rob}}) \leq \tau(\pi^{\text{nom}}).$$

□

Theorem 4 highlights the crucial role of threshold policies in ICU admission decision-making. In the framework of our single-MDP for modeling the patient dynamics, both an optimal nominal policy and an optimal robust policy can be found in this class of simple and implementable policies. Moreover, there exists a natural ordering on the threshold of a nominal policy and a policy that accounts for parameter misspecification. In particular, the robust optimal policy is more aggressive in proactively transferring patients.

5. Numerical experiments.

In this section, we utilize real data from 21 Northern California Kaiser Permanente hospitals to examine the potential implications of our theoretical results in practice. We utilize this data to estimate the nominal parameters and uncertainty set of our hospital model (Figure 1) and our single-patient MDP (Figure 2). We then compare the performances of optimal nominal and optimal robust policies on several metrics of interest: mortality, Length-Of-Stay (LOS) and average ICU occupancy.

5.1. Dataset.

Our dataset consists of 296,381 unique patient hospitalizations at one of 21 Northern California Kaiser Permanente hospitals. For each hospitalization, we have patient-level data which is assigned at the time of hospital admission: age, gender, admitting hospital, admitting diagnosis, classification of diseases codes, and three scores that quantify the severity of the illness of the patient (CHMR, COPS2, LAPS2, see Hu et al. (2018) for more details). During the patient’s hospitalization, we can track each unit (i.e., ICU, Transitional Care Unit, general medical-surgical ward, operating room, or post-anesthesia care unit) the patient stayed in and when. Additionally, we have a sequence

of early warning scores, known as Advance Alert Monitor (AAM) scores, that are updated every six hours. This early warning score uses the LAPS2, COPS2, individual vital signs and laboratory tests, interaction terms, temporal markers, and location indicators to estimate the probability of in-hospital deterioration (requiring ICU transfer or leading to death on the ward) within the next 12 hours, with an alert issued at a probability of $\geq 8\%$ (Kipnis et al. 2016). Similar to Hu et al. (2018), we focus on medical patients who were admitted to the hospital through the emergency department (this comprises more than 60% of all patients). Our final dataset consists of 174,632 hospitalizations, each corresponding to a patient trajectory that evolves across $n = 10$ severity scores, possible ICU visit(s), and terminates with the patient either recovering and leaving the hospital or dying and leaving the hospital. In order to have a more accurate estimate for the dynamics of the most severe patients, the number of patients in each severity group is nonuniform. Summary statistics (partition, mortality rate, average length-of-stay, etc.) for the patients across the different severity scores is given in Appendix A.

5.2. Transition matrix and model of uncertainty.

We first calibrate the transition matrix which determines the evolution of patient severity score while in the general ward.

5.2.1. Nominal transition matrix. We use the AAM scores as our severity scores. The matrix \mathbf{T}^0 has dimension $n \times (n+3)$ where $n = 10$ is the number of severity scores. \mathbf{T}^0 is constructed as follows. Let $i \in [n]$ and $j \in [n] \cup \{CR, RL, D\}$. The coefficient T_{ij}^0 represents the probability that a patient in severity score i will transfer to state j in the next period. We use the empirical mean as the nominal value for T_{ij}^0 . We use the method in Sison and Glaz (1995) to obtain the 95%-confidence intervals for the matrix \mathbf{T}^0 :

$$[T_{ij}^0 - \alpha_i, T_{ij}^0 + 2 \cdot \alpha_i], \forall (i, j) \in [n] \times [n+3]. \quad (5.1)$$

This expression highlights the skewness of the confidence intervals, which follows from the skewness of the multinomial distribution. Also, note that for a given severity score $i \in [n]$, the parameter uncertainty in T_{ij}^0 is uniform across all $j \in [n+3]$. See Appendix G for the values for $\alpha_1, \dots, \alpha_n$, which are in the order of 10^{-4} to 10^{-3} .

5.2.2. Nominal factor matrix. In order to construct a factor matrix uncertainty set (4.2), one needs to compute the coefficients $(u_i^\ell)_{(i,\ell)} \in \mathbb{R}^{n \times r}$, the nominal factors $\mathbf{W} = (\mathbf{w}_1, \dots, \mathbf{w}_r) \in \mathbb{R}_+^{(n+3) \times r}$ such that $\mathbf{T}^0 \approx \mathbf{U}\mathbf{W}^\top$, and the confidence regions \mathcal{W}^i for each factor $\mathbf{w}_i, i = 1, \dots, r$. To do this, we solve the following Nonnegative Matrix Factorization (NMF) problem:

$$\begin{aligned} \min \|\mathbf{T}^0 - \mathbf{U}\mathbf{W}^\top\|_2^2 \\ \mathbf{U}\mathbf{e}_r = \mathbf{e}_n, \\ \mathbf{e}_{n+3}^\top \mathbf{W} = \mathbf{e}_r, \\ \mathbf{U} \in \mathbb{R}_+^{n \times r}, \mathbf{W} \in \mathbb{R}_+^{(n+3) \times r}. \end{aligned} \tag{5.2}$$

Although (5.2) is a non-convex optimization problem, there exist fast algorithms for efficiently computing local minima. We adapt the block-coordinate descent method of Xu and Yin (2013), starting from 10^6 different random matrices and keep the best solution². For $r = 8$, our solution $\hat{\mathbf{T}} = \mathbf{U}\hat{\mathbf{W}}^\top$ achieves the following errors: $\|\mathbf{T}^0 - \hat{\mathbf{T}}\|_1 = 0.0811$, $\|\mathbf{T}^0 - \hat{\mathbf{T}}\|_\infty = 0.0074$, $\|\mathbf{T}^0 - \hat{\mathbf{T}}\|_{\text{relat}, \mathbf{T}^0} = 0.3385$, where $\|\cdot\|_{\text{relat}, \mathbf{T}^0}$ stands for the maximum relative deviation from a parameter of \mathbf{T}^0 :

$$\|\mathbf{T}^0 - \hat{\mathbf{T}}\|_{\text{relat}, \mathbf{T}^0} = \max_{(i,j) \in [n] \times [n+3]} \frac{|T_{ij}^0 - \hat{T}_{ij}|}{T_{ij}^0}.$$

Table 1 summarizes the errors across the $n \times (n + 3) = 130$ elements of \mathbf{T}^0 .

	max.	mean	median	95% percentile
absolute deviation	0.0074	0.0006	0.0003	0.0022
relative deviation	0.3385	0.0565	0.0204	0.2656

Table 1 Statistics of the absolute and relative deviations of \hat{T}_{ij} from T_{ij}^0 , for all $(i, j) \in [n] \times [n + 3]$.

The maximum *absolute* deviation between T_{ij}^0 and \hat{T}_{ij} is less than 0.01 (0.0891 instead of 0.0817). Moreover, the maximum relative deviation is about 34%, with a coefficient of $4.8527 \cdot 10^{-4}$ instead of $7.34 \cdot 10^{-4}$ for \mathbf{T}^0 . This occurs for $T_{3,9}^0$, which represents a sudden, dramatic, and relatively rare health deterioration from state 3 to state 9.

² This takes less than 5 minutes on a laptop with 2.2 GHz Intel Core i7 and 8 GB of RAM.

By construction, any two rows $\hat{\mathbf{T}}_{i_1,\cdot}$ and $\hat{\mathbf{T}}_{i_2,\cdot}$ are convex combinations of the same factors $\hat{\mathbf{w}}_1, \dots, \hat{\mathbf{w}}_r$, with coefficients $\mathbf{U}_{i_1,\cdot}$ and $\mathbf{U}_{i_2,\cdot}$. We compute $\Delta(i_1, i_2) = \|\mathbf{U}_{i_1,\cdot} - \mathbf{U}_{i_2,\cdot}\|_1$ as a measure of relatedness between the uncertainty on $\mathbf{T}_{i_1,\cdot}$ and $\mathbf{T}_{i_2,\cdot}$. We note that our NMF decomposition captures the intuition that close severity scores are more related than very different severity scores. To see this, consider any severity score $i \in [n]$ and any two alternative severity scores $(j, k) \in [n] \times [n]$ which are different from i . Then, we observe that $|i - j| < |i - k| \Rightarrow \Delta(i, j) < \Delta(i, k)$.

Errors related to the confidence intervals. The intent of the structured nominal transition matrix (and the subsequent uncertainty sets) is to capture the parameter uncertainty inherent in the estimation process from real data. As such, it is of interest to understand whether our nominal matrix is consistent with the confidence intervals of our parameter estimates. To do this, we consider a relative error of $\hat{\mathbf{T}}$ compared to \mathbf{T}^0 , measured in terms of the confidence bounds $\alpha_1, \dots, \alpha_n$. That is, we compute the ratios

$$ratio_{(i,j)} = \frac{T_{ij}^0 - \hat{T}_{ij}}{\alpha_i}, \forall (i, j) \in [n] \times [n+3]. \quad (5.3)$$

For $r = 8$, we find that all coefficients are in the confidence intervals as defined by (5.1), i.e., $ratio_{(i,j)} \in [-1, 2], \forall (i, j) \in [n] \times [n+3]$. In particular, the mean over (i, j) of the absolute values of the ratios (5.3) is 0.2345, with a median of 0.1579. Moreover, 95% of these absolute values are below 0.6729. Thus, we conclude that $\hat{\mathbf{T}}$ (our NMF solution of rank 8) is a plausible approximation for \mathbf{T}^0 .

For completeness, we also computed the solutions to the optimization problem (5.2) for lower ranked matrices: $r = 5, 6, 7$. While the errors in the 1- and ∞ -norms remain small, the relative errors increased substantially, up to 0.43 for $r = 7$, 0.98 for $r = 6$ and 5.8 for $r = 5$. For rank 7, we have 10 coefficients outside of the confidence intervals, with a maximum deviation of $4.840 \cdot \alpha_9$ for $T_{9,9}^0$. Despite that one coefficient being well out of its confidence interval, we find that the mean of the absolute value of the ratios is 0.4818, with a median at 0.2380. Therefore, the NMF solution for $r = 7$ appears to be a reasonable approximation for \mathbf{T}^0 . However, it does not seem reasonable to decrease the rank even further. For instance, for a rank $r = 5$, our NMF solution has 70 coefficients

that are outside the 95% confidence intervals, with a maximum ratio of 48.4931. Similarly, for $r = 6$, there are still 54 coefficients outside of the confidence intervals, with a maximum ratio of 44.0267. Therefore, in the rest of the paper we will focus on NMF solutions corresponding to rank $r = 8$. Detailed experiments for $r = 7$ are provided in Appendix I.

5.2.3. Choice of uncertainty sets. Recall that in our main analysis, we considered a factor matrix uncertainty set. That is, the coefficient matrix \mathbf{U} is fixed and the factors $\mathbf{w}_1, \dots, \mathbf{w}_r$ belong to $\mathcal{W}^1 \times \dots \times \mathcal{W}^r$. We consider several uncertainty sets:

- \mathcal{U}_{\min} : We construct a factor matrix uncertainty set based on the 95% confidence interval in the most optimistic manner. Specifically, for $\alpha_{\min} = \min_{i \in [n]} \alpha_i$, we consider $\mathcal{U}_{\min} = \{\mathbf{T} \mid \mathbf{T} = \mathbf{U}\mathbf{W}^\top, \mathbf{W} \in \mathcal{W}_{\min}\}$, where $\mathcal{W}_{\min} = \mathcal{W}_{\min}^1 \times \dots \times \mathcal{W}_{\min}^r$,

$$\mathcal{W}_{\min}^\ell = \{\mathbf{w}_\ell \mid \forall j \in [n+3], w_{\ell,j} - \hat{w}_{\ell,j} \in [-\alpha_{\min}, +2 \cdot \alpha_{\min}], \mathbf{w}_\ell \geq \mathbf{0}, \mathbf{w}_\ell^\top \mathbf{e}_{n+3} = 1\}, \forall \ell \in [r].$$

Thus, the deviation on each component of the factor vectors must be within $[-\alpha_{\min}, +2 \cdot \alpha_{\min}]$, which in turn implies that $\mathbf{T} = \mathbf{U}\mathbf{W}^\top$ is within $[-\alpha_{\min}, +2 \cdot \alpha_{\min}]$ from the matrix $\hat{\mathbf{T}}$ (in $\|\cdot\|_\infty$).

- \mathcal{U}_{emp} : We also consider another possibly less restrictive uncertainty set that is constructed empirically from the 95% confidence intervals. To do this, we generate 95% confidence intervals of the *factor vectors*. First, we sample q transition matrices $\mathbf{T}^1, \dots, \mathbf{T}^q$ uniformly in the 95% confidence intervals around \mathbf{T}^0 , for $q = 10^4$. For each sampled matrix, we use Nonnegative Matrix Factorization to compute factor vectors $\mathbf{W}^1, \dots, \mathbf{W}^q$ such that $\mathbf{T}^m \approx \mathbf{U}\mathbf{W}^{m\top}$, $m = 1, \dots, q$. Let σ_j^ℓ be the empirical standard deviations of each coefficients w_j^ℓ , for $(\ell, j) \in [r] \times [n+3]$, from the resulting $\mathbf{W}^1, \dots, \mathbf{W}^q$. We then define the uncertainty set $\mathcal{U}_{\text{emp}} = \{\mathbf{T} \mid \mathbf{T} = \mathbf{U}\mathbf{W}^\top, \mathbf{W} \in \mathcal{W}_{\text{emp}}\}$, where $\mathcal{W}_{\text{emp}} = \mathcal{W}_{\text{emp}}^1 \times \dots \times \mathcal{W}_{\text{emp}}^r$ represents the bootstrapped 95% confidence intervals for the factor vectors. That is:

$$\mathcal{W}_{\text{emp}}^\ell = \{\mathbf{w}_\ell \mid \forall j \in [n+3], |w_{\ell,j} - \hat{w}_{\ell,j}| \leq \sigma_j^\ell \cdot \frac{1.96}{\sqrt{q}}, \mathbf{w}_\ell \geq \mathbf{0}, \mathbf{w}_\ell^\top \mathbf{e}_{n+3} = 1\}, \forall \ell \in [r].$$

- \mathcal{U}_{sa} : Finally, for completeness, we also consider the following (s, a) -rectangular uncertainty set, where transitions from each state can be chosen unrelated to the transitions out of any other states:

$$\mathcal{U}_{\text{sa}} = \{\mathbf{T} \mid T_{ij} - T_{ij}^0 \in [-\alpha_i, +2 \cdot \alpha_i], \sum_{j=1}^{13} T_{ij} = 1, \forall i \in [n]\}.$$

This uncertainty set is unable to capture the fact that there are likely characteristics which introduce correlations across different states.

5.3. Robustness analysis for the single-patient MDP.

We first give the details about the parameters of our single-patient MDP.

5.3.1. Choice of MDP Parameters.

Nominal transition matrix. The probability that the patient transitions from severity score $i \in [n]$ to next state $j \in [n+3]$ is T_{ij}^0 . The probability that a patient dies after having crashed in the ward is given by $d_C = 0.4761$, and is estimated by sample mean in our dataset. The probability that a patient dies after having been proactively transferred to the ICU is estimated similarly and is $d_A = 0.0009$, which is significantly lower than d_C .

Initial distribution and rewards. We set the initial distribution $\mathbf{p}_0 \in \mathbb{R}_+^{n+4}$ as the long-run average occupation of patients in each severity score group according to the data (see Appendix A, table A). We choose a discount factor of 0.95 to capture the importance of future outcomes for the decision-maker. We choose the following rewards, satisfying Assumption 1 and Assumption 3,

$$\begin{aligned} r_W = 100, r_{RL} = \frac{1}{1-\lambda} \cdot 250, r_{PT-RL} = \frac{1}{1-\lambda} \cdot 190, r_{CR-RL} = \frac{1}{1-\lambda} \cdot 160, \\ r_D = \frac{1}{1-\lambda} \cdot 30, r_{CR-D} = \frac{1}{1-\lambda} \cdot 20, r_{PT-D} = \frac{1}{1-\lambda} \cdot 10, \end{aligned} \quad (5.4)$$

We would like to note that the following natural ordering conditions are satisfied:

$$r_{RL} \geq r_{PT-RL} \geq r_{CR-RL},$$

$$r_D \geq r_{CR-D} \geq r_{PT-D}.$$

Certainly, different choices of rewards may lead to different thresholds for the optimal nominal and the optimal robust policies. Appendix H summarizes a detailed sensitivity analysis of our numerical results for different rewards. The single MDP is most valuable in identifying candidate worst-case transition matrices. Thus, while the thresholds of the optimal and nominal robust policies for the single MDP are highly dependent on the rewards, the performance of the hospital (in terms of mortality, LOS and average ICU occupancy) based on the resulting worse-case matrix is fairly consistent across different rewards, including those in (5.4).

5.3.2. Empirical results for the single-patient MDP. From Theorems 1 and 3, we know the optimal nominal and robust policies are of threshold type. Therefore, we consider all threshold policies, denoted by $\pi^{[1]}, \dots, \pi^{[11]}$, and compare their nominal and worst-case rewards in the single-patient MDP for the different uncertainty sets. Figure 3 summarizes these results.

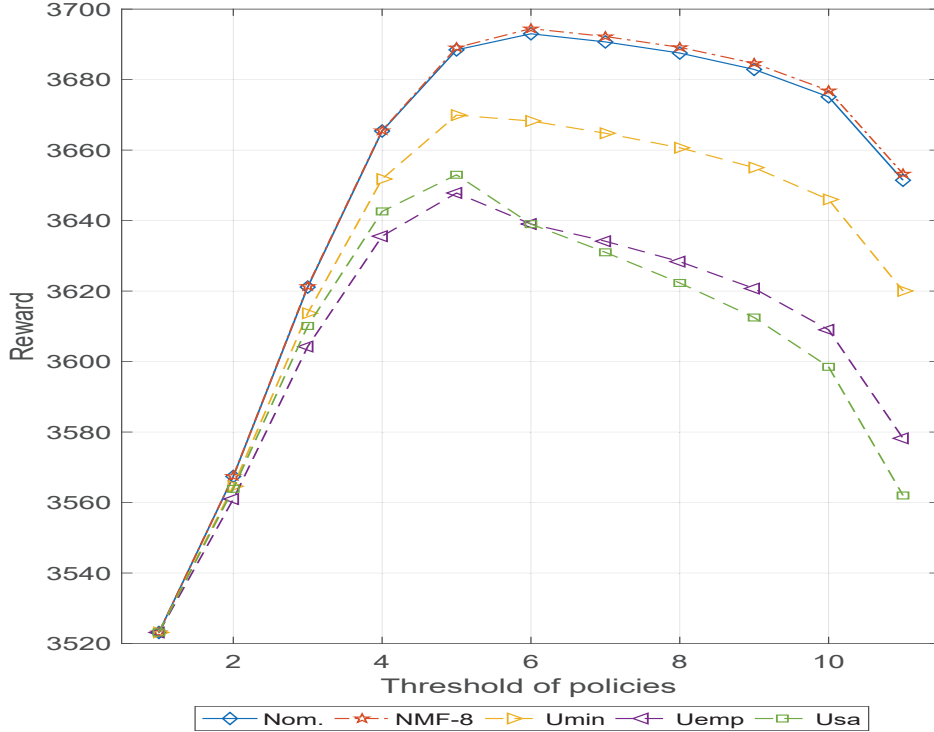


Figure 3 Nominal and worst-case performance of threshold policies for an NMF approximation of rank 8. For any threshold $\tau = 1, \dots, 11$, ‘Nom.’ stands for $R(\pi^{[\tau]}, \mathbf{T}^0)$, while ‘NMF-8’ stands for $R(\pi^{[\tau]}, \hat{\mathbf{T}})$. The other three curves represent the worst-case reward of $\pi^{[\tau]}$ for the specified uncertainty set $(\mathcal{U}_{\min}, \mathcal{U}_{\text{emp}}, \mathcal{U}_{\text{sa}})$.

Note that for all threshold policies $\pi^{[\tau]}$, the corresponding reward using the estimated transition matrix, $R(\pi^{[\tau]}, \mathbf{T}^0)$, and that using the NMF approximation of the transition matrix, $R(\pi^{[\tau]}, \hat{\mathbf{T}})$, are practically indistinguishable. This provides additional support for using our NMF solution as an approximation for \mathbf{T}^0 .

We observe that the optimal nominal policy ($\pi^{[6]}$) is different than the optimal robust policy ($\pi^{[5]}$) for the three different uncertainty sets. Our primary goal with the single MDP is not to provide direct recommendations for the hospital system, but rather to determine candidate transition matrices under which the hospital system can be evaluated. As we will see later, the performance of these policies under their corresponding transition matrices is quite different in the hospital simulation.

We would like to note that the worst-case matrices in \mathcal{U}_{\min} and \mathcal{U}_{emp} do not belong to the 95% confidence intervals (5.1), even though \hat{T} belongs to (5.1). That said, only a few of the coefficients are outside of the 95% confidence regions, and the violations are small. For instance, for the worst-case matrix in \mathcal{U}_{\min} associated with $\pi^{[5]}$, only five coefficients (out of 130) are outside of (5.1), and the worst-case deviation is $-1.2179 \cdot \alpha_3$ (instead of $-\alpha_3$), while the mean of the absolute values of the deviations is 0.3381 and 95% of these absolute values are below 1.0690. The results are similar for \mathcal{U}_{emp} . For instance, for the worst-case matrix in \mathcal{U}_{emp} associated with $\pi^{[5]}$, 20 coefficients out of 130 are outside the confidence intervals. While the coefficient (1,1) is about $10 \cdot \alpha_1$ away from $T_{1,1}^0$ (instead of $2\alpha_1$), the mean of the absolute values of the deviations is 0.4430, and 95% of these absolute values are below 2.3057. Therefore, we can still consider the worst-case matrices for \mathcal{U}_{\min} and \mathcal{U}_{emp} as plausible transition matrices for our hospital model.

5.4. Robustness analysis for the hospital.

The primary purpose of our single-patient MDP model is to develop insights into the management of the full hospital system. To that end, we use our single-patient MDP to generate transition matrices that are candidates for a worst-case deterioration of the hospital performances. Given the complexity and multi-objective nature of the hospital system (i.e., minimize mortality rate, LOS, and average ICU occupancy), defining, let alone deriving, an optimal policy is highly complex. As such, we focus on the class of threshold policies given their desirable theoretical properties (see Theorems 1 and 3) and their simplicity which can help facilitate implementation in practice. For

each threshold policy $\pi^{[\tau]}$ and each uncertainty set \mathcal{U} (among $\mathcal{U}_{\min}, \mathcal{U}_{\text{emp}}, \mathcal{U}_{\text{sa}}$), we compute $\mathbf{T}^{[\tau, \mathcal{U}]}$ a worst-case transition matrix for the single-patient MDP in \mathcal{U} :

$$\mathbf{T}^{[\tau, \mathcal{U}]} \in \arg \min_{\mathbf{T} \in \mathcal{U}} R(\pi^{[\tau]}, \mathbf{T}).$$

Then, we use the pair $(\pi^{[\tau]}, \mathbf{T}^{[\tau, \mathcal{U}]})$ to simulate the hospital performance as measured by the mortality rate, length-of-stay, and average occupancy of the ICU.

5.4.1. In-hospital mortality and Length-Of-Stay.

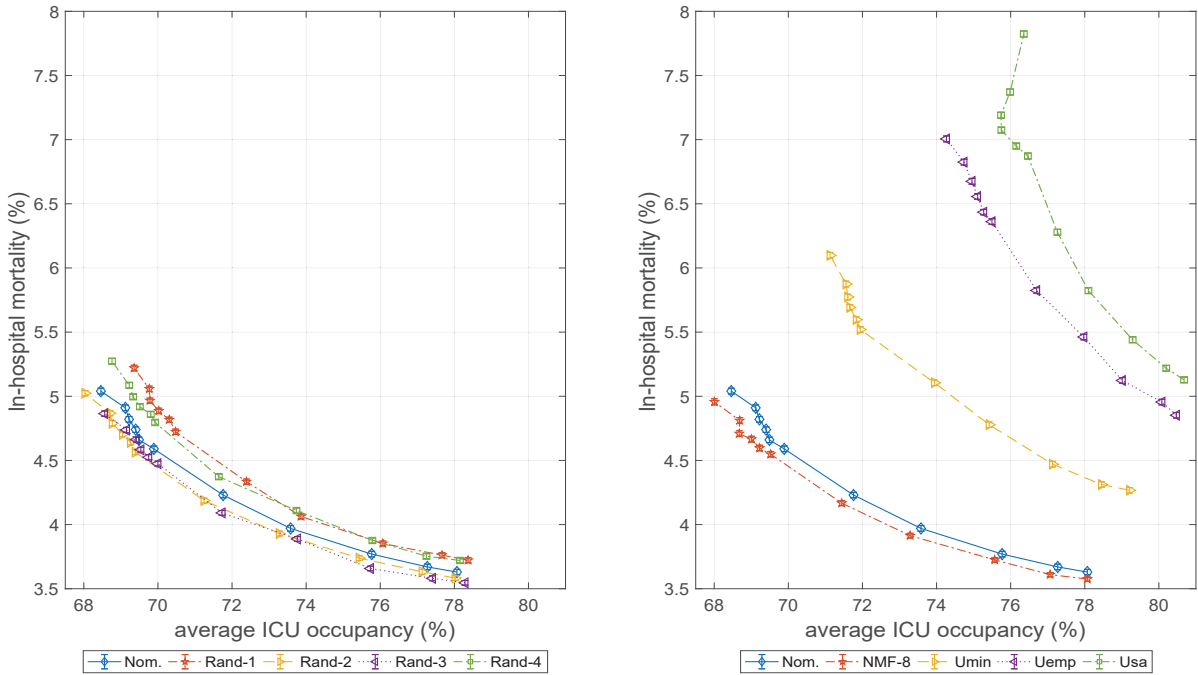
In-hospital mortality. In Figure 4a we study the variation of the hospital performance over the 95% confidence intervals for the nominal transition matrix \mathbf{T}^0 . In particular, we sample $N = 20$ transition matrices in the confidence intervals (5.1) and plot the nominal performance (mortality rate versus average ICU occupancy) of all threshold policies as well as the performances for 4 of the N sampled matrices. The mortality rate for all sampled transition matrices are very close to the nominal mortality rate. The maximum *relative* observed deviation from the nominal mortality rate is 8.82%, with an average relative deviation of 3.84%. We present more details about the statistics of the random deviations from the nominal performance in Appendix J.

In Figure 4b, we compare the worst-case performance of all threshold policies with the nominal performance. For each threshold policy, we construct a worst-case transition matrix that minimizes the single-patient MDP reward and compute the hospital performances for this particular matrix and threshold policy. As we saw in the single-patient MDP experiments, the hospital performance for \mathbf{T}^0 and $\hat{\mathbf{T}}$ are very close, again suggesting that the NMF approximation is reasonable. We consider the three uncertainty sets $\mathcal{U}_{\min}, \mathcal{U}_{\text{emp}}$ and \mathcal{U}_{sa} . Note that \mathcal{U}_{\min} and \mathcal{U}_{emp} are centered around our NMF approximation $\hat{\mathbf{T}}$. Under uncertainty set \mathcal{U}_{\min} , the mortality rate can significantly increase, with relative increases from 18% to 23%. This substantial degradation occurs even though \mathcal{U}_{\min} is our most-optimistic uncertainty set, with variations in the order of 10^{-4} from $\hat{\mathbf{T}}$. For worst-case matrices in \mathcal{U}_{emp} or \mathcal{U}_{sa} , the mortality rate of any threshold policy increases by 40% to 50%. Therefore, our worst-case analysis (Figure 4b) shows that the mortality may severely deteriorate,

even for very small parameters deviations from the nominal transition matrix \mathbf{T}^0 . Note that this is not the case in our random sample analysis (Figure 4a). This suggests that not considering worst-case deviations may lead to overly optimistic estimations of the hospital performances.

As a thought experiment, suppose the decision-maker determined that the average ICU occupancy should not exceed 72%. The decision-maker then chooses the threshold policy with the lowest mortality and average ICU occupancy lower than 72%. Based on the nominal performance, the decision-maker will choose $\pi^{[5]}$ which proactively transfers 27.1% of the patients. However, our analysis demonstrates there exists a ‘worst-case’ transition matrix that is consistent with the available data which, under the selected policy $\pi^{[5]}$, would result in a higher average ICU occupancy of 74.0%. In contrast, if the decision-maker were to account for the parameter uncertainty and consider the worst-case performance in \mathcal{U}_{\min} , the decision-maker would choose $\pi^{[6]}$, which proactively transfers 10.2% of the patients and results in a worst-case average ICU occupancy of 71.9%.

In general, as the threshold decreases, and proactive transfers are used more aggressively, the ICU occupancy increases while the mortality rate decreases. This behavior does not generalize to the uncertainty set \mathcal{U}_{sa} . In particular, we notice in Figure 4b that for \mathcal{U}_{sa} , the worst-case mortality rate *and* the ICU occupancy decrease from $\pi^{[11]}$ to $\pi^{[8]}$. Therefore, the (worst-case) average ICU occupancy decreases when the decision-maker decides to transfer more patients to the ICU. In principle, this could be explained by the fact that the patients with severity scores in $\{8, 9, 10\}$ are the sickest patients. Thus, proactively transferring them may actually be Pareto improving. That said, we are somewhat cautious about the interpretations of \mathcal{U}_{sa} . The worst-case transition matrices in \mathcal{U}_{sa} are extreme perturbations from \mathbf{T}^0 . For instance, the coefficients $T_{1,RL}^0, \dots, T_{n,RL}^0$ become $T_{1,RL}^0 - \alpha_1, \dots, T_{n,RL}^0 - \alpha_n$, and the coefficients $T_{1,D}^0, \dots, T_{n,D}^0$ become $T_{1,D}^0 + 2 \cdot \alpha_1, \dots, T_{n,D}^0 + 2 \cdot \alpha_n$. In that sense, such coordinated, structured parameter misspecification appears unlikely. This is due to the ability to arbitrarily perturb the coefficients of \mathbf{T}^0 , provided that the resulting rows still form a transition kernel, rather than accounting for potential correlations across states that our factor matrix approach incorporates. Such extreme perturbations are unlikely to arise in practice, which is why we focus our attention on the model of factor matrix uncertainty set.



(a) Random samples analysis.

(b) Worst-case analysis.

Figure 4 In-hospital mortality of the 11 threshold policies for the nominal estimated matrix, randomly sampled matrices in the 95% confidence intervals (left-hand side), and the worst-case matrices found by our single MDP model (right-hand side). Each point corresponds to a threshold policy: the policy with highest mortality rate corresponds to threshold $\tau = 11$ (top-left of each curve) and the threshold decreases until the bottom-right point of each curve, corresponding to threshold $\tau = 1$. We consider the uncertainty sets \mathcal{U}_{\min} , \mathcal{U}_{emp} and \mathcal{U}_{sa} when the rank $r = 8$. On the right-hand side, we also report the hospital mortality rate when the transition matrix is our NMF approximation with rank 8.

In comparing the worst-case matrices in \mathcal{U}_{\min} and \mathcal{U}_{emp} to that in \mathcal{U}_{sa} , we find the magnitude of the deviations from \mathbf{T}^0 , as measured by the 1-norm, are of the same order. Despite having similar amount of deviations with \mathcal{U}_{sa} , we see the behavior of the hospital system under the worst-case matrices from \mathcal{U}_{\min} and \mathcal{U}_{emp} is consistent with that of the nominal matrix, which is not the case with \mathcal{U}_{sa} .

Underlying physiologic characteristics dictate the evolution of a patient’s health. As such, it is reasonable to assume these similar medical factors manifest themselves in our model through correlated dynamics across the different severity scores; thus, we expect the uncertainty to be reasonably captured by a low-rank deviation from the nominal estimation \mathbf{T}^0 . Therefore, we expect the true worst-case performances of the threshold policies to be somewhere in between the performances in \mathcal{U}_{\min} and the performances in \mathcal{U}_{emp} .

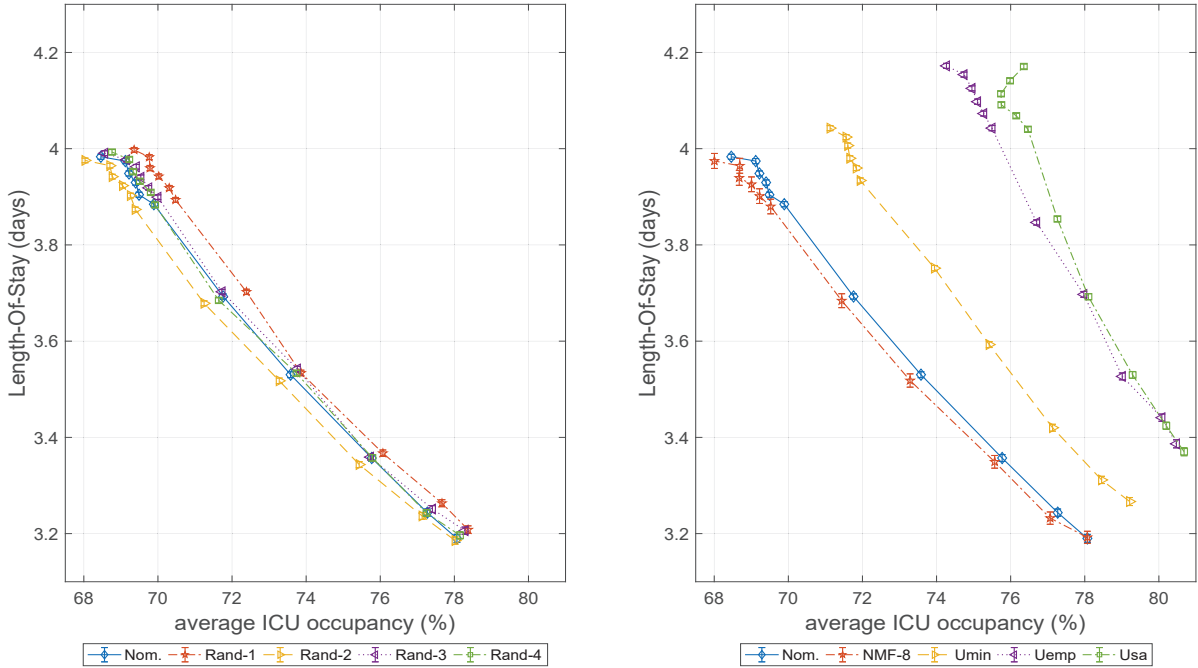
From these experiments we see that 1) ignoring parameter uncertainty may result in overly optimistic expectations of system performances; and 2) the type of parameter uncertainty (e.g. correlated or arbitrary) can have a substantial impact on the insights derived from the robust analysis.

Length-Of-Stay. In the case of Length-Of-Stay (LOS), we notice similar trends as compared to the in-hospital mortality rate. Figure 5a shows the deviations in performance for some randomly sampled matrices. The average deviation ranges from 0.34% for $\pi^{[3]}$ to 1.04% of deviation for threshold $\pi^{[11]}$. Therefore, the hospital flow seems stable as regards to parameters deviations from the nominal matrix \mathbf{T}^0

We compare the worst-case LOS with the nominal performance. We see that the LOS can increase by up to 2.5% in \mathcal{U}_{\min} , and up to 5.0% in \mathcal{U}_{emp} and \mathcal{U}_{sa} . The impact of worst-case parameter deviations is less severe for the Length-Of-Stay than for the mortality rate. However, worst-case deviations are still more substantial than random deviations from the nominal transition (Figure 5a). As for in-hospital mortality rate, in Figure 5b we notice that under uncertainty set \mathcal{U}_{sa} , it appears to be Pareto improving to be more aggressive in proactively transferring patients with threshold policy $\pi^{[8]}$ rather than $\pi^{[11]}$. However, as discussed before, we believe that \mathcal{U}_{sa} is not able to fully capture reasonable types of uncertainty one would expect to see in practice.

6. Conclusions and Discussion.

Interest in preventative and proactive care has been growing. With the advancements in machine-learning, the ability to conduct proactive care based on predictive analytics is quickly becoming a



(a) Random samples analysis.

(b) Worst-case analysis.

Figure 5 Length-Of-Stay of the 11 threshold policies for the nominal estimated matrix, randomly sampled matrices in the 95% confidence intervals (left-hand side), and the worst-case matrices found by our single MDP model (right-hand side). Each point corresponds to a threshold policy: the policy with highest LOS corresponds to threshold $\tau = 11$ (top-left of each curve) and the threshold decreases until the bottom-right point of each curve, corresponding to threshold $\tau = 1$. We consider the uncertainty sets \mathcal{U}_{\min} , \mathcal{U}_{emp} and \mathcal{U}_{sa} when the rank $r = 8$. On the right-hand side, we also report the hospital mortality rate when the transition matrix is our NMF approximation with rank 8.

reality. In this work, we consider the decision to proactively admit patients to the ICU based on a severity score before they suffer a sudden health deterioration in the ward and require even more resources. In practice, an early warning system alert could trigger many potential interventions such as placing the patient in an evaluation state where admission decisions could be made from. While a threshold policy for proactive admission is a simplification of what could happen in practice when an alert is triggered, our analysis facilitates the derivation of valuable insights on the performances

of this simple class of transfer policies and the impact of parameter uncertainty. We explicitly account for parameter uncertainty that arises naturally in practice due to the need to estimate model parameters from finite, real data. Since the severity scores are likely influenced by common underlying medical factors, we introduce a robust model that accounts for potential correlations in the uncertainty related to different severity scores. Under mild and interpretable assumptions, our model shows that the optimal nominal and optimal robust transfer policies are of threshold type. Our extensive simulations show that not accounting for parameter misspecification may lead to overly optimistic estimations of the hospital performances, even for very small deviations. Moreover, we find that unrelated uncertainty may lead to extreme perturbations from the nominal parameters and unreliable insights for the impact of threshold policies on the patient flow in the hospital. Our work suggests that it is crucial for the decision-makers to account for parameters uncertainty when basing their decisions on predictive models where some parameters are estimated from real data.

One limitation of our work is the choice of worst-case as a relevant metric for the decision-maker. While worst-cases performances may be unlikely in practice, it is worth noting that the resulting parameter values for some worst-case performances are in the confidence intervals and therefore are as likely as the nominal parameters. Moreover, in the field of healthcare operations where the goal is to save the lives of the patients, it is still relevant to obtain an estimation of the potential deterioration of the performances of the hospital, especially if the deviation from the nominal parameters is small (i.e., in the confidence intervals). We would like to highlight that our work gives insight on potential misestimations of the metrics (average in-hospital mortality, length-of-stay and ICU occupancy) on which the physicians may base their decisions of a transfer policy. More specifically, we provide a tool to estimate worst-case deteriorations, within the confidence intervals given by some statistical estimators. It is to the merit of the physicians to decide what levels of risks are acceptable

Another limitation is tied to the ability of the severity score to fully capture the patient potential health deterioration. We must recognize that the impact of proactive transfer policies will be highly

dependent on the quality of the evaluation of the severity scores. As such, proactive transfer policies could also be beneficial if based on other metrics, such as LAPS2, MEWS or others, as long as these scores accurately describe a predicted potential patient health deterioration.

There are various interesting directions for future research that arise from our work. For instance, one could consider various level of actions for the proactive policies, ranging from a simple alert to the physicians for more continuous monitoring, to an immediate ICU transfer (the action considered in this paper). Moreover, the proactive transfer policies considered in this work do not account for the number of empty beds in the ICU. One could consider *adaptive thresholds*, varying with the number of free beds in ICU. While Hu et al. (2018) shows with simulations that the performances of such adaptive threshold policies are comparable to the non-adaptive ones, it could be of interest to investigate the theoretical guarantees of such adaptive policies in the framework of our single-patient MDP. Finally, given the vast amount of patients trajectories available in our dataset, one could utilize recent methods from the *off-policy evaluation* literature (Kallus and Uehara (2019) for a review) to obtain a model-free performance estimator, in contrast to our model-based analysis using our single-patient robust MDP.

References

- Alagoz O, Hsu H, Schaefer AJ, Roberts MS (2010) Markov decision processes: a tool for sequential decision making under uncertainty. *Medical Decision Making* 30(4):474–483.
- Alagoz O, Maillart LM, Schaefer AJ, Roberts MS (2004) The optimal timing of living-donor liver transplantation. *Management Science* 50(10):1420–1430.
- Altman E, Jiménez T, Koole G (2001) On optimal call admission control in resource-sharing system. *IEEE Transactions on Communications* 49(9):1659–1668.
- Ayer T, Alagoz O, Stout NK (2012) Or forum—a pomdp approach to personalize mammography screening decisions. *Operations Research* 60(5):1019–1034.
- Barnett M, Kaboli P, Sirio C, G R (2002) Day of the week of intensive care admission and patient outcomes: A multisite regional evaluation. *Medical Care* 40(6):530–539, ISSN 00257079, URL <http://www.jstor.org/stable/3768133>.

-
- Bilben B, Grandal L, Søvik S (2016) National early warning score (news) as an emergency department predictor of disease severity and 90-day survival in the acutely dyspneic patient—a prospective observational study. *Scandinavian journal of trauma, resuscitation and emergency medicine* 24(1):80.
- Bountourelis T, Eckman D, Luangkesorn L, Schaefer A, Nabors SG, Clermont G (2012) Sensitivity analysis of an icu simulation model. *Proceedings of the 2012 Winter Simulation Conference (WSC)*, 1–12 (IEEE).
- Brunelli A, Ferguson MK, Rocco G, Pieretti P, Vigneswaran WT, Morgan-Hughes NJ, Zanello M, Salati M (2008) A scoring system predicting the risk for intensive care unit admission for complications after major lung resection: a multicenter analysis. *The Annals of thoracic surgery* 86(1):213–218.
- Butcher BW, Vittinghoff E, Maselli J, Auerbach AD (2013) Impact of proactive rounding by a rapid response team on patient outcomes at an academic medical center. *Journal of hospital medicine* 8(1):7–12.
- Chan CW, Farias VF, Bambos N, Escobar GJ (2012) Optimizing intensive care unit discharge decisions with patient readmissions. *Operations research* 60(6):1323–1341.
- Cheng G, Xie J, Zheng Z (2019) Optimal stopping for medical treatment with predictive information. *Available at SSRN 3397530* .
- Chrusch C, Olafson K, McMillan P, Roberts D, Gray P (2009) High occupancy increases the risk of early death or readmission after transfer from intensive care. *Critical care medicine* 37:2753–8, URL <http://dx.doi.org/10.1097/CCM.0b013e3181a57b0c>.
- Delage E, Mannor S (2010) Percentile optimization for markov decision processes with parameter uncertainty. *Operations Research* 58(1):203 – 213.
- Escobar G, Gardner N, Greene D, Draper D, Kipnis P (2013) Risk-adjusting hospital mortality using a comprehensive electronic record in an integrated health care delivery system. *Medical Care* 51(5):446–453.
- Goh J, Bayati M, Zenios SA, Singh S, Moore D (2018) Data uncertainty in markov chains: Application to cost-effectiveness analyses of medical innovations. *Operations Research* 66(3):697–715.
- Goyal V, Grand-Clement J (2018) Robust Markov Decision Process: Beyond Rectangularity. *ArXiv e-prints* URL <https://arxiv.org/abs/1811.00215>.
- Green L (2002) How many hospitals beds? *Inquiry* 39(4):400–412.

- Guirgis FW, Gerdik C, Wears RL, Williams DJ, Kalynych CJ, Sabato J, Godwin SA (2013) Proactive rounding by the rapid response team reduces inpatient cardiac arrests. *Resuscitation* 84(12):1668–1673.
- Higgins TL, Estafanous FG, Loop FD, Lee JC, Starr NJ, Knaus WA, Cosgrove III DM (1997) Icu admission score for predicting morbidity and mortality risk after coronary artery bypass grafting. *The Annals of thoracic surgery* 64(4):1050–1058.
- Hu W, Chan CW, Zubizarreta JR, Escobar GJ (2018) An examination of early transfers to the icu based on a physiologic risk score. *Manufacturing & Service Operations Management* 20(3):531–549.
- Iyengar G (2005) Robust dynamic programming. *Mathematics of Operations Research* 30(2):257–280.
- Kallus N, Uehara M (2019) Double reinforcement learning for efficient off-policy evaluation in markov decision processes. *arXiv preprint arXiv:1908.08526* .
- Kc DS, Terwiesch C (2012) An econometric analysis of patient flows in the cardiac intensive care unit. *Manufacturing & Service Operations Management* 14(1):50–65.
- Kim SH, Chan CW, Olivares M, Escobar G (2014) Icu admission control: An empirical study of capacity allocation and its implication for patient outcomes. *Management Science* 61(1):19–38.
- Kipnis P, Turk BJ, Wulf DA, LaGuardia JC, Liu V, Churpek MM, Romero-Brufau S, Escobar GJ (2016) Development and validation of an electronic medical record-based alert score for detection of inpatient deterioration outside the icu. *Journal of biomedical informatics* 64:10–19.
- Mannor S, Mebel O, Xu H (2016) Robust mdps with k-rectangular uncertainty. *Mathematics of Operations Research* 41(4):1484–1509.
- Mannor S, Simester D, Sun P, Tsitsiklis JN (2007) Bias and variance approximation in value function estimates. *Management Science* 53(2):308–322.
- McClellan S, Millard P (2006) Using markov models to manage high occupancy hospital care. *2006 3rd International IEEE Conference Intelligent Systems*, 256–260 (IEEE).
- Milbrandt B, A K, Rahim T, Dremsizov T, Clermont G, Cooper M, Angus C, Linde-Zwirble T (2008) Growth of intensive care unit resource use and its estimated cost in medicare. *Critical Care Medicine* 36(9):2504–2510.

-
- Mullins P, Goyal M, Pines M (2013) National growth in intensive care unit admissions from emergency departments in the united states from 2002 to 2009. *Academic Emergency Medicine* 20(5):479–486.
- Nilim A, El Ghaoui L (2005) Robust control of markov decision processes with uncertain transition matrices. *Operations Research* 53(5):780–798.
- Opgenorth D, Stelfox H, Gilfoyle E, Gibney R, Meier M, Boucher P, Mckinlay D, McIntosh C, Wang X, Zygun D, Bagshaw S (2018) Perspectives on strained intensive care unit capacity: A survey of critical care professionals. *Plos one* 13(8).
- Özekici S, Pliska SR (1991) Optimal scheduling of inspections: A delayed markov model with false positives and negatives. *Operations Research* 39(2):261–273.
- Peck JS, Benneyan JC, Nightingale DJ, Gaehde SA (2012) Predicting emergency department inpatient admissions to improve same-day patient flow. *Academic Emergency Medicine* 19(9):E1045–E1054.
- Puterman M (1994) *Markov Decision Processes : Discrete Stochastic Dynamic Programming* (John Wiley and Sons).
- Putnam KG, Buist DS, Fishman P, Andrade SE, Boles M, Chase GA, Goodman MJ, Gurwitz JH, Platt R, Raebel MA, et al. (2002) Chronic disease score as a predictor of hospitalization. *Epidemiology* 13(3):340–346.
- Rapsang AG, Shyam DC (2014) Scoring systems in the intensive care unit: a compendium. *Indian journal of critical care medicine: peer-reviewed, official publication of Indian Society of Critical Care Medicine* 18(4):220.
- Satia J, Lave R (1973) Markov decision processes with uncertain transition probabilities. *Operations Research* 21(3):728–740.
- Schaefer AJ, Bailey MD, Shechter SM, Roberts MS (2005) Modeling medical treatment using markov decision processes. *Operations research and health care*, 593–612 (Springer).
- Shechter SM, Bailey MD, Schaefer AJ, Roberts MS (2008) The optimal time to initiate hiv therapy under ordered health states. *Operations Research* 56(1):20–33.
- Shmueli A, Baras M, Sprung CL (2004) The effect of intensive care on in-hospital survival. *Health Services and Outcomes Research Methodology* 5(3-4):163–174.

- Shmueli A, Sprung CL, Kaplan EH (2003) Optimizing admissions to an intensive care unit. *Health Care Management Science* 6(3):131–136.
- Sison C, Glaz J (1995) Simultaneous confidence intervals and sample size determination for multinomial proportions. *Journal of the American Statistical Association* 90(429):366–369.
- Steimle LN, Denton BT (2017) Markov decision processes for screening and treatment of chronic diseases. *Markov Decision Processes in Practice*, 189–222 (Springer).
- Steimle LN, Kaufman DL, Denton BT (2018) Multi-model markov decision processes. *Optimization Online* URL http://www.optimization-online.org/DB_FILE/2018/01/6434.pdf.
- Wiesemann W, Kuhn D, Rustem B (2013) Robust markov decision processes. *Operations Research* 38(1):153–183.
- Xu H, Mannor S (2006) The robustness-performance tradeoff in markov decision processes. *Neural Information Processing Systems* 1537–1544.
- Xu K, Chan CW (2016) Using future information to reduce waiting times in the emergency department via diversion. *Manufacturing & Service Operations Management* 18(3):314–331.
- Xu Y, Yin W (2013) A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion. *SIAM Journal on imaging sciences* 6(3):1758–1789.
- Yauney G, Shah P (2018) Reinforcement learning with action-derived rewards for chemotherapy and clinical trial dosing regimen selection. *Machine Learning for Healthcare Conference*, 161–226.

Appendix A: Details about the hospital model of Figure 1.

We give details on the hospital model from Figure 1, as introduced by the authors of Hu et al. (2018). As in Hu et al. (2018), the parameters of the model were calibrated using sample means across all 21 hospitals and/or estimates from regression models.

Direct admits. The arrivals of the patients who are directly admitted to the ICU follows a non-homogeneous Poisson process. The empirical arrival rates are estimated using 12 months of data across all 21 hospitals. The LOS of a direct admit patient is lognormally distributed with mean $1/\mu_E$ and standard deviation σ_E , and a proportion p_E (following a distribution with density f_{p_E}) of this LOS is spent in the ICU, while the remaining time is spent in the ward. The rate of readmission

to the ICU is ρ_E . At the end of this LOS, the patients are discharged with probability $1 - d_E$. The value of the parameters are the following: $d_E = 9.41\%$, $\rho_E = 15.76\%$, $1/\mu_E = 5.49$ (days), $\sigma_E = 5.71$ (days), $\mathbb{E}[p_E] = 50.79\%$. The density f_{p_E} is the empirical distribution derived from the dataset.

Transfer from the ward. These patients can be divided into the patients who have versus have not already been to the ICU. Consider the patients who have never been to the ICU. A patient arrives in the ward with a severity score of $i \in \{1, \dots, n\}$ following a non-homogeneous Poisson process. Every 6 hours, s/he then transitions to another risk score j with probability T_{ij}^0 , or s/he may ‘crash’ and require ICU admission, recover and leave the hospital, or die. After a patient has crashed, a LOS is chosen which is lognormally distributed with mean $1/\mu_C$ and standard deviation σ_C , and a proportion p_W (following a distribution with density f_{p_W}) of this LOS is spent in the ICU, while the remaining proportion $1 - p_W$ of time is spent in the ward. At the end of this LOS, the patient is discharged with probability $1 - d_C$. If there are no available beds in the ICU when a crashed patient requires an admission, the ICU patient with the shortest remaining service time in the ICU will be discharged, and this is called a ‘demand-driven discharge’. Such a patient will have a readmission rate of ρ_D , higher than the readmission rate ρ_C of ward patients who were naturally discharged from the ICU after finishing their service time in the ICU. In particular, we have the following values, estimated through empirical averages in our dataset: $\rho_C = 16.88\%$, $\rho_D = 18.13\%$, $1/\mu_C = 12.54$ (days), $\sigma_C = 10.13$ (days), $\mathbb{E}[p_W] = 46.92\%$, and $d_C = 57.28\%$. Similarly, the density f_{p_W} is derived as the empirical distribution from the dataset.

Proactive transfer. Every 6 hours, the doctors might perform a proactive transfer and transfer a patient in the ward to the ICU, if there is an available bed. When a patient is proactively transferred, the hospital LOS is lognormally distributed with mean $1/\mu_{A,i}$ and standard deviation $\sigma_{A,i}$, while a proportion $p_W \sim f_{p_W}$ of this LOS is spent in the ICU. The patient will then survive the hospital discharge with a probability $1 - d_{A,i}$. If this patient is naturally discharged from the ICU, the readmission rate is $\rho_{a,i} = \rho_C$, otherwise it is ρ_D . We indicate below the proportion, the mortality rate and the LOS related to each $n = 10$ severity scores.

Appendix B: Proof of Lemma B

Proof. We will prove that for any policy π and for any transition matrix \mathbf{T} , for all severity score $i \in [n]$, the value vector \mathbf{V} of policy π satisfies

$$V_i \leq r_W + \lambda \cdot r_{RL}.$$

Let $i \in [n]$. By definition, V_i is the expected infinite-horizon reward starting from state i :

$$V_i = E^{\pi, \mathbf{T}} \left[\sum_{t=0}^{\infty} \lambda^t r_{i_t} \mid i_0 = i \right].$$

Severity score i	1	2	3	4	5	6	7	8	9	10
Proportion (%)	17.6	20.3	20.0	14.9	16.9	2.2	2.0	2.0	2.0	2.0
Mortality $d_{A,i}$ (%)	0.01	0.02	0.04	0.05	0.11	0.18	0.28	0.39	0.70	6.84
LOS average $1/\mu_{A,i}$ (days)	0.85	0.91	0.97	1.04	1.17	1.36	1.45	1.57	1.85	3.77
LOS std $\sigma_{A,i}$ (days)	0.68	0.74	0.78	0.84	0.95	1.10	1.17	1.27	1.50	3.04

Table 2 Statistics of patients across the 10 severity scores.

Let us consider a trajectory O of the Markov chain on \mathbb{S} associated with (π, \mathbf{T}) . Then either the patient stays infinitely in the ward, in which case the reward is $r_W \cdot (1 - \lambda)^{-1}$, which is smaller than $r_W + \lambda \cdot r_{RL}$ by Assumption 1; otherwise during the trajectory O , there is a time t at which the patient leaves the ward and reaches the state $n + 1 = CR$, $n + 2 = RL$, $n + 3 = D$, or $n + 4 = PT$. In that case the reward is smaller than

$$\frac{r_W \cdot (1 - \lambda^t)}{1 - \lambda} + \lambda^{t+1} \cdot \max\{r_{CR}, r_{RL}, r_D, r_{PT}\}.$$

Since the maximum instantaneous reward is r_{RL} , the reward associated with the trajectory O is smaller than

$$\frac{r_W \cdot (1 - \lambda^t)}{1 - \lambda} + \lambda^{t+1} \cdot r_{RL}.$$

Now

$$\frac{r_W \cdot (1 - \lambda^t)}{1 - \lambda} + \lambda^{t+1} \cdot r_{RL} \leq (r_W + \lambda \cdot r_{RL}) \cdot (1 - \lambda^t) + \lambda^{t+1} \cdot r_{RL} \quad (\text{B.1})$$

$$\leq r_W + \lambda \cdot r_{RL} - \lambda^t \cdot r_W - \lambda^{t+1} \cdot r_{RL} + \lambda^{t+1} \cdot r_{RL} \quad (\text{B.2})$$

$$\leq r_W + \lambda \cdot r_{RL} - \lambda^t \cdot r_W \quad (\text{B.3})$$

$$\leq r_W + \lambda \cdot r_{RL}, \quad (\text{B.4})$$

where Inequality (B.1) follows from Assumption 1. Therefore, the reward associated with any trajectory O is smaller than $r_W + \lambda \cdot r_{RL}$. We can thus conclude that the value vector \mathbf{V} satisfies

$$V_i \leq r_W + \lambda \cdot r_{RL}, \forall i \in [n].$$

□

Appendix C: Homogeneity of conditions (3.3) and (3.4).

We prove the following two lemmas.

LEMMA 2. *Let $(r_W, r_{CR}, r_{RL}, r_D, r_{PT}) \in \mathbb{R}_+^5$ denote some rewards and \mathbf{T} a transition matrix such that condition (3.3) and condition (3.4) hold.*

1. Let $\alpha \geq 0$. For $\alpha \cdot (r_W, r_{CR}, r_{RL}, r_D, r_{PT})$ and \mathbf{T} , condition (3.3) and condition (3.4) still hold.
2. Let $\alpha \geq 0$. For $(r_W + \alpha, r_{CR} + \alpha, r_{RL} + \alpha, r_D + \alpha, r_{PT} + \alpha)$ and \mathbf{T} , condition (3.4) still holds.

Proof. 1. This follows from $\alpha \geq 0$ and

$$\alpha \cdot r_{CR} \cdot T_{i,n+1}^0 + \alpha \cdot r_{RL} \cdot T_{i,n+2}^0 + \alpha \cdot r_D \cdot T_{i,n+3}^0 = \alpha \cdot (r_{CR} \cdot T_{i,n+1}^0 + r_{RL} \cdot T_{i,n+2}^0 + r_D \cdot T_{i,n+3}^0),$$

and

$$\frac{\alpha \cdot r_W + \lambda \cdot \alpha \cdot r_{PT}}{\alpha \cdot r_W + \lambda \cdot \alpha \cdot r_{RL}} = \frac{r_W + \lambda \cdot r_{PT}}{r_W + \lambda \cdot r_{RL}}.$$

2. Let us assume that

$$\frac{r_W + \lambda \cdot r_{PT}}{r_W + \lambda \cdot r_{RL}} \geq \frac{\left(\sum_{j=1}^n T_{i+1,j}^0\right)}{\left(\sum_{j=1}^n T_{ij}^0\right)}, \forall i \in [n-1].$$

We write ϕ the function of \mathbb{R} such that for any scalar α ,

$$\phi(\alpha) = \frac{r_W + \alpha + \lambda \cdot (r_{PT} + \alpha)}{r_W + \alpha + \lambda \cdot (r_{RL} + \alpha)}.$$

We will prove that ϕ is non-decreasing and therefore that

$$\phi(\alpha) \geq \phi(0) = \frac{r_W + \lambda \cdot r_{PT}}{r_W + \lambda \cdot r_{RL}} \geq \frac{\left(\sum_{j=1}^n T_{i+1,j}^0\right)}{\left(\sum_{j=1}^n T_{ij}^0\right)}, \forall i \in [n-1].$$

Indeed, ϕ has a derivative in \mathbb{R}_+ and

$$\begin{aligned} \phi'(\alpha) &= \frac{(1 + \lambda) \cdot (r_W + \alpha + \lambda(r_{RL} + \alpha)) - r_W - \alpha - \lambda \cdot (r_{PT} + \alpha)}{(r_W + \alpha + \lambda \cdot (r_{RL} + \alpha))^2} \\ &= \frac{\lambda(1 + \lambda) \cdot (r_{RL} - r_{PT})}{(r_W + \alpha + \lambda \cdot (r_{RL} + \alpha))^2} \\ &\geq 0, \end{aligned}$$

since $r_{RL} \geq r_{PT}$. Therefore ϕ is a non-decreasing function, and for all $\alpha \geq 0$, the condition (3.4) holds. \square

LEMMA 3. Let $(r_W, r_{CR}, r_{RL}, r_D, r_{PT}) \in \mathbb{R}_+^5$ denote some rewards and \mathbf{T} a transition matrix such that condition (3.5) holds.

1. Let $\alpha \geq 0$. For $\alpha \cdot (r_W, r_{CR}, r_{RL}, r_D, r_{PT})$ and \mathbf{T} , condition (3.5) still holds.
2. Let $\alpha \geq 0$ and let us assume that

$$\sum_{j=1}^n T_{ij} \geq \sum_{j=1}^n T_{i+1,j}, \forall i \in [n-1].$$

Then for $(r_W + \alpha, r_{CR} + \alpha, r_{RL} + \alpha, r_D + \alpha, r_{PT} + \alpha)$ and \mathbf{T} , condition (3.5) still holds.

Proof. 1. Let α be a non-negative scalar. For the same reason as in Lemma 2, condition (3.5) still holds for $\alpha \cdot (r_W, r_{PT}, r_{RL}, r_{CR})$ and \mathbf{T} .

2. Let $\alpha \geq 0$. For any $i \in [n-1]$, we have

$$\left(\sum_{j=1}^n T_{ij}^0 \right) \cdot (r_W + \lambda \cdot r_{PT}) + out(i) \geq \left(\sum_{j=1}^n T_{i+1,j}^0 \right) \cdot (r_W + \lambda \cdot r_{RL}) + out(i+1), \quad (\text{C.1})$$

where

$$out(i) = r_{CR} \cdot T_{i,n+1}^0 + r_{RL} \cdot T_{i,n+2}^0 + r_D \cdot T_{i,n+3}^0.$$

Since $\sum_{j=1}^{n+3} T_{\ell,j} = 1$ for any severity score $\ell \in [n-1]$, we notice that adding α to all rewards is equivalent to adding $\alpha + \lambda \cdot \alpha \cdot \left(\sum_{j=1}^{10} T_{ij} \right)$ to the left-hand side of (C.1) and $\alpha + \lambda \cdot \alpha \cdot \left(\sum_{j=1}^{10} T_{i+1,j} \right)$ to the right-hand side of (C.1). Therefore, condition (3.5) holds for all $\alpha \geq 0$, as long as

$$\sum_{j=1}^n T_{ij} \geq \sum_{j=1}^n T_{i+1,j}, \forall i \in [n-1].$$

□

Appendix D: Proof of Theorem 1.

Proof. Let $\mathbf{V}^0 \in \mathbb{R}^{n+4}$ such that

$$\begin{aligned} V_i^0 &= 0, \forall i \in [n], \\ V_{n+1}^0 &= V_{n+2}^0 = V_{n+3}^0 = 0, \\ V_{n+4}^0 &= r_{PT}. \end{aligned}$$

Let $F: \mathbb{R}^{n+4} \rightarrow \mathbb{R}^{n+4}$ denote the function that maps $\mathbf{V} \in \mathbb{R}^{n+4}$ to $F(\mathbf{V})$, where

$$\begin{aligned} F(V)_i &= \max\{r_W + \lambda \cdot \sum_{j=1}^{n+3} T_{ij}^0 V_j, r_W + \lambda \cdot r_{PT}\}, \forall i \in [n], \\ F(V)_{n+1} &= r_{CR}, F(V)_{n+2} = r_{RL}, F(V)_{n+3} = r_D, \\ F(V)_{n+4} &= r_{PT}. \end{aligned}$$

The function F is the Bellman operator associated with our single-patient MDP with transition kernel \mathbf{T}^0 . Therefore, we know that the value iteration algorithm finds an optimal policy (Puterman (1994)): if π^* is an optimal policy and \mathbf{V}^{π^*} is its value vector, then

$$\lim_{t \rightarrow \infty} F^t(\mathbf{V}^0) = \lim_{t \rightarrow \infty} \mathbf{V}^t = \mathbf{V}^{\pi^*},$$

and $\lim_{t \rightarrow \infty} \pi^t \rightarrow \pi^*$ where π^t is the sequence of deterministic policies such that

$$\pi^t(i) = 1 \iff r_W + \lambda \cdot \sum_{j=1}^{n+3} T_{ij}^0 V_j^t < r_W + \lambda \cdot r_{PT}, \forall i \in [n].$$

We will prove by induction that the policy π^t is a threshold policy at every iteration $t \geq 1$.

At $t = 1$, for $i \in [n]$,

$$\begin{aligned} F(\mathbf{V})_i &= \max\{r_W + \lambda \cdot \sum_{j=1}^{n+3} T_{ij} V_j^0, r_W + \lambda \cdot r_{PT}\} \\ &= \max\{r_W + \lambda \cdot 0, r_W + \lambda \cdot r_{PT}\}, \end{aligned}$$

and therefore we have

$$\pi^1(i) = 1, \forall i \in [n].$$

Therefore, π^1 is a threshold policy, and its threshold is 1: $\pi^1 = \pi^{[1]}$.

Let $t \geq 1$ and let us assume that π^t is a threshold policy, and let its threshold be $\tau \in [n + 1]$. We prove that the policy π^{t+1} is a threshold policy. In order to do so, we will prove that for any $i \in [n - 1]$,

$$\pi^{t+1}(i) = 1 \Rightarrow \pi^{t+1}(i + 1) = 1.$$

Let $i \in [n - 1]$ such that $\pi^{t+1}(i) = 1$. From the definition of π^{t+1} ,

$$r_W + \lambda \cdot \left(\sum_{j=1}^n T_{ij}^0 V_j^t + out(i) \right) < r_W + \lambda \cdot r_{PT}.$$

Moreover, since $\pi^t = \pi^{[\tau]}$, we know that the vector \mathbf{V}^t is such that

$$V_\ell^t > r_W + \lambda \cdot r_{PT}, \forall \ell \in \{1, \dots, \tau - 1\}, \quad (\text{D.1})$$

$$V_\ell^t = r_W + \lambda \cdot r_{PT}, \forall \ell \in \{\tau, \dots, n\}, \quad (\text{D.2})$$

$$V_{n+1}^t = r_{CR}, V_{n+2}^t = r_{RL}, V_{n+3}^t = r_D, \quad (\text{D.3})$$

$$V_{n+4}^t = r_{PT}. \quad (\text{D.4})$$

Now following the value iteration algorithm, we know that

$$F(\mathbf{V}^t)_\ell \in \max\{r_W + \lambda \cdot \sum_{j=1}^{n+3} T_{\ell,j}^0 V_j^t, r_W + \lambda \cdot r_{PT}\}, \forall \ell \in [n].$$

We have

$$r_W + \lambda \cdot r_{PT} \geq r_W + \lambda \cdot \sum_{j=1}^{n+3} T_{ij}^0 V_j^t \quad (\text{D.5})$$

$$\geq r_W + \lambda \cdot \sum_{j=1}^n T_{ij}^0 V_j^t + \lambda \cdot out(i) \quad (\text{D.6})$$

$$\geq r_W + \lambda \cdot \sum_{j=1}^{\tau-1} T_{ij}^0 V_j^t + \lambda \cdot \sum_{j=\tau}^n T_{ij}^0 V_j^t + \lambda \cdot out(i) \quad (\text{D.7})$$

$$\geq r_W + \lambda \cdot \sum_{j=1}^{\tau-1} T_{ij}^0 V_j^t + \lambda \cdot \sum_{j=\tau}^n T_{ij}^0 (r_W + \lambda \cdot r_{PT}) + \lambda \cdot out(i) \quad (\text{D.8})$$

$$\geq r_W + \lambda \cdot \sum_{j=1}^{\tau-1} T_{ij}^0 (r_W + \lambda \cdot r_{PT}) + \lambda \cdot \sum_{j=\tau}^n T_{ij}^0 (r_W + \lambda \cdot r_{PT}) + \lambda \cdot out(i) \quad (\text{D.9})$$

$$\geq r_W + \lambda \cdot \sum_{j=1}^{\tau-1} T_{i+1,j}^0 (r_W + \lambda \cdot r_{RL}) + \lambda \cdot \sum_{j=\tau}^n T_{i+1,j}^0 (r_W + \lambda \cdot r_{PT}) + \lambda \cdot out(i+1) \quad (\text{D.10})$$

$$\geq r_W + \lambda \cdot \sum_{j=1}^{\tau-1} T_{i+1,j}^0 V_j^t + \lambda \cdot \sum_{j=th}^n T_{i+1,j}^0 (r_W + \lambda \cdot r_{PT}) + \lambda \cdot out(i+1) \quad (\text{D.11})$$

$$\geq r_W + \lambda \cdot \sum_{j=1}^{n+3} T_{i+1,j}^0 V_j^t, \quad (\text{D.12})$$

where Inequality (D.5) follows from $\pi^{t+1}(i) = 1$, Inequality (D.8) follows from the fact that the policy π^t is a threshold policy and therefore the vector \mathbf{V}^t satisfies Inequalities (D.1). Inequality (D.10) follows from Assumption 2, because Assumption 2 and $r_{PT} \leq r_{RL}$ imply that for all threshold $\tau \in \{1, \dots, n+1\}$,

$$\left(\sum_{j=1}^n T_{ij}^0 \right) \cdot (r_W + \lambda \cdot r_{PT}) + \lambda \cdot out(i) \geq \left(\sum_{j=1}^{\tau-1} T_{i+1,j}^0 \right) \cdot (r_W + \lambda \cdot r_{RL}) + \left(\sum_{j=\tau}^n T_{i+1,j}^0 \right) \cdot (r_W + \lambda \cdot r_{PT}) + \lambda \cdot out(i+1).$$

Inequality (D.11) follows from Lemma 1: $\forall \ell \in [n]$,

$$\begin{aligned} V_\ell^t &\leq V_\ell^{\pi^*} \\ &\leq r_W + \lambda \cdot r_{RL}, \end{aligned}$$

where the first inequality follows from the fact that the operator F is a *non-increasing* mapping and

$$\lim_{t \rightarrow \infty} \mathbf{V}^t = \lim_{t \rightarrow \infty} F^t(\mathbf{V}^0) = \mathbf{V}^{\pi^*}.$$

Therefore, we conclude that

$$r_W + \lambda \cdot r_{PT} \geq r_W + \lambda \cdot \sum_{j=1}^{n+3} T_{i+1,j}^0 V_j^t,$$

which implies that $\pi^{t+1}(i+1) = 1$. We can thus conclude that the policy π^{t+1} is a threshold policy.

Therefore, for all $t \geq 1$, the policy π^t is threshold. We can conclude that there exists an optimal policy that is a threshold policy. \square

Appendix E: Non-threshold optimal policies.

We provide an example of a single-patient MDP which does not satisfy Assumption 2 and for which the optimal nominal policy is not threshold; in particular, the optimal policy in the MDP of Figure 6 is to proactively transfer a patient in state 1 and to not proactively transfer a patient in state 2.

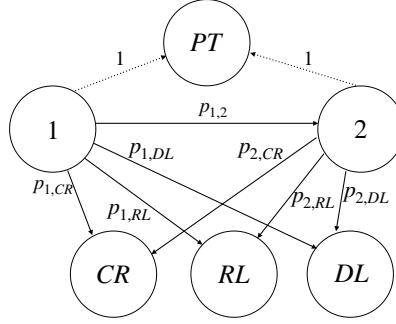


Figure 6 Example of an MPD where the optimal policy is not threshold. There is no self-transition in state 1 or 2. In state 1, the patient transitions to state 2, *CR*, *RL* or *D* (solid arcs), or is proactively transferred (dashed arc). In state 2, the patient has to exit the ward, either by proactive transfer, in which case s/he transitions to *PT* with probability 1 (dashed arc), either by transitioning *CR*, *RL* or *D* (solid arcs); s/he can not transition back to state 1. We provide values of the rewards and transitions for which the optimal policy is not threshold.

Condition (3.3) is not satisfied, i.e., $out(1) < out(2)$. We also set the rewards $r_W, r_{PT}, r_{RL}, r_{CR}, r_D$ such that $out(2) > r_{PT}$, which means that the optimal nominal policy will not proactively transfer the patient when in state 2: $\pi^*(1) = 0$. However,

$$\begin{aligned} \pi^*(1) = 1 &\iff r_W + \lambda \cdot r_{PT} > r_W + \lambda \cdot (out(1) + p_{1,2}(r_W + \lambda \cdot out(2))) \\ &\iff r_{PT} > out(1) + p_{1,2}(r_W + \lambda \cdot out(2)). \end{aligned}$$

Therefore, when $out(1) < r_{PT} < out(2)$ and the discount factor λ is small enough, the decision-maker has an incentive to proactively transfer the patient in state 1. In particular, this is the case for the following set of parameters:

$$(p_{1,RL}, p_{1,CR}, p_{1,D}) = (0.3, 0, 0.3), p_{1,2} = 0.4, (p_{2,RL}, p_{2,CR}, p_{2,D}) = (0.3, 0.4, 0.3), \lambda = 0.01,$$

$$r_W = 1.6, r_{RL} = 3, r_{CR} = 2, r_D = 1.5, r_{PT} = 2.15.$$

We detail the computation of an optimal policy for the single-patient MDP of Figure 6. We start with the Bellman Equation in state 2:

$$V_2^* = \max\{r_W + \lambda \cdot out(2), r_W + \lambda \cdot r_{PT}\}.$$

Since $out(2) = (0.3, 0.4, 0.3)^\top(3, 2, 1.5) = 2.15 > r_{PT} = 2$, we know that $\pi^*(2) = 0$, and the optimal policy does not transfer the patient with a severity score of 2. Moreover, $V_2^* = r_W + \lambda \cdot out(2) = 1.6 + 0.01 \cdot 2.15 = 1.6215$. Let us compute V_1^* . The Bellman Equation in state 1 gives

$$V_1^* = \max\{r_W + \lambda \cdot (p_{1,2} \cdot V_2^* + out(1)), r_W + \lambda \cdot r_{PT}\}.$$

Moreover, $out(1) = (0.3, 0, 0.3)^\top(3, 2, 1.5) = 1.35 < out(2) = 2.15$. Therefore,

$$V_1^* = \max\{1.6 + 0.01 \cdot (0.4 \cdot 1.6215 + 1.35), 1.6 + 0.01 \cdot 2\} = \max\{1.619986, 1.62\} = 1.62,$$

from which we conclude that $\pi^*(1) = 1$, and the optimal policy proactively transfers the patients with severity score of 1. Therefore, the optimal nominal policy is not threshold. We would like to note that we could have chosen any set of parameters for which

$$out(2) > r_{PT} > out(1),$$

$$r_{PT} > p_{1,2} \cdot (r_W + \lambda \cdot out(2)) + out(1).$$

In practice, the discount factor is likely to be significantly higher than 0.01, since the decision-maker in the hospital likely does care about the long-term impacts of the transfer policies.

Appendix F: Non-uniform rewards for the single-patient MDP.

We now consider the case where the rewards in the ward $(r_{W,i})_{i \in [n]}$ and the rewards for proactive transfers $(r_{PT,i})_{i \in [n]}$ depend on the severity of the patient. In this case, we will prove that in our single-patient MDP, the optimal nominal policy is still threshold. In this section, we write $\mathbf{r}_W = (r_{W,i})_{i \in [n]}$ and $\mathbf{r}_{PT} = (r_{PT,i})_{i \in [n]}$. We will also write $\mathbf{r}_{RL} = (r_{RL})_{i \in [n]}$ the vector of \mathbb{R}^n whose components are r_{RL} everywhere. Note that we need to introduce the states $PT_i, i \in [n]$ which represent a patient who has been proactively transferred from severity score $i \in [n]$. Note that there are now $n + n + 3 = 2n + 3$ states. The states $i \in \{1, \dots, n\}$ correspond to the severity scores of a patient in the ward, the states $i \in \{n + 1, \dots, 2n\}$ correspond to the terminal states after the proactive transfer of a patient from state $i - n$. The $(2n + 1)$ -th state corresponds to CR , the $(2n + 2)$ -th state corresponds to RL and the $(2n + 3)$ -th state corresponds to D .

Similarly as for the case of uniform rewards (Theorem 1), we need two assumptions to prove that there exists an optimal nominal policy that is threshold (Theorem 5). We start with the following

assumption which is an extension of Assumption 1 to non-uniform rewards. Condition (F.1) means that it is better for the patient to recover and leave than to be proactively transferred (from any state). Condition (F.2) means that it is better for the patient to recover and leave (from any state) than to stay in the ward forever (in any state). We use Assumption 4 to bound the reward of the value vector of any policy in our proof of Theorem 5.

ASSUMPTION 4.

$$r_{PT,i} \leq r_{RL}, \forall i \in [n], \quad (\text{F.1})$$

$$\frac{1}{1-\lambda} r_{W,i} \leq r_{W,\ell} + \lambda \cdot r_{RL}, \forall (i, \ell) \in [n] \times [n]. \quad (\text{F.2})$$

We also need the following assumption, which is a generalization of Assumption 2 for non-uniform rewards. Note that condition (F.3) and condition (F.4) in Assumption 5 have the same interpretations as condition (3.3) and condition (3.4) in Assumption 2.

ASSUMPTION 5. *We assume that*

$$\text{out}(i) \geq \text{out}(i+1), \forall i \in [n-1], \quad (\text{F.3})$$

and that

$$\mathbf{T}_{i,\cdot}^\top (\mathbf{r}_W + \lambda \cdot \mathbf{r}_{PT}) - r_{PT,i} \geq \mathbf{T}_{i+1,\cdot}^\top (\mathbf{r}_W + \lambda \cdot \mathbf{r}_{RL}) - r_{PT,i+1}, \forall i \in [n-1]. \quad (\text{F.4})$$

The following theorem holds.

THEOREM 5. *Under Assumption 4 and Assumption 5, there exists an optimal nominal policy that is a threshold policy.*

Proof. We adapt the proof of Theorem 1 to the case of non-uniform rewards for the states in the ward ($i \in [n]$) and the proactive transfers ($PT_i, i \in [n]$.) Let $\mathbf{V}^0 \in \mathbb{R}^{2n+3}$ such that

$$\begin{aligned} V_i^0 &= 0, \forall i \in [n], \\ V_{n+i}^0 &= r_{PT,i}, \forall i \in [n], \\ V_{2n+1}^0 &= V_{2n+2}^0 = V_{2n+3}^0 = 0. \end{aligned}$$

Let $F: \mathbb{R}^{2n+3} \rightarrow \mathbb{R}^{2n+3}$ be the mapping such that

$$\begin{aligned} F(V)_i &= \max\{r_{W,i} + \lambda \cdot \left(\sum_{j \in [n] \cup \{CR, RL, D\}} T_{ij}^0 V_j \right), r_{W,i} + \lambda \cdot r_{PT}\}, \forall i \in [n], \\ F(V)_{n+i} &= r_{PT,i}, \forall i \in [n], \\ F(V)_{2n+1} &= r_{CR}, F(V)_{2n+2} = r_{RL}, F(V)_{2n+3} = r_D. \end{aligned}$$

The function F is the Bellman operator associated for our single-patient MDP with transition kernel \mathbf{T} , when the rewards are not uniform. Similarly as in the proof of Theorem 1, we will prove by induction that the sequence of policies $(\pi^t)_{t \geq 1}$ generated by the value iteration algorithm is threshold for all periods $t \geq 1$.

At $t = 1$, for $i \in [n]$,

$$\begin{aligned} F(\mathbf{V})_i &= \arg \max \{ r_{W,i} + \lambda \cdot \left(\sum_{j \in [n] \cup \{CR, RL, D\}} T_{ij}^0 V_j^0 \right), r_{W,i} + \lambda \cdot r_{PT,i} \} \\ &= \arg \max \{ r_{W,i} + \lambda \cdot 0, r_{W,i} + \lambda \cdot r_{PT,i} \}. \end{aligned}$$

We can conclude that π^1 is a threshold policy, and its threshold is 1.

Let us assume that $\pi^t = \pi^{[\tau]}$ for some $t \geq 1$ and some threshold $\tau \in [n+1]$. Let $i \in [n-1]$ such that $\pi^{t+1}(i) = 1$. It holds that

$$r_{W,i} + \lambda \cdot \left(\sum_{j=1}^n T_{ij}^0 V_j^t + out(i) \right) < r_{W,i} + \lambda \cdot r_{PT,i}.$$

Moreover, since π^t is $\pi^{[\tau]}$, by construction the vector \mathbf{V}^t is such that

$$V_\ell^t > r_{W,\ell} + \lambda \cdot r_{PT,\ell}, \forall \ell \in \{1, \dots, \tau-1\}, \quad (\text{F.5})$$

$$V_\ell^t = r_{W,\ell} + \lambda \cdot r_{PT,\ell}, \forall \ell \in [\tau, n], \quad (\text{F.6})$$

$$V_{n+\ell}^t = r_{PT,\ell}, \forall \ell \in [n], \quad (\text{F.7})$$

$$V_{2n+1}^t = r_{CR}, V_{2n+2}^t = r_{RL}, V_{2n+3}^t = r_D. \quad (\text{F.8})$$

We know that

$$F(\mathbf{V}^t)_\ell \in \arg \max \{ r_{W,\ell} + \lambda \cdot \sum_{j \in [n] \cup \{CR, RL, D\}} T_{\ell,j}^0 V_j^t, r_{W,\ell} + \lambda \cdot r_{PT,\ell} \}, \forall \ell \in [n].$$

We will prove that

$$r_{W,i+1} + \lambda \cdot \sum_{j \in [n] \cup \{CR, RL, D\}} T_{i+1,j}^0 V_j^t < r_{W,i+1} + \lambda \cdot r_{PT,i+1}.$$

We have

$$r_{W,i+1} + \lambda \cdot r_{PT,i+1} = r_{W,i} + \lambda \cdot r_{PT,i} + r_{W,i+1} + \lambda \cdot r_{PT,i+1} - r_{W,i} - \lambda \cdot r_{PT,i} \quad (\text{F.9})$$

$$\geq r_{W,i} + \lambda \cdot \sum_{j \in [n] \cup \{CR, RL, D\}} T_{ij}^0 V_j^t + r_{W,i+1} + \lambda \cdot r_{PT,i+1} - r_{W,i} - \lambda \cdot r_{PT,i} \quad (\text{F.10})$$

$$\geq r_{W,i} + \lambda \cdot \sum_{j=1}^n T_{ij}^0 V_j^t + \lambda \cdot out(i) + r_{W,i+1} + \lambda \cdot r_{PT,i+1} - r_{W,i} - \lambda \cdot r_{PT,i} \quad (\text{F.11})$$

$$\geq r_{W,i} + \lambda \cdot \sum_{j=1}^{\tau-1} T_{ij}^0 V_j^t + \lambda \cdot \sum_{j=\tau}^n T_{ij}^0 V_j^t + \lambda \cdot out(i) \quad (\text{F.12})$$

$$+ r_{W,i+1} + \lambda \cdot r_{PT,i+1} - r_{W,i} - \lambda \cdot r_{PT,i}$$

$$\geq r_{W,i} + \lambda \cdot \sum_{j=1}^{\tau-1} T_{ij}^0 V_j^t + \lambda \cdot \sum_{j=\tau}^n T_{ij}^0 (r_{W,j} + \lambda \cdot r_{PT,j}) + \lambda \cdot out(i) \quad (\text{F.13})$$

$$+ r_{W,i+1} + \lambda \cdot r_{PT,i+1} - r_{W,i} - \lambda \cdot r_{PT,i}$$

$$\geq r_{W,i} + \lambda \cdot \sum_{j=1}^{\tau-1} T_{ij}^0 (r_{W,j} + \lambda \cdot r_{PT,j}) + \lambda \cdot \sum_{j=1}^n T_{ij}^0 (r_{W,j} + \lambda \cdot r_{PT,j}) + \lambda \cdot out(i) \quad (\text{F.14})$$

$$+ r_{W,i+1} + \lambda \cdot r_{PT,i+1} - r_{W,i} - \lambda \cdot r_{PT,i}$$

$$\geq r_{W,i} + \lambda \cdot \sum_{j=1}^{\tau-1} T_{i+1,j}^0 (r_{W,j} + \lambda \cdot r_{RL}) + \lambda \cdot \sum_{j=\tau}^n T_{i+1,j}^0 (r_{W,j} + \lambda \cdot r_{PT,j}) - \lambda \cdot r_{PT,i+1} \quad (\text{F.15})$$

$$+ \lambda \cdot out(i+1) + r_{W,i+1} + \lambda \cdot r_{PT,i+1} - r_{W,i}$$

$$\geq r_{W,i+1} + \lambda \cdot \sum_{j=1}^{\tau-1} T_{i+1,j}^0 V_j^t + \lambda \cdot \sum_{j=\tau}^n T_{i+1,j}^0 (r_{W,j} + \lambda \cdot r_{PT,j}) + \lambda \cdot out(i+1) \quad (\text{F.16})$$

$$\geq r_{W,i+1} + \lambda \cdot \sum_{j \in [n] \cup \{CR, RL, D\}} T_{i+1,j}^0 V_j^t, \quad (\text{F.17})$$

where Inequality (F.10) follows from $\pi^{t+1}(i) = 1$, Inequality (F.13) follows from the fact that the policy π^t is a threshold policy and therefore the vector \mathbf{V}^t satisfies Inequalities (F.5). Inequality (F.15) follows from Assumption 5, because Assumption 5 and $\forall \ell \in [10], r_{PT,\ell} \leq r_{RL}$ imply that for all threshold $\tau \in \{1, \dots, n+1\}$,

$$\begin{aligned} \sum_{j=1}^n T_{ij}^0 (r_{W,j} + \lambda \cdot r_{PT,j}) + out(i) - r_{PT,i} &\geq \sum_{j=1}^{\tau-1} T_{i+1,j}^0 (r_{W,j} + \lambda \cdot r_{RL}) + \lambda \cdot \sum_{j=\tau}^n T_{i+1,j}^0 (r_{W,j} + \lambda \cdot r_{PT,j}) \\ &\quad + \lambda \cdot out(i+1) - \lambda \cdot r_{PT,i+1} - r_{W,i}. \end{aligned}$$

In order to show F.16, we can extend Lemma 1 to the case of non-uniform rewards:

$$\frac{1}{1-\lambda} \max_{i \in [n]} r_{W,i} \leq \min_{i \in [n]} r_{W,i} + \lambda \cdot r_{RL} \Rightarrow V_i^t \leq V_i^{\pi^*} \leq r_{W,i} + \lambda \cdot r_{RL}, \forall i \in [n]. \quad (\text{F.18})$$

Therefore, we conclude that

$$r_{W,i+1} + \lambda \cdot r_{PT,i+1} \geq r_{W,i+1} + \lambda \cdot \sum_{j \in [n] \cup \{CR, RL, D\}} T_{i+1,j}^0 V_j^t,$$

and $\pi^{t+1}(i+1) = 1$. We can thus conclude that the policy π^{t+1} is a threshold policy, and similarly as in the proof of Theorem 1, there exists an optimal nominal policy that is a threshold policy. \square

Appendix G: Details about the nominal matrix.

Confidence intervals. We use the method in Sison and Glaz (1995) to compute 95% confidence intervals around the nominal matrix \mathbf{T}^0 . This method yields

$$[T_{ij}^0 - \alpha_i, T_{ij}^0 + 2 \cdot \alpha_i], \forall (i, j) \in [10] \times [13], \boldsymbol{\alpha} = 10^{-4} \cdot (4, 8, 10, 14, 15, 43, 46, 47, 46, 45). \quad (\text{G.1})$$

We notice that the confidence intervals are larger for small severity scores than for larger severity scores (up to one order of magnitude). This is because large severity scores correspond to more serious health conditions, which are less likely to be observed than smaller severity scores.

Details on nominal factor matrix. We want to know if the errors between \mathbf{T}^0 and $\hat{\mathbf{T}}$ are more important for some severity scores than others. Therefore, we compute the maximum absolute and relative deviations between each row of $\hat{\mathbf{T}}$ and each row of \mathbf{T}^0 . In general, we notice that the absolute errors are higher for high severity scores. For instance, for severity score 1 the maximum absolute error is 0.0023. On the other hand, for severity score of 9, the maximum error is 0.0049. However, the maximum *relative* error is higher for low severity scores (from 1 to 5). Even though the absolute deviations are small, they amount to large relative deviations because they occur on coefficients that are already small. For instance, $T_{1,7}^0 = 5.40 \cdot 10^{-5}$ and $\hat{T}_{1,7} = 6.23 \cdot 10^{-5}$, which gives a relative deviation of about 15%, even though the absolute deviation is in the order of 10^{-5} .

Appendix H: Sensitivity Analysis for our single-parameter MDP and our hospital simulations.

In this section we present a detailed sensitivity analysis for our single-patient MDP and our hospital simulations. We have mentioned that the *ordering* of the rewards of our single-patient MDP can be inferred from the outcomes of the patients and the use of ICU resources (see (5.4)). We consider the impact of a change in the reward parameters presented in Section 5.3.1. We have seen in Section 5.3.2 that for this setting of rewards, the optimal nominal policy is $\pi^{[6]}$, while the optimal robust policy is $\pi^{[5]}$ (for \mathcal{U}_{\min} and \mathcal{U}_{emp}). We choose to study the variations in r_{RL} and r_{PT-RL} , since these rewards have reversed influences on the thresholds of the optimal policies. Indeed, r_{RL} is associated with a patient recovering from the ward, i.e., a patient who has not been proactively transferred, while r_{PT-RL} is the reward associated with a patient recovering *after* having been proactively transferred.

We present in Table 3 and Table 4 the variations in the hospital performances of the optimal robust policies, for changes in the value of r_{RL} (from 240 to 260) and in the value of r_{PT-RL} (from 180 to 200). We would like to note that both the thresholds of the optimal robust policies and the associated worst-case transition matrix may vary when the rewards parameters change.

Table 3 Change in the worst-case mortality, LOS and average ICU occupancy of the optimal robust policies for the uncertainty sets \mathcal{U}_{\min} , \mathcal{U}_{emp} and \mathcal{U}_{sa} , for variations in the reward r_{PT-RL} , associated with a patient who has been proactively transferred to the ICU and recovers.

r_{PT-RL}	Mortality (%)			LOS (days)			ICU occupancy (%)		
	\mathcal{U}_{\min}	\mathcal{U}_{emp}	\mathcal{U}_{sa}	\mathcal{U}_{\min}	\mathcal{U}_{emp}	\mathcal{U}_{sa}	\mathcal{U}_{\min}	\mathcal{U}_{emp}	\mathcal{U}_{sa}
180	5.73	6.47	7.02	3.98	4.07	4.06	71.88	75.34	77.25
185	5.57	6.35	7.02	3.94	4.04	4.06	72.34	75.62	77.34
190	5.11	5.83	6.40	3.75	3.85	3.85	73.95	76.70	77.80
195	5.11	5.46	6.02	3.75	3.70	3.73	73.94	77.98	78.96
200	4.78	5.12	5.64	3.59	3.53	3.57	75.39	79.03	80.02

Table 4 Change in the worst-case mortality, LOS and average ICU occupancy of the optimal robust policies for the uncertainty sets \mathcal{U}_{\min} , \mathcal{U}_{emp} and \mathcal{U}_{sa} , for variations in the reward r_{RL} , associated with a patient who is in the ward and exits the hospital recovering without entering the ICU.

r_{RL}	Mortality (%)			LOS (days)			ICU occupancy (%)		
	\mathcal{U}_{\min}	\mathcal{U}_{emp}	\mathcal{U}_{sa}	\mathcal{U}_{\min}	\mathcal{U}_{emp}	\mathcal{U}_{sa}	\mathcal{U}_{\min}	\mathcal{U}_{emp}	\mathcal{U}_{sa}
240	4.78	5.46	6.02	3.60	3.70	3.73	75.4	77.97	78.96
245	5.11	5.82	6.41	3.75	3.85	3.87	73.95	76.70	77.81
250	5.11	5.83	6.40	3.75	3.85	3.85	73.95	76.70	77.80
255	5.57	6.36	6.41	3.94	4.04	3.87	72.33	75.63	77.81
260	5.57	6.36	7.02	3.94	4.04	4.06	72.33	75.63	77.24

In both Table 3 and Table 4, we notice that the hospital performances of the optimal robust policies can vary when the set of reward parameters of our single-patient MDP does change. However, these changes are mostly due to the fact that we are comparing the optimal robust policies for different values of the rewards, and therefore the thresholds of the policies that we are comparing do vary. For instance, for \mathcal{U}_{\min} , the optimal robust policy is $\pi^{[8]}$ for $r_{PT-RL} = 240$ but it is $\pi^{[4]}$ for $r_{PT-RL} = 260$. On the contrary, we notice that when the optimal robust threshold is the

same, *variations in rewards value do not impact the hospital worst-case performances*. For instance, the optimal robust policies are the same (equal to $\pi^{[5]}$ across all three uncertainty sets) when $r_{RL} = 250$ and when $r_{RL} = 245$. We can see in Table 4 that the hospital worst-case performances (mortality, LOS, ICU occupancy) are also the same, when $r_{RL} = 250$ and when $r_{RL} = 245$. However, these two rows of worst-case performances are computed for *different* worst-case matrices (since the reward parameters for our single-patient MDP were different).

Therefore, we can conclude that even though the optimal robust and nominal policies can vary in our single-patient MDP, the hospital performances remain stable for each threshold policy individually. The ordering of the rewards does yield worst-case transition matrices for our single-patient MDP that are *good* and *robust* candidate worst-case transition matrices for the hospital worst-case performances, since variations in the rewards parameters on the single-patient MDP side still yield worst-case hospital performances that are very similar.

Appendix I: Numerical results for rank $r = 7$.

In this section we present our numerical results for the performances of the hospital, when the NMF approximation $\hat{\mathbf{T}}$ is of rank $r = 7$.

Errors of the NMF approximations. For $r = 7$, we compute a new $\hat{\mathbf{T}}$ solution of the NMF program (5.2). Our solution $\hat{\mathbf{T}} = \mathbf{U}\hat{\mathbf{W}}^\top$ achieves the following errors: $\|\mathbf{T}^0 - \hat{\mathbf{T}}\|_1 = 0.1932$, $\|\mathbf{T}^0 - \hat{\mathbf{T}}\|_\infty = 0.0224$, $\|\mathbf{T}^0 - \hat{\mathbf{T}}\|_{\text{relat}, \mathbf{T}^0} = 0.4093$. In more details, $\hat{\mathbf{T}}$ achieves the following errors.

	max.	mean	median	95% percentile
absolute deviation	0.0224	0.0015	0.0004	0.0069
relative deviation	0.4093	0.0856	0.0432	0.3247

Table 5 Statistics of the absolute and relative deviations of $\hat{\mathbf{T}}$ from \mathbf{T}^0 for a rank $r = 7$.

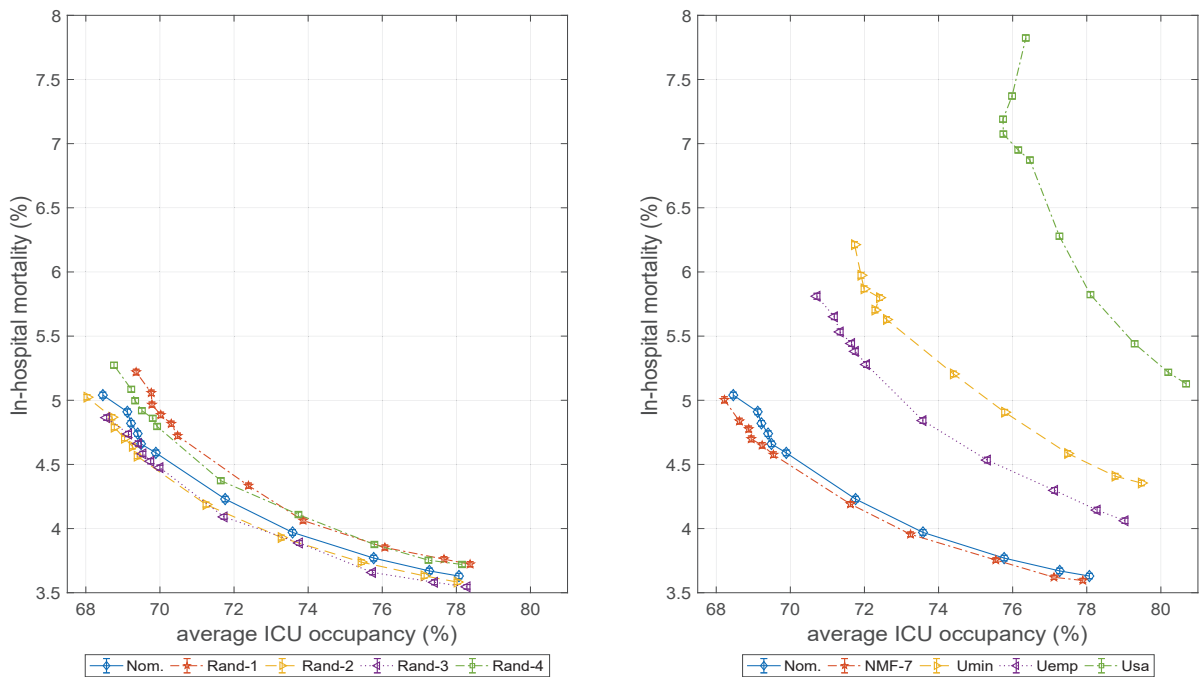
As we can see in Table 5, the absolute deviations remains small. Additionally, the relative differences between the coefficients are moderate, with half being less than 8.56%. That said, the maximum relative different is 40.93%. This occurs with $\hat{T}_{4,6} = 0.0035$, while $T_{4,6}^0 = 0.0060$; so while the relative deviation is quite large, the absolute variation is only in the order of 10^{-3} .

Mortality and Length-Of-Stay. We present the worst-case performances of the 11 threshold policies, for our uncertainty sets \mathcal{U}_{\min} and \mathcal{U}_{emp} , when the rank is $r = 7$. For references we still show the performances for the nominal transition kernel \mathbf{T}^0 (nominal performances), for the uncertainty set \mathcal{U}_{sa} and for our NMF solution of rank $r = 7$.

We first note that the hospital performances with our NMF approximation of rank $r = 7$ are very close to the hospital performances for \mathbf{T}^0 , which provides support that $\hat{\mathbf{T}}$ is a plausible transition

matrix. We notice that the performances of the threshold policies can still significantly deteriorate, even for small variations from the nominal matrix \mathbf{T}^0 . In particular, there is a 20% increase in the average mortality, for some worst-cases matrices in \mathcal{U}_{\min} and \mathcal{U}_{emp} . Interestingly, the uncertainty set \mathcal{U}_{\min} yields worst-case mortality rates that are higher than for worst-cases matrices in \mathcal{U}_{emp} , contrary to what we noticed in Section 5 for rank $r = 8$. However, these two uncertainty sets still yield the same insights, which are that the performances can significantly degrade even for small deviations, and that in worst-case, the initial decrease for proactively transferring the patients with the highest severity scores (policy $\pi^{[11]}$, top-left of each curve, to policy $\pi^{[6]}$, the sixth point of each curve, starting from the left) is steeper than the initial decrease for the nominal performances. Moreover, these insights are still different from the worst-cases performances in \mathcal{U}_{sa} , since the results for \mathcal{U}_{sa} are independent of the rank chosen for our NMF approximation. In particular, for worst-cases in \mathcal{U}_{sa} , the decision-maker appears to be able to proactively transfer the patients with severity scores in $\{8, 9, 10\}$, *without* increasing the ICU occupancy.

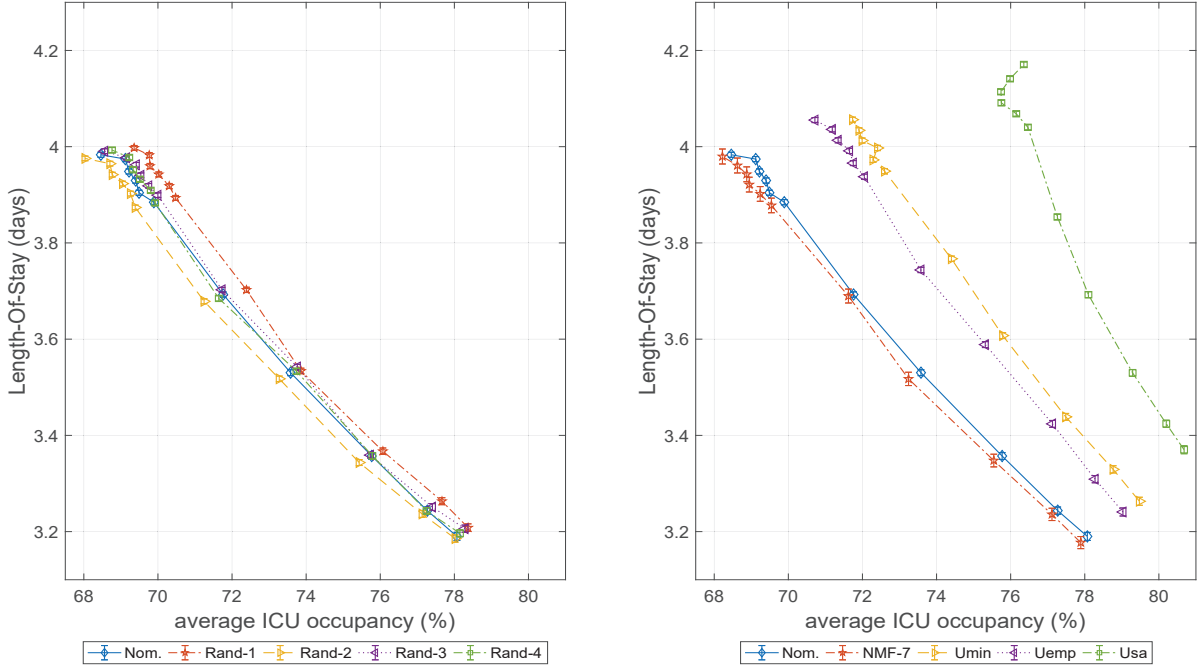
Therefore, our numerical simulations for rank $r = 7$ are corroborating our numerical simulations of Section 5 for rank $r = 8$. We do not present the hospital simulations for lower ranks, since the NMF approximations become very poor for rank r lower than 7. For instance, for $r = 6$, there are 54 coefficients (out of 130 coefficients) outside of the confidence intervals, and for a rank $r = 5$, our NMF solution has 70 coefficients that are outside the 95% confidence intervals.



(a) Random samples analysis.

(b) Worst-case analysis.

Figure 7 In-hospital mortality of the 11 threshold policies for the nominal estimated matrix, randomly sampled matrices in the 95% confidence intervals (left-hand side), and the worst-case matrices found by our single MDP model (right-hand side). Each point corresponds to a threshold policy: the policy with highest mortality rate corresponds to threshold $\tau = 11$ (top-left of each curve) and the threshold decreases until the bottom-right point of each curve, corresponding to threshold $\tau = 1$. We consider the uncertainty sets \mathcal{U}_{\min} , \mathcal{U}_{emp} and \mathcal{U}_{sa} when the rank $r = 7$. On the right-hand side figure, we also report the hospital mortality rate when the transition matrix is our NMF approximation with rank 7.



(a) Random samples analysis.

(b) Worst-case analysis.

Figure 8 Length-Of-Stay of the 11 threshold policies for the nominal estimated matrix, randomly sampled matrices in the 95% confidence intervals (left-hand side), and the worst-case matrices found by our single MDP model (right-hand side). Each point corresponds to a threshold policy: the policy with highest LOS corresponds to threshold $\tau = 11$ (top-left of each curve) and the threshold decreases until the bottom-right point of each curve, corresponding to threshold $\tau = 1$. We consider the uncertainty sets $\mathcal{U}_{\min}, \mathcal{U}_{\text{emp}}$ and \mathcal{U}_{sa} when the rank $r = 7$. On the right-hand side figure, we also report the hospital mortality rate when the transition matrix is our NMF approximation with rank 7.

Appendix J: Hospital performances for random deviations around the nominal kernel.

We sample at random 20 matrices in the confidence intervals (5.1). In order to do so, we first sample a matrix of deviations $\mathbf{D} \in \mathbb{R}^{10 \times 13}$, with $D_{ij} \in [-\alpha_i, +2 \cdot \alpha_i], \forall (i, j) \in [10] \times [13]$. Note that the matrix $\mathbf{T}^0 + \mathbf{D}$ is not necessarily a transition matrix, because each of its row does not necessarily sum up to 1. Therefore, we project each of the rows of the matrix $\mathbf{T}^0 + \mathbf{D}$ onto the simplex and we obtain a matrix a new matrix $\tilde{\mathbf{T}}$. If the corresponding matrix $\tilde{\mathbf{T}}$ is inside the confidence-intervals,

we compute the hospital performances of the 11 threshold policies; otherwise, we reject \tilde{T} and sample a new deviation matrix D .

Using this method, we compute the performances of the threshold policies for 20 matrices chosen randomly inside the confidence intervals (5.1). Out of these 20 performances, 8 were pessimistic (higher mortality / Length-Of-Stay / ICU occupancy than in the nominal case) and 12 were optimistic (lower mortality / Length-Of-Stay / ICU occupancy than in the nominal case)

Mortality. For the in-hospital mortality, the average relative deviations from the nominal performances ranged from 3.00% to 3.84% (from threshold 0 to threshold 11). For each policy, the maximum relative deviation from the nominal performances ranged from 6.19% to 8.82% (again for threshold 0 to threshold 11).

LOS. For the Length-Of-Stay, the average relative deviations from the nominal performances ranged from 0.34% to 1.04% (for threshold 3 and threshold 11). For each policy, the maximum relative deviation from the nominal performances ranged from 0.91% to 2.79% (for threshold 0 and threshold 11).

ICU occupancy. For the average ICU occupancy, the average relative deviations from the nominal performances ranged from 0.37% to 0.57% (for threshold 0 and threshold 11). For each policy, the maximum relative deviation from the nominal performances ranged from 0.99% to 1.57% (for threshold 11 and threshold 0).