Robustness of proactive ICU transfer policies

Julien Grand-Clément
Information Systems and Operations Management Department, HEC Paris, grand-clement@hec.fr
Carri W. Chan
Columbia Business School, Columbia University, cwchan@columbia.edu
Vineet Goyal
Industrial Engineering and Operations Research Department, Columbia University vgoyal@ieor.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. 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 the important problem of 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. This is an important issue, which can lead to choosing significantly suboptimal policies. 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 are able to show that the robust policy also has a threshold structure under fairly general assumptions, and that 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* 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, 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 this work, as a first step towards better theoretical understanding of the pathways involved in early warning systems and 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 (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. 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 (see Section 6 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 this important real-world 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 performance 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-Clément (2018)) to develop a theoretical and empirical basis for our proposed transfer policies.

Our main contributions, both methodological and practical, 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". We rigorously show that our single-patient MDP can be interpreted as a relaxation of a more expressive, but intractable, multi-patient MDP which directly incorporates the ICU capacity constraints in the decision-making.

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, under some mild assumptions, there exists an optimal policy that transfers all patients above a certain severity score. This threshold structure is particularly important because of its interpretability and implementability. In the event the assumptions do not hold, we bound the extend of the suboptimality of the class of threshold policies based on the amount of violation of our main assumptions.

Robustness of transfer policies. Building upon the nominal model, we incorporate the real-world limitations of parameter uncertainty, by considering parameter misspecification for the transition matrix. We take a robust optimization approach, where the goal is to compute a robust transfer policy with guarantees of good performance on a set of all plausible transition matrices. Naturally, the conservativeness of a robust transfer policy is directly related to the choice of transition matrices that are considered "plausible". Prior approaches to parameter uncertainty (Iyengar 2005, Wiesemann et al. 2013) assume *rectangularity*, where transitions out of different health states are unrelated. In a healthcare setting, the health evolution of the patients is likely to be dictated

by underlying factors, such as genetics, demographics, and/or physiologic characteristics of certain diseases, which have a common influence on the transitions across different health states. Therefore, rectangular models may result in overly conservative estimations of the worst-case performance of the transfer decisions. In contrast, we consider a model of uncertainty where the transition probabilities of different health states are related and depend on a factor model (Goh et al. 2018, Goyal and Grand-Clément 2018). In this model of uncertainty, a common but uncertain *factor matrix* influences the transition probabilities out of every health state, resulting in a more constrained set of possible transition rates that lends itself to less conservative performance estimations.

We present an efficient algorithm to compute an *optimal robust policy* that maximizes the worst-case possible outcomes over all plausible transition matrices. Moreover, we prove structural results for the optimal robust policy and compare it to the nominal policy. In particular, an optimal robust policy is always deterministic and – under the same assumptions as in our nominal model – of threshold type. This is in contrast to the general situation in robust MDPs, where there may not exist a optimal robust policy that is deterministic. Additionally, we show that 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. We also introduce a robust version of the *Whittle index*, which characterizes the monotonicity of the optimal threshold of the robust policy as regards the beneficial impact of proactive transfers. To the best of our knowledge, our paper is the first to explore robust Whittle indices and to prove the indexability of a robust MDP.

Numerical experiments. We present detailed numerical experiments to compare the performance 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 polices could have substantial differences

in the proportion of patients who are proactively transferred. When considering the full hospital model, we observe similar deterioration in performance (as measured by mortality, length-of-stay, average ICU occupancy) even for small parameters deviations. Additionally, we find that factor model of uncertainty in the transition matrix results in different and more useful insights than when considering rectangular uncertainty which can be overly conservative by optimizing for worst-case scenarios that are extremely unlikely. We also highlight the contrast between this worst-case analysis and more standard sensitivity analysis approaches. Our work suggests, even in the worst-case, proactively transferring the patients with the riskiest severity scores has the potential to improve the hospital mortality and LOS, without significantly increasing the ICU occupancy.

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 6, 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 $n \in \mathbb{N}$, we denote by $[n]$ the set $\{1, ..., n\}$. 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 $\boldsymbol{e}$ 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 (Chrusch et al. 2009a, Kc and Terwiesch 2012) and theoretically (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. Threshold policies have been investigated in various admission control settings (including the ICU), but most prior works either focus on a simple transition model (e.g., a patient in severity class i can only transition to severity class $i+1$ or $i-1$), or only focus on the *empirical* performance of threshold policies (e.g., Barron (2016, 2018) for related approaches in machine maintenance). Altman et al. (2001) prove the optimality of threshold policies, but rely on a submodularity assumption, which is less interpretable than ours in a healthcare setting. In contrast to these works, we consider a setting where patients can be admitted to the ICU at any point while they are in the general medical/surgical ward and can transition to any other severity condition. In such a setting we give interpretable conditions for the optimality of threshold policies.

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, including LOS, death and readmission rates to the ICU. For instance, Putnam et al. (2002) develop a risk-adjustment metric to predict patient hospitalization. Bilben et al. (2016) study the performance 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, while Brunelli et al. (2008) develop a scoring

system for predicting ICU admission after major lung resection. We refer to Rapsang and Shyam (2014) for a review of the medical severity scores for ICU patients.

The operations community has studied how 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 (Kipnis et al. 2016). The particular problem we study is related to that in Hu et al. (2018), where the authors use 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. Most of the healthcare literature primarily focuses on sensitivity analysis of the selected policy. Unfortunately, sensitivity analysis is unable to capture possible *adversarial* deviations of these parameters and, as we will show in this work, such deviations can substantially impact system performance. In contrast, robust optimization explicitly accounts for the uncertainty (and adversarial deviations) when deriving good policies.

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. Among others, MDPs have been used 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 to Alagoz et al. (2010) for a review of applications of MDPs in healthcare.

Robust Markov Decision Processes. 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. Partially Observable MDPs (POMDPs) assume that the system dynamics are determined by an MDP, but the agent cannot fully observe the underlying states. Instead, the decisions must be based on the observed states. However, POMDPs are generally hard to solve (Madani et al. 1999), and robust POMDPs are known to be even harder to solve and may lead to randomized optimal policies (Rasouli and Saghafian 2018), making POMDPs unusable in our healthcare application. Robust MDPs address the issue of parameter misspecification in the MDP (Iyengar 2005, Nilim and El Ghaoui 2005, Wiesemann et al. 2013, 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. 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. In principle one could also use *distributionally* robust MDPs (Xu and Mannor 2010) to ameliorate the issues with parameter uncertainty, but this leads to harder optimization problems than the linear program of our own robust MDP formulations. Additionally, it is not clear how to build ambiguity sets around the nominal density estimation for distributionally robust MDPs, while it is reasonably easy to incorporate confidence intervals in the uncertainty sets of robust MDPs.

Robust Markov Decision Processes 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 cholesterol management. However, computing the optimal robust policy of a 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-Clément (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. They show how to compute the worst-case reward for a given policy, and apply these methods to a cost-effectiveness analysis of fecal immunochemical testing for detecting colorectal cancer. Goyal and Grand-Clément (2018) show that for a factor matrix uncertainty set, one can also compute the optimal robust policy, i.e., the policy that maximizes the worst-case reward. They also prove important structural properties on the optimal value functions, which we can use to prove structural properties on the optimal nominal and robust policies. Our work builds upon Goh et al. (2018) and Goyal and Grand-Clément (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. The estimation errors in the transitions rates for different health states may be related, as they are influenced by the characteristics of the patients (e.g., demographics and/or comorbidities). Overlooking this fact may yield transition matrices that are unlikely to be realized in practice. Since the robust policy attempts to achieve good performance on every transition matrix in the uncertainty set, a rectangular uncertainty set may yield robust policies that take into account overly pessimistic realizations of the parameters, resulting in misleading insights and poor performance when deployed. In contrast, in the factor matrix uncertainty set, all plausible transition probabilities depend on a common underlying factor matrix, which is itself uncertain. The factor matrix uncertainty set is less conservative to model the parameter uncertainty in our healthcare setting, as a deviation in the factor matrix results in a constrained deviation for the transition probabilities out of all states, which is impossible to capture in the classical rectangular model of uncertainty. More specifically, we are able to (i) compute an optimal robust policy, and (ii) give theoretical guarantees on the *structure* of the optimal policies, namely the optimal nominal

and the optimal robust policies both have a threshold structure. In contrast, while Goh et al. (2018) consider a similar uncertainty formulation and present an algorithm to compute the *worst-case* performance of a given policy, they do not consider the problem of finding *optimal* robust policies. Building upon the results in Goyal and Grand-Clément (2018), we focus on the specific problem of proactive transfer to the ICU. Our single-patient and multi-patient MDP models, as well as our interpretable assumptions, appear to be novel in the literature. Our detailed numerical experiments with our hospital model emphasize the advantage of our robustness analysis (compared to a simple sensitivity analysis) in highlighting the potential detrimental impact of adversarial parameter deviations. We also present extensive details into the construction of a factor matrix uncertainty sets directly from the data, in contrast to the simulations in Goyal and Grand-Clément (2018) which consider that the factor matrix uncertainty set is already given.

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 $C < +\infty$.

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 $i \in \{1, ..., n\}$, 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 $\boldsymbol{T}^0 \in \mathbb{R}^{n \times (n+3)}$.

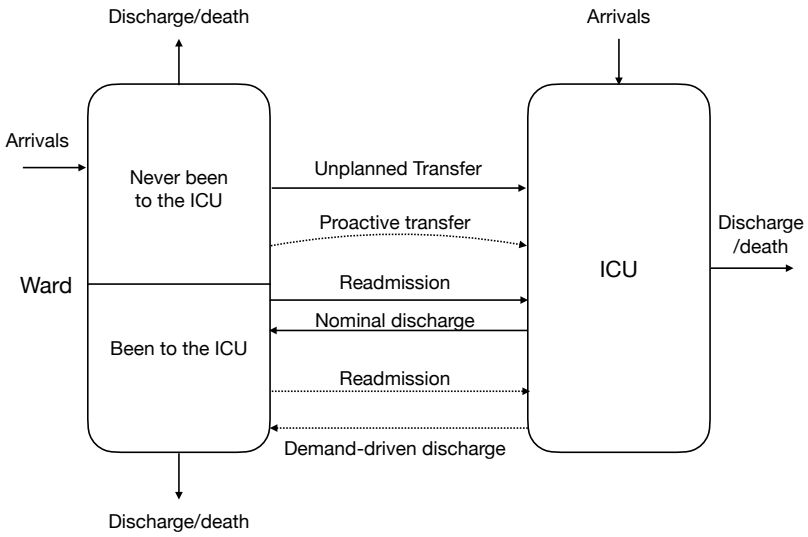


Figure 1 Simulation model for the hospital.

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^0_{ij}. In addition, one of the three following events may happen at the end of each time slot:

1) With probability $T^0_{i,n+1}$, the patient may *crash* and require a *reactive* transfer to the ICU.

2) With probability $T^0_{i,n+2}$, the patient may fully recover and leave the hospital.

3) With probability $T^0_{i,n+3}$, the patient may die.

We verify the consistency of this approach by comparing the empirical probability of crashing, dying in the ward, and surviving to hospital discharge in the data with the results of our hospital simulations using our Markov chain. We find that with our parameters calibrated as in Appendix A, the key metrics are comparable. For instance the mortality rate in the ward is 1.93 % in our simulations, which is comparable to the 2.2 % computed from the data directly.

Reactive transfer to the ICU. If a patient crashes on the ward, s/he will be admitted to the ICU. Upon ICU admission, the patient's remaining hospital Length-Of-Stay (LOS) is modeled as being log-normally 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 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. In Chan et al. (2012), the ratio ρ_D/ρ_E fluctuates between 1.11 and 1.18. Therefore, we set $\rho_D = 1.15 \cdot \rho_E$. Note that this choice of parameters is stress-tested in the numerical experiments in Section 5 and Appendix B in Hu et al. (2018).

Remark 1 (Dynamics of patients in the ward). We do not model the health dynamics of patients discharged from the ICU with a Markov chain. This would require using a transition matrix different than $\boldsymbol{T}^0$, since these patients are sicker than the population of patients who have not been to the ICU. This would have two issues. First, there are many fewer observations for patients who have been to the ICU than those who have never been to the ICU, and this would result in unreasonably large confidence intervals for the coefficients of the transition matrix. Second, and perhaps most importantly, using a different Markov chain would not allow us to explicitly capture the impact of ICU discharge on LOS/mortality (whereas our current approach does). Therefore, we find it more relevant to directly use simpler statistics of mortality rate and LOS to summarize the impact of ICU admission onto these patients, rather than using a different Markov chain.

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 log-normally 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 $\rho_{A,i} = \rho_C$, as these two types of patients are transferred to the ICU from the ward, in contrast to direct admits patients.

Note that in practice when a patient reaches an alert threshold, s/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. 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 log-normally 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. Note that proactive transfers to the ICU can only happen when there are beds available in the ICU. If the ICU is full, no proactive transfers to the ICU can be performed, even if the policy recommends a proactive transfer for a patient.

2.3. Challenges and limitations of the hospital model.

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, e.g., Puterman (1994)), the state space is prohibitively large. For instance, with N patients in the ward and n severity scores, one needs a state space of cardinality n^N 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 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 simulations[1].

Parameter uncertainty. The hospital model is specified by the stochastic processes detailed above, including a transition matrix $\boldsymbol{T}^0 \in \mathbb{R}_+^{n\times(n+3)}$. 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 $\boldsymbol{T}^0$ and therefore,

[1] For 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. 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. We will connect the insights developed from this model to one with multiple patients in Section 5.

3.1. Single-patient MDP.

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

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

The states $i \in \{1, ...n\}$ 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 $i = 1, ..., n$, 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 $\pi : \mathbb{S} \to [0, 1]$, 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 $\pi(i) = 0$.

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.

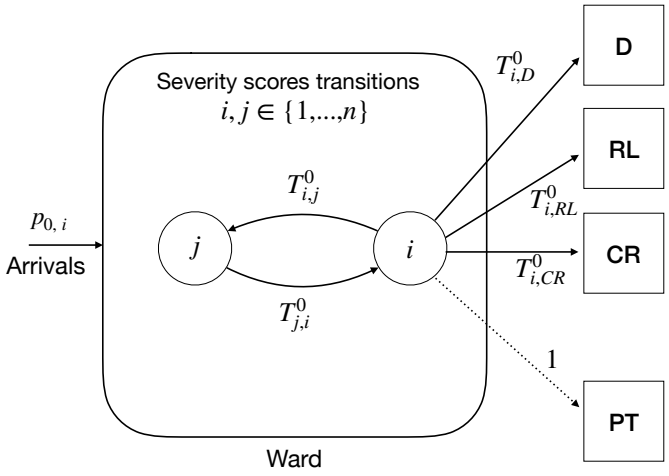


Figure 2 Single-patient MDP model. Terminal states are indicated as square. The patient arrives in the ward with a severity score of $i \in \{1, ..., n\}$ with an initial probability $p_{0,i}$. The solid arcs correspond to transitions where no transfer decision is taken (action 0), and the patient can transition to another severity score j with probability $T^0_{i,j}$ 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).

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 are 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. Every six hours (a *period*), 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^0_{ij} in the next 6 hours (where $\boldsymbol{T}^0$ is the transition matrix among severity scores). Alternatively, the patient can either transition to state CR with probability $T^0_{i,n+1}$, to state RL, with probability $T^0_{i,n+2}$ or the patient transitions to D with probability $T^0_{i,n+3}$. 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 $\theta = \min_{i\in[n]} \min\{T^0_{i,CR}, T^0_{i,RL}, T^0_{i,D}\} > 0$.

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 $R(\pi, \boldsymbol{T}^0) = E^{\pi,\boldsymbol{T}^0}\left[\sum_{t=0}^{\infty} \lambda^t r_{i_t a_t}\right]$, 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 function* $\boldsymbol{V}^\pi \in \mathbb{R}^{n+4}$, defined as

$$V^\pi_i = E^{\pi,\boldsymbol{T}}\left[\sum_{t=0}^{\infty} \lambda^t r_{i_t a_t} \,\middle|\, i_0 = i\right], \forall\, i \in \{1, ..., n+4\}. \tag{3.1}$$

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 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., $r_{i,0} = r_{i,1} = r_i, \forall\, i \in \mathbb{S}$, and the actions dictate 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.

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 (McClean and Millard 2006, Yauney and Shah 2018). Therefore, we focus on the relative *ordering* of these rewards, to capture the trade-off between better health outcomes by proactive transfers and increased congestion in ICU. First, the decision-maker prioritizes the outcome (survival versus death). Second, for a given outcome, the decision-maker favors the policy which uses the fewest ICU resources. These considerations imply the following ordering of the rewards. 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

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

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 the following mild assumption.

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

This 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. This is a natural assumption and implies 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 function)**. *Let $\boldsymbol{V}^\pi$ be the value function 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. In the next section, we will 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*

$$out(i) \geq out(i+1), \forall\, i \in [n-1]. \tag{3.2}$$

and we assume that

$$\frac{r_W + \lambda \cdot r_{PT}}{r_W + \lambda \cdot r_{RL}} \geq \frac{\left(\sum_{j=1}^n T^0_{i+1,j}\right)}{\left(\sum_{j=1}^n T^0_{ij}\right)}, \forall\, i \in [n-1]. \tag{3.3}$$

We can interpret condition (3.2) as follows. Condition (3.2) 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 with a given health condition is worse than the outside option of a patient with a better health condition, i.e., the healthier patient is more likely to leave the ward in a better state. Note that $T^0_{i,CR}$ and $T^0_{i,D}$ should be increasing in i and $T^0_{i,RL}$ should be decreasing in i. Therefore, $out(i) \geq out(i+1)$ is implicitly assuming that the reward r_{RL} (reward for recovering) is significantly larger than r_{CR} ("reward" for crashing) and r_D ("reward" for dying), as we expect to hold in practice.

We can interpret condition (3.3) as follows. Condition (3.3) assumes that the probability of staying in the system in risk score i, $\sum_{j=1}^{n} T_{ij}^0$, is non-increasing in severity score i. Additionally, the rate at which the probability 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 (3.2), we expect condition (3.3) 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. We are implicitly assuming that the increase in $T_{i,CR}^0 + T_{i,D}^0$ outweighs the decrease in $T_{i,RL}^0$ when i increases. We can expect this in practice if the changes in these coefficients are of the same order of magnitude. In particular, it holds for our dataset of nearly $300,000$ patients across 21 hospitals.

Note that both condition (3.2) and condition (3.3) are homogeneous: they hold if we scale all rewards by the same (positive) scalar. Moreover, condition (3.3) is invariant by translation, i.e., it holds if we add the same scalar to all rewards. However, this is not the case for condition (3.2) (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 relating Assumption 2 with the Bellman optimality equation for the value function of the optimal policy. 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. 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 (3.2) and (3.3) 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 (3.2)), while the probability to exit the ward is increasing (condition (3.3)). 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.

Note that the main assumption for Theorem 1 is Assumption 2. In particular, Assumption 2 may fail to hold, because of the inherent tension between condition (3.2), where $out(i) = r_{CR} \cdot T^0_{i,n+1} + r_{RL} \cdot T^0_{i,n+2} + r_D \cdot T^0_{i,n+3}$ is non-increasing, and condition (3.3), which implies that $T^0_{i,n+1} + T^0_{i,n+2} + T^0_{i,n+3}$ is non-decreasing. In the case that Assumption 2 fails to hold, there may not exist an optimal policy that is threshold in general. We provide an example of this in Appendix E. Still, it is possible to bound the suboptimality of the class of threshold policies if Assumption 2 is violated by at most $\epsilon > 0$ on each of the equations in condition (3.2) and condition (3.3). In particular, we have the following proposition.

PROPOSITION 1. *Consider that Assumption 1 holds but that Assumption 2 is violated by at most $\epsilon > 0$, in the sense that for all $i \in [n-1]$, we have*

$$\epsilon + out(i) \geq out(i+1), \tag{3.4}$$

$$\epsilon + (r_W + \lambda \cdot r_{PT}) \left(\sum_{j=1}^{n} T^0_{i,j} \right) \geq (r_W + \lambda \cdot r_{RL}) \left(\sum_{j=1}^{n} T^0_{i+1,j} \right). \tag{3.5}$$

Then there exists a threshold policy π such that

$$R(\pi) \leq R(\pi^\star) \leq R(\pi) + \lambda \frac{2(n-1)\epsilon}{1-\lambda},$$

where $\pi^\star$ is an optimal nominal policy in the single-patient MDP.

We provide the proof of Proposition 1 in Appendix F, where we construct a threshold policy π such that $\pi(i) = 1$ if there exists $i' \leq i$ for which $\pi^\star(i') = 1$. Note that the factor $1/(1-\lambda)$ results from the sequential nature of the single-patient MDP, where the future reward is discounted by a factor of λ compared to the current reward. The factor $2(n-1)\epsilon$ follows from the fact that conditions (3.4)-(3.5) are in fact $2(n-1)$ inequalities (one for each $i \in [n-1]$), each violated by at most $\epsilon > 0$, so that $2(n-1)\epsilon$ is the total amount of violation of Assumption 2 when conditions (3.4)-(3.5) are not satisfied. Also, it is straightforward to compute $\epsilon > 0$, given the values of the transitions and reward parameters. Overall, Proposition 1 precisely quantifies the suboptimality of the class of threshold policies and shows that this suboptimality is proportional to the level of violation of Assumption 2. We conclude this section with the following remarks.

Remark 2 (A more general assumption). Theorem 1 holds if we replace Assumption 2 by the following weaker, but less interpretable condition:

$$\left(\sum_{j=1}^{n} T^0_{ij}\right) \cdot (r_W + \lambda \cdot r_{PT}) + out(i) \geq \left(\sum_{j=1}^{n} T^0_{i+1,j}\right) \cdot (r_W + \lambda \cdot r_{RL}) + out(i+1), \forall\, i \in [n-1]. \quad (3.6)$$

Condition (3.6) is implied by Assumption 2, simply by summing up condition (3.2) and condition (3.3). Moreover, condition (3.6) is homogeneous and holds under translation of the rewards (see Lemma 3 in Appendix C). However, even though condition (3.6) is more general than Assumption 2, it is much less interpretable.

Remark 3 (Non-homogeneous rewards). By assumption, the rewards in the ward and for proactively transferring a patient do not depend of the current severity score. We can relax this assumption 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. It is straightforward to extend our proof for uniform rewards to the case of non-uniform rewards.

3.3. Challenges and limitations of the single-patient MDP.

In this section we discuss the main limitations of the single-patient MDP. These limitations relate to (a) the difficulty to account for ICU capacity directly in a tractable MDP model, and (b) what can happen when our main assumption (Assumption 2) is violated.

Incorporating the ICU capacity. Our single-patient MDP does not *explicitly* account for the ICU occupancy or the total ICU capacity. To account for capacity constraints, our single-patient MDP penalizes aggressive proactive transfers by assuming a lower reward r_{PT} compared to r_{RL}. Our single-patient MDP attempts to find a good balance between a model that is (i) complex and expressive enough to capture important trade-offs (e.g., transfer or wait), (ii) sufficiently tractable to allow studying structural properties of optimal policies, and (iii) simple enough so that we can evaluate its parameters with satisfying accuracy (see Section 6). In Section 5, we introduce a more general *multi-patient* MDP, which models the dynamics for $N \in \mathbb{N}$ patients in the hospital and accounts for ICU capacity through a constraint on the number of simultaneous proactive transfers. Because of the very large number of states and actions for a multi-patient model, this model is numerically intractable. However, we show how relaxing the constraint on the number of transfers in the multi-patient MDP yields a single-patient MDP with penalized reward for proactive transfers. This provides a rigorous justification for our single-patient model.

What happens when Assumption 2 is violated. Assumption 2 is at the core of our theoretical analysis of optimal nominal and optimal robust policies. We can interpret condition (3.2) and condition (3.3) as imposing monotonicity of two explainable quantities, the outside option $out(i)$ and the probability of staying in the ward $\sum_{j=1}^{n} T_{ij}^{0}$ for the severity condition $i \in [n]$. We note that to obtain simpler conditions, one would also require simpler dynamics (e.g., transitioning from severity condition i to either $i+1$ or $i-1$), which would greatly diminish the modeling power of the single-patient MDP. In Proposition 1, we characterize the performance of the class of threshold policies in single-patient MDP instances where Assumption 2 fails to hold. In particular, we show that if Assumption 2 is violated, the suboptimality of the class of threshold policies is bounded by

a factor proportional to the level of violation of Assumption 2. Note that it is straightforward to compute the level of violation of Assumption 2, given the values of the transition probabilities and the rewards. Overall, our results in Section 3.2 provide insights on threshold policies, both when Assumption 2 is satisfied (Theorem 1) and when it is violated (Proposition 1).

4. Robustness analysis of the single-patient MDP.

Thus far, we have assumed the health state of a patient evolves according to a Markov chain with known transition matrix $\boldsymbol{T}^0$. However, the parameters of the transition matrix are estimated from historical data and are subject to statistical estimation errors. 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 $\boldsymbol{T}^0$ is known so that one finds a policy that maximizes the expected nominal reward and solves the optimization problem: $\max_{\pi\in\Pi} R(\pi, \boldsymbol{T}^0)$. 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: $\max_{\pi\in\Pi} \min_{\boldsymbol{T}\in\mathcal{U}} R(\pi, \boldsymbol{T})$. A solution π^{rob} to this optimization problem 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-Clément 2018). In this model we consider that the transition probabilities are convex combination of some common factors, which can themselves be uncertain. This is in contrast to rectangular

uncertainty sets (Iyengar 2005, Wiesemann et al. 2013) that assume unrelated adversarial deviations. In particular, we assume that $\mathcal{U}$ is of the following form:

$$\mathcal{U} = \left\{ \boldsymbol{T} \in \mathbb{R}_+^{n\times(n+3)} \;\middle|\; T_{ij} = \sum_{\ell=1}^{r} u_i^\ell w_{\ell,j} \forall\, (i,j) \in [n]\times[n+3],\ \boldsymbol{w}_\ell \in \mathcal{W}^\ell, \forall\, \ell \in [r] \right\} \tag{4.1}$$

where $\boldsymbol{u}_1, ..., \boldsymbol{u}_n$ are fixed vectors in $\mathbb{R}_+^r$ and $\mathcal{W}^\ell, \ell = 1, ..., r$ are convex, compact subsets of $\mathbb{R}_+^{(n+3)\times r}$ such that $\sum_{\ell=1}^r u_i^\ell = 1,\ \forall\, i \in [n], \sum_{j=1}^{n+3} w_{\ell,j} = 1,\ \forall\, \ell \in [r], \forall\, \boldsymbol{w}_\ell \in \mathcal{W}^\ell$. To understand better the implications of this uncertainty set, consider $\boldsymbol{T}$, a transition matrix in our factor matrix uncertainty set $\mathcal{U}$. Each of the rows of the matrix $\boldsymbol{T}$ is a convex combination of the *factors* $\boldsymbol{w}_1, ..., \boldsymbol{w}_r \in \mathbb{R}^{n+3}$: $\boldsymbol{T}_i = u_i^1 \boldsymbol{w}_1 + ... + u_i^r \boldsymbol{w}_r$, for $i \in [n]$. Therefore, when one of the factor $\boldsymbol{w}_\ell$ varies in the set $\mathcal{W}^\ell$, this impacts *all* the rows of the matrix $\boldsymbol{T}$ at the same time in a coherent manner. Typically, the sets $\mathcal{W}^1, ..., \mathcal{W}^r$ can be polytopes (Iyengar 2005) or ellipsoidal uncertainty sets (Wiesemann et al. 2013). Also, note that for $\boldsymbol{U} = (u_i^\ell)_{(i,\ell)} \in \mathbb{R}^{n\times r}$ and $\boldsymbol{W} = (\boldsymbol{w}_1, ..., \boldsymbol{w}_r) \in \mathbb{R}^{(n+3)\times r}$, we have $\boldsymbol{T} = \boldsymbol{U}\boldsymbol{W}^\top$.

Comparison with rectangular uncertainty sets. The factor model of uncertainty 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 $\boldsymbol{T}$ are chosen independently, when $r = n$ (where n is the number of rows of $\boldsymbol{T}$). This corresponds to a *full-rank* deviation from the nominal parameter $\boldsymbol{T}^0$. Additionally, as different rows of a matrix $\boldsymbol{T} \in \mathcal{U}$ are convex combinations of the same r factor vectors, we can use our factor matrix uncertainty set to model relations between the probability distributions for different states when r is smaller than n. This corresponds to a *rank-constrained* deviation from the nominal parameter $\boldsymbol{T}^0$. This contrast between full-rank and rank-constrained deviations is the main difference between rectangular and factor matrix uncertainty sets, and this is unrelated to the geometry of the sets $\mathcal{W}^1, ..., \mathcal{W}^r$ (e.g., ellipsoidal or polytope). The full-rank deviations of a rectangular uncertainty set allows for a larger set of plausible transition matrices, resulting in robust policies that take into account more scenarios. This directly influences the conservativeness of the robust policy for rectangular uncertainty, which may be too pessimistic for our healthcare applications. It is conceivable that there exists some relations across the transitions from different severity conditions, and the factor matrix uncertainty set provides an approximate

way to model this through *rank-constrained* deviations from the nominal parameters. We explore the different level of conservativeness afforded by rectangular and factor matrix uncertainty sets in our numerical experiments in Section 6.

In Section 3, we assume that the nominal matrix $\boldsymbol{T}^0$ satisfies Assumption 2, while our uncertainty set $\mathcal{U}$ models small parameters variations around $\boldsymbol{T}^0$. Therefore, it is reasonable to have the following assumption.

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

REMARK 4 (VERIFYING ASSUMPTIONS 1-3). Assumptions 1-3 can be verified numerically. Assumption 1 is simply an inequality on the set of rewards. Similarly, verifying Assumption 2 requires the computation of the cumulative sums $\left(\sum_{j=1}^{n} T_{ij}\right)_{i\in[n]}$. To verify Assumption 3 for an uncertainty set $\mathcal{U}$, we can check the nonnegativity of

$$\min\left\{\frac{r_W+\lambda r_{PT}}{r_W+\lambda r_{RL}}\sum_{j=1}^{n} T_{ij} - \sum_{j=1}^{n} T_{i+1,j} \,\middle|\, \boldsymbol{T}\in\mathcal{U}\right\},$$

for each severity condition $i \in [n]$. As the objective function is linear in the matrix $\boldsymbol{T}$, this optimization program can be solved efficiently when the uncertainty set $\mathcal{U}$ is defined by linear, or convex quadratic, or conic inequalities. Note that the wider and the more unconstrained the uncertainty set, the more difficult it is to satisfy Assumption 3.

For factor matrix uncertainty sets, the authors in Goh et al. (2018) show that one can compute the worst-case reward of a given policy. Goyal and Grand-Clément (2018) show 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 (4.1), the *robust maximum principle* holds. Since our analysis relies on the maximum principle, we state it formally for completeness.

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

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

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

THEOREM 2 **(Robust Maximum Principle (Goyal and Grand-Clément 2018))**. *Let $\mathcal{U}$ be a factor matrix uncertainty set as in* (4.1).

1. *Let $\hat{\boldsymbol{T}} \in \mathcal{U}$ and $\hat{\pi} \in \arg\max_{\pi \in \Pi} R(\pi, \hat{\boldsymbol{T}})$. Then*

$$v_i^{\pi, \hat{\boldsymbol{T}}} \leq v_i^{\hat{\pi}, \hat{\boldsymbol{T}}}, \forall\, \pi \in \Pi, \forall i \in [n]. \tag{4.2}$$

2. *Let $\hat{\pi}$ be a policy and $\hat{\boldsymbol{T}} \in \arg\min_{\boldsymbol{T} \in \mathcal{U}} R(\hat{\pi}, \boldsymbol{T})$. Then*

$$v_i^{\hat{\pi}, \hat{\boldsymbol{T}}} \leq v_i^{\hat{\pi}, \boldsymbol{T}}, \forall\, \boldsymbol{T} \in \mathcal{U}, \forall\, i \in [n]. \tag{4.3}$$

3. *Let $(\pi^\star, \boldsymbol{T}^\star) \in \arg\max_{\pi \in \Pi} \min_{\boldsymbol{T} \in \mathcal{U}} R(\pi, \boldsymbol{T})$. For all policy $\hat{\pi}$, for all transition matrix $\hat{\boldsymbol{T}} \in \arg\min_{\boldsymbol{T} \in \mathcal{U}} R(\hat{\pi}, \boldsymbol{T})$, we have*

$$v_i^{\hat{\pi}, \hat{\boldsymbol{T}}} \leq v_i^{\pi^\star, \boldsymbol{T}^\star}, \forall i \in [n]. \tag{4.4}$$

Inequality (4.2) implies that in a classical MDP setting, the value function of the optimal nominal policy is component-wise higher than the value function 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 (4.3), when a policy is fixed but the transition matrix varies in the uncertainty set $\mathcal{U}$, the worst-case value function of the policy is component-wise lower than the value function of the policy for any other transition matrix. Finally, when we consider an optimal robust policy, Inequality (4.4) implies that the worst-case value function of the optimal robust policy is component-wise higher than the worst-case value function of any other policy. Importantly, this shows that the optimal robust policy is maximizing the worst-case expected reward *starting from any state.*

4.2. Theoretical guarantees.

We now prove our structural properties for the optimal robust policy. In particular, we start with the following theorem.

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

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

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

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

$$(\pi^{\mathsf{rob}}, \boldsymbol{T}^{\mathsf{rob}}) \in \arg\max_{\pi \in \Pi} \min_{\boldsymbol{T} \in \mathcal{U}} R(\pi, \boldsymbol{T}) \iff \pi^{\mathsf{rob}} \in \arg\max_{\pi \in \Pi} R(\pi, \boldsymbol{T}^{\mathsf{rob}}). \tag{4.5}$$

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

Remark 5. It is possible to extend Proposition 1, which holds for the nominal model, to the robust single-patient MDP. In particular, if for a given $\epsilon > 0$ the conditions (3.4)-(3.5) are valid inequalities for any matrix $\boldsymbol{T} \in \mathcal{U}$, we can find a threshold policy that is suboptimal by at most $2\epsilon(n-1)\lambda/(1-\lambda)$ in the worst-case. This is because the proof of Proposition 1 works in the nominal case, and in a factor model, an optimal robust policy π^{rob} is also optimal (in the nominal sense) for a given transition matrix $\boldsymbol{T}^{\mathsf{rob}} \in \mathcal{U}$ (see (4.5) in the proof of Theorem 3). Therefore, even in the worst-case, it is possible to precisely quantify the suboptimality of threshold policies in the case that Assumption 3 is violated.

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* $\tau(\pi^{\mathsf{rob}}) \leq \tau(\pi^{\mathsf{nom}})$, *where* π^{rob} *is the optimal robust policy and* π^{nom} *is the optimal nominal policy.*

Proof. Let $\hat{\Pi}$ be the set of policies that are optimal for some transition kernel in $\mathcal{U}$: $\hat{\Pi} = \{\pi \mid \exists\, \boldsymbol{T} \in \mathcal{U}, \pi \in \arg\max_{\pi \in \Pi} R(\pi, \boldsymbol{T})\}$. Note that $\pi^{\mathsf{nom}} \in \hat{\Pi}$. We will prove that $\tau(\pi^{\mathsf{rob}}) \leq \tau(\pi), \forall\, \pi \in \hat{\Pi}$.

Following Theorem 3, we can pick π^{rob} to be an optimal robust policy that is a threshold policy. We denote $\boldsymbol{T}^{\mathsf{rob}}$ a matrix in $\mathcal{U}$ such that $(\pi^{\mathsf{rob}}, \boldsymbol{T}^{\mathsf{rob}}) \in \max_{\pi \in \Pi} \min_{\boldsymbol{T} \in \mathcal{U}} R(\pi, \boldsymbol{T})$. Let $\hat{\pi} \in \hat{\Pi}$. There

exists a transition matrix $\hat{\boldsymbol{T}} \in \mathcal{U}$ such that $\hat{\pi} \in \arg\max_{\pi \in \Pi} R(\pi, \hat{\boldsymbol{T}})$. Let us assume that $\hat{\pi}(i) = 1$ for some $i \in [n]$. We will prove that $\pi^{\mathsf{rob}}(i) = 1$. We have

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

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

where Inequality (4.6) follows from the Bellman Equation for the MDP with transition matrix $\hat{\boldsymbol{T}}$ and Inequality (4.7) follows from Inequality (4.2) of Theorem 2:

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

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

$$r_W + \lambda \cdot \hat{\boldsymbol{T}}_{i,\cdot}^{\top} \boldsymbol{V}^{\pi^{\mathsf{rob}}, \hat{\boldsymbol{T}}} = V_i^{\pi^{\mathsf{rob}}, \hat{\boldsymbol{T}}}. \tag{4.8}$$

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

$$r_W + \lambda \cdot r_{PT} > V_i^{\pi^{\mathsf{rob}}, \hat{\boldsymbol{T}}} \tag{4.9}$$

$$\geq V_i^{\pi^{\mathsf{rob}}, \boldsymbol{T}^{\mathsf{rob}}}, \tag{4.10}$$

where the strict Inequality (4.9) follows from (4.8) and (4.6), and Inequality (4.10) follows from (4.3) in the robust maximum principle:

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

We can therefore conclude that

$$r_W + \lambda \cdot r_{PT} > V_i^{\pi^{\mathsf{rob}}, \boldsymbol{T}^{\mathsf{rob}}}. \tag{4.11}$$

But since π^{rob} is an optimal robust policy, we know following Theorem 2 that $\pi^{\mathsf{rob}} \in \arg\max_{\pi \in \Pi} R(\pi, \boldsymbol{T}^{\mathsf{rob}})$. Therefore, from the Bellman Equation we know that $\pi^{\mathsf{rob}}(i) = 1$ if $r_W + \lambda \cdot r_{PT} > r_W + \lambda \cdot \boldsymbol{T}_{i,\cdot}^{\mathsf{rob}\,\top} \boldsymbol{V}^{\pi^{\mathsf{rob}}, \boldsymbol{T}^{\mathsf{rob}}}$ and $\pi^{\mathsf{rob}}(i) = 0$ if $r_W + \lambda \cdot r_{PT} \leq r_W + \lambda \cdot \boldsymbol{T}_{i,\cdot}^{\mathsf{rob}\,\top} \boldsymbol{V}^{\pi^{\mathsf{rob}}, \boldsymbol{T}^{\mathsf{rob}}}$. This implies that $V_i^{\pi^{\mathsf{rob}}, \boldsymbol{T}^{\mathsf{rob}}} \geq r_W + \lambda \cdot r_{PT}$, which contradicts Inequality (4.11), and therefore it is impossible that $\pi^{\mathsf{rob}}(i) = 0$. Since π^{rob} is a deterministic policy, $\pi^{\mathsf{rob}}(i) \neq 0 \Rightarrow \pi^{\mathsf{rob}}(i) = 1$. We have proved that

if $\hat{\pi}(i) = 1$ for some $\hat{\pi}$ in $\hat{\Pi}$ and some $i \in [n]$, then $\pi^{\mathsf{rob}}(i) = 1$. Therefore, we can conclude that $\tau(\pi^{\mathsf{rob}}) \leq \tau(\pi), \forall\, \pi \in \hat{\Pi}$. Since π^{nom} is the optimal policy for the nominal transition kernel $\boldsymbol{T}^0$, we can conclude that $\pi^{\mathsf{nom}} \in \hat{\Pi}$ and therefore in particular $\tau(\pi^{rob}) \leq \tau(\pi^{\mathsf{nom}})$. □

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.

4.3. Whittle indexability of the robust single-patient MDP.

Policies based on *indexes* (e.g., Gittins index (Gittins 1979) and Whittle index (Whittle 1988)) provide simple and interpretable characterizations of optimal policies. In this section, we use the Whittle index to shed new lights on the structure of optimal nominal and optimal robust policies. In particular, we have the following definition.

Definition 1 (Robust Whittle index). For a given state i, the robust Whittle index $\omega^{\mathsf{rob}}(i)$ at state i is the choice of reward r_{PT} such that it is equally desirable (in the worst-case) to proactively transfer the patient in state i and to not proactively transfer the patient.

Note that the robust Whittle index is dependent on the state i. The optimal robust policy proactively transfers the patient in state i if and only if $r_{PT} \geq \omega^{\mathsf{rob}}(i)$. We can also define the *nominal* Whittle index as follows. From Whittle (1988), the nominal Whittle index $\omega^{\mathsf{nom}}(i)$ of the single-patient MDP at a state i is the value of r_{PT} such that it is equally desirable (for the nominal decision-maker) to proactively transfer the patient in state i and to not proactively transfer the patient. Therefore, our robust Whittle index relates to the classical definition of the Whittle index but for the optimal *robust* policy (instead of the optimal *nominal* policy). To the best of our knowledge, our paper is the first to introduce a robust version of the Whittle index, the robust versions of the Gittins index being studied in Caro and Gupta (2015) and Kim and Lim (2016).

We now study the *indexability* of the robust single-patient MDP.

DEFINITION 2 (INDEXABILITY). Let $\mathcal{I}(r_{PT})$ be the set of severity conditions for which the optimal robust policy (for reward r_{PT}) proactively transfers the patient to the ICU. The robust single-patient MDP is *indexable* if $\mathcal{I}(r_{PT})$ is monotonically increasing from the empty set $\emptyset$ to the set of all severity conditions $[n]$, when the reward r_{PT} is increasing from $r_{PT} = 0$ to $r_{PT} = r_{RL}$.

The indexability of the robust single-patient MDP is equivalent to the robust Whittle indices $i \mapsto \omega^{\mathsf{rob}}(i)$ being non-increasing, i.e., to $\omega^{\mathsf{rob}}(n) \leq \omega^{\mathsf{rob}}(n-1) \leq ... \leq \omega^{\mathsf{rob}}(1)$. In this case, for a given value of $r_{PT} \in [0, r_{RL}]$, there exists an integer $\tau \in \{0, ..., n+1\}$ such that $\omega^{\mathsf{rob}}(\tau - 1) \leq \tau \leq \omega^{\mathsf{rob}}(\tau)$ (with the convention that $\omega^{\mathsf{rob}}(n+1) = 0, \omega^{\mathsf{rob}}(0) = r_{RL}$), and the optimal robust policy transfers all patients with a risk score higher (or equal) than τ, i.e., the optimal robust policy has threshold τ. Note that both the cases $r_{PT} = 0$ and $r_{PT} = r_{RL}$ are easy to interpret: when there is no reward for proactive transfer, we do not have any incentive to proactively transfer any patients, whereas when there is as much reward for proactive transfer as for recovering and leaving the ward (state RL), we proactively transfer every severity condition. We prove in the following proposition that the robust single-patient MDP is indexable. We present a detailed proof in Appendix G.

PROPOSITION 2. *Consider that Assumption 1 holds, and that Assumption 3 holds for $r_{PT} = 0$. Then the robust single-patient MDP is indexable.*

Compared to the results of Section 3.2 and Section 4.2, note the relatively stronger assumption for Proposition 2, which requires that Assumption 3 holds for $r_{PT} = 0$. This is a sufficient condition to ensure that an optimal robust policy can be chosen threshold for all $r_{PT} \in [0, r_{RL}]$. When Assumption 3 fails to hold for $r_{PT} = 0$, the robust single-patient MDP may not be indexable, since for some values of $r_{PT} \in (0, r_{RL})$, there may not exist threshold policies that are optimal, as highlighted in our counter-example in Appendix E. Still, Proposition 2 provides practical insights on the sensitivity of the threshold of optimal robust policies as regards r_{PT}, as highlighted in the following immediate corollary.

COROLLARY 1. *Consider that Assumption 1 holds and that Assumption 3 holds for $r_{PT}=0$. Then the threshold of an optimal robust policy in the single-patient MDP is decreasing with r_{PT}.*

The proofs of Proposition 2 and Corollary 1 rely on the following facts:

- At every state i, the robust decision-maker compares the value for proactively transferring the patient: r_{PT}, and the value obtained in the next states if the patients is not proactively transferred: $\boldsymbol{T}_i^\top \boldsymbol{V}$, where $\boldsymbol{V}$ is the value function of the decision-maker and $\boldsymbol{T}$ is the transition matrix.
- When the reward r_{PT} increases to $r_{PT}+\epsilon$ (for $\epsilon>0$), the value for proactively transferring the patient increases to $r_{PT}+\epsilon$.
- When the reward r_{PT} increases to $r_{PT}+\epsilon$, the value function $\boldsymbol{V}$ increases by less (or equal) than ϵ on each component j related to a severity condition $j\in[n]$, because the transition to PT happens after at least one period, in which case there is a discount factor of (at least) $\lambda<1$ on the reward for PT.
- Therefore, if for a value r_{PT} it is optimal for the robust decision-maker to proactively transfer a patient in state i, it will remain optimal to do so for any $r'_{PT}\geq r_{PT}$.

REMARK 6 (NOMINAL WHITTLE INDEX). We would like to highlight that all the results in this section (Proposition 2 and Corollary 1 for a robust single-patient MDP with uncertainty set $\mathcal{U}$) can be applied to the *nominal* single-patient MDP with nominal transition matrix $\boldsymbol{T}^0$, simply by choosing an uncertainty set $\mathcal{U}=\{\boldsymbol{T}^0\}$. In particular, the nominal single-patient MDP is also indexable (for the nominal Whittle index), under the assumption that Assumption 2 holds for $r_{PT}=0$. Since we know from Theorem 4 that the threshold of an optimal robust policy is always smaller than the threshold of an optimal nominal policy, this also shows that the robust Whittle index at a state i is always smaller than the nominal Whittle index at the same state.

REMARK 7 (NUMERICAL METHODS FOR COMPUTING WHITTLE INDICES.). Note that computing a closed-form solution for the robust Whittle index at state i in our setting is challenging. Closed-form expressions of the Whittle index appear in problems with simpler dynamics where the agent in state $i\in[n]$ can transition only to $i+1$ or $i-1$ (Whittle 1988, Hsu 2018, Tripathi and Modiano

2019). In contrast to these models, in our single-patient MDP model, a patient in severity condition $i \in [n]$ can transition to any other severity condition $j \in [n]$, and terminal states RL, CR and D. Despite not being tractable to derive closed-form expressions, it is still possible to compute the nominal and the robust Whittle indices using numerical methods based on binary searches to search for the thresholds of the optimal nominal and robust policies for each value of r_{PT}. These searches can be simplified using our theoretical results: for any value of r_{PT}, we can search for an optimal nominal policy by finding the threshold policy with the highest nominal reward (Theorem 1). We can then search for an optimal robust policy among threshold policies (Theorem 3) with a lower threshold than the optimal nominal policy (Theorem 4).

5. A robust multi-patient MDP model.

In this section, we provide a theoretical justification of the relation between ICU capacity and penalizing rewards in our single-patient MDP. In particular, we introduce a multi-patient MDP, which we call *N-patient MDP* below. The complexity of this model is intermediary between the (simpler) single-patient MDP of Section 3 and the (more complex) full hospital model of Section 2. It consists of $N \in \mathbb{N}$ patients, whose individual dynamics of the severity condition follows the single-patient MDP described in Figure 2.

Model of patient dynamics. The states are the tuples representing the severity conditions of each of the N patients. The actions are whether or not to transfer patients to the ICU, and if so, which ones. Crucially, there is a constraint on the number of patients that can be transferred at the same time to the ICU.

Set of states. A state in the N-patient MDP is an N-tuple $(i_1, ..., i_N)$, where for each $\ell \in \{1, ..., N\}$, i_ℓ describes the severity condition of patient ℓ (in which case $i_\ell \in [n]$), or one of the four absorbing states $r_{RL}, r_{CR}, r_D, r_{PT}$. Note that the number of states, $(n+4)^N$, is exponential in the number of patients.

Set of actions and policy. An action is an N-tuple $\left(\pi^1_{i_1}, ..., \pi^N_{i_N}\right)$ where $\pi^\ell_{i_\ell}$ represents the probability to proactively transfer a patient with severity score i_ℓ. The number of (deterministic) actions

is also exponential in N. A policy is an N-tuple $(\pi^1, ..., \pi^N)$ where each π^ℓ is a map from the set of severity conditions $[n]$ to $[0, 1]$, i.e. *each π^ℓ is a policy in the single-patient MDP for patient $\ell \in [N]$.*

Constraints on transfers to the ICU. To incorporate the effect of proactive transfers on other patients in the hospital, we choose to add a constraint in our N-patient MDP. In particular, the chosen action $\left(\pi^1_{i_1}, ..., \pi^N_{i_N}\right)$ for state $(i_1, ..., i_N)$ has to satisfy

$$\sum_{\ell=1}^{N} \pi^\ell_{i_\ell} \leq m, \tag{5.1}$$

where $m \in \mathbb{N}$ is the maximum number of patients that can be proactively transferred at the same time. Note that in practice, the person (or group of persons) who makes the decision to transfer patients from the ward to the ICU may not have access to an accurate estimate of the ICU capacity. Our analysis provides insights into the case that an accurate estimate for the capacity constraint is available.

Rewards. The rewards depend on the state $(i_1, ..., i_N)$ and are defined as the sum of the individual rewards across all N patients:

$$r_{(i_1,...,i_N)} = \sum_{\ell=1}^{N} r_{i_\ell},$$

where r_{i_ℓ} is the reward as in the single-patient MDP for a patient in severity condition i_ℓ, described at the beginning of Section 3.

Transitions. Once action $\left(\pi^1_{i_1}, ..., \pi^N_{i_N}\right)$ is chosen in state $(i_1, ..., i_N)$, the patients who are proactively transferred transition to state PT, and other patients all transition to a next severity condition (or RL, CR, D) following the same transition matrix $\boldsymbol{T}^0$ as for the single-patient MDP.

Uncertainty set. We consider a *robust* version of the N-patient MDP, where the transition matrix $\boldsymbol{T}$ describing the health evolution of each patient is uncertain and belongs to a factor matrix uncertainty set $\mathcal{U}$ of the form (4.1).

Relation with single-patient dynamics. We now show that our robust single-patient MDP is a Lagrangian relaxation of this robust N-patient MDP, in the sense of Adelman and Mersereau (2008). In particular, the numbers of states and actions of the N-patient MDP are exponential

in N, rendering it intractable. While the robust Bellman equation for the value function $\boldsymbol{J}$ of the N-patient MDP is

$$J_{(i_1,\ldots,i_N)} = \max_{\left(\pi^1_{i_1},\ldots,\pi^N_{i_N}\right);\sum_{\ell=1}^N \pi^\ell_{i_\ell}\leq m} \min_{\boldsymbol{T}\in\mathcal{U}} \sum_{\ell=1}^N r_{i_\ell} + \lambda \sum_{\left(i'_1,\ldots,i'_N\right)\in\mathbb{S}} \left(\Pi_{\ell=1}^N T_{i_\ell,i'_\ell}\right) J_{\left(i'_1,\ldots,i'_N\right)}, \forall\, i_1,\ldots,i_N \in [n], \tag{5.2}$$

the authors in Adelman and Mersereau (2008) suggest to approximate this equation by another equation on a vector $\boldsymbol{J}^\mu$, for a Lagrange multiplier $\mu \geq 0$:

$$\begin{aligned} J^\mu_{(i_1,\ldots,i_N)} = \max_{\left(\pi^1_{i_1},\ldots,\pi^N_{i_N}\right)} \min_{\boldsymbol{T}\in\mathcal{U}} \sum_{\ell=1}^N r_{i_\ell} + \lambda \sum_{\left(i'_1,\ldots,i'_N\right)\in\mathbb{S}} \left(\Pi_{\ell=1}^N T_{i_\ell,i'_\ell}\right) J^\mu_{\left(i'_1,\ldots,i'_N\right)} \\ + \mu\left(m - \sum_{\ell=1}^N \pi^\ell_{i_\ell}\right), \forall\, i_1,\ldots,i_N \in [n]. \end{aligned} \tag{5.3}$$

In this case, we obtain that

$$J^\mu_{(i_1,\ldots,i_N)} = \frac{m\cdot\mu}{1-\lambda} + \sum_{\ell=1}^N \boldsymbol{V}^\mu_\ell(i_\ell), \forall\, i_1,\ldots,i_N \in [n], \tag{5.4}$$

where $\boldsymbol{V}^\mu_\ell \in \mathbb{R}^n$ are the robust value functions for the (penalized) robust single-patient MDPs:

$$V^\mu_\ell(i_\ell) = \max_{\pi^\ell_{i_\ell}\in[0,1]} \min_{\boldsymbol{T}\in\mathcal{U}} r_{i_\ell} - \mu\pi^\ell_{i_\ell} + \lambda\left((1-\pi^\ell_{i_\ell})\boldsymbol{T}_i^\top \boldsymbol{V}^\mu_\ell + \pi^\ell_{i_\ell} r_{PT}\right), \forall\, i_\ell \in [n]. \tag{5.5}$$

Note that we can reformulate (5.5) as

$$V^\mu_\ell(i_\ell) = \max\{r_{i_\ell} + \lambda \min_{\boldsymbol{T}\in\mathcal{U}} \boldsymbol{T}_i^\top \boldsymbol{V}^\mu_\ell, r_{i_\ell} + \lambda\left(r_{PT} - \frac{\mu}{\lambda}\right)\}, \forall\, i_\ell \in [n]. \tag{5.6}$$

Equation (5.6) is the robust Bellman equation for our single-patient MDP, except for the term $-\frac{\mu}{\lambda}$. Note that in (5.6), the first term in the maximization program corresponds to the worst-case reward for *not* proactively transferring the patient from state i_ℓ, while the second term corresponds to the reward associated with proactively transferring the patients from state i_ℓ. This shows the relations between our robust single-patient MDP and the robust N-patient MDP: dualizing the binding constraint on the number of transfers in the robust N-patient MDP yields a robust single-patient MDP, with a penalty on r_{PT}, the reward for proactively transferring the patient.

Remark 8 (Non-homogeneous transition matrices). In the N-patient MDP, we have chosen to model the dynamics of the severity conditions of each of the N patients with the same transition matrix $\boldsymbol{T} \in \mathcal{U}$. We note that the Lagrangian relaxation of Equations (5.3)-(5.6) still holds if we consider different transitions matrices $\boldsymbol{T}_1, ..., \boldsymbol{T}_N$ in different factor matrix uncertainty sets $\mathcal{U}_1, ..., \mathcal{U}_N$ for the patients in $\{1, ..., N\}$. We present our results with $\boldsymbol{T}_1 = ... = \boldsymbol{T}_N = \boldsymbol{T} \in \mathcal{U}$ for the sake of conciseness. We finally note that in the case of non-homogeneous transition matrices, the (robust) Whittle indexes from Section 4.3 can be used to prioritize ICU transfers among patients with the same risk scores but different transition matrices.

Sensitivity to the maximum number of patients transferred. Our robust N-patient MDP relies on an exogenous, static parameter $m \in \mathbb{N}$, the maximum number of patients that can be proactively transferred at any time period. We have shown how the N-patient MDP can be approximated with the single-patient MDP, via Lagrangian relaxation. Here, we leverage our results of Section 4.3 to investigate the impact of the parameter m on the robust policies of the single-patient MDP. In particular, we show the following result.

Proposition 3. *The threshold of an optimal robust policy in the Lagrangian relaxation of the N-patient MDP is decreasing with m.*

Proof. Remember that $\boldsymbol{p}_0$ is the initial distribution over the set of states $\{1, ..., n\}$. From Equation (5.4), the goal of the decision-maker in the relaxation of the N-patient MDP is to solve

$$\min_{\mu \geq 0} \ \frac{m \cdot \mu}{1 - \lambda} + \sum_{\ell=1}^{N} \boldsymbol{p}_0^\top \boldsymbol{V}_\ell^\mu.$$

By a standard duality argument, this is a convex minimization problem. Now, let us consider $m \in \mathbb{N}$ and $m' \in \mathbb{N}$ such that $m \leq m'$. Let $f : \mu \mapsto \sum_{\ell=1}^{N} \boldsymbol{p}_0^\top \boldsymbol{V}_\ell^\mu$. We write $g_m : \mu \mapsto \frac{m \cdot \mu}{1 - \lambda} + f(\mu)$ and $g_{m'} : \mu \mapsto \frac{m' \cdot \mu}{1 - \lambda} + f(\mu)$. Note that for any $\mu \geq 0$, we have $g'_{m'}(\mu) = m' + f'(\mu) \geq m + f'(\mu) = g'_m(\mu)$. Now, let $\mu^\star$ and $\mu^{'\star}$ be the optimal Lagrange multipliers for m and m', characterized by $g'_m(\mu^\star) = 0$ and $g'_{m'}(\mu^{'\star}) = 0$. We have $0 = g'_{m'}(\mu^{'\star}) \geq g'_m(\mu^{'\star})$. Since g is convex, this means that $\mu^\star \geq \mu^{'\star}$. Therefore, we have proved that the optimal Lagrange multiplier $\mu^\star$ is a *non-increasing* function

of m. This shows that larger values of m results in higher reward for proactive transfer in the robust single-patient MDP, since a Lagrange multiplier of μ in the N-patient MDP results in a penalty of $-\mu/\lambda$ for r_{PT} in the robust single-patient MDP. Now, we have proved in Corollary 1 that the threshold of an optimal robust policy is non-increasing as r_{PT} is increasing. Therefore, we can conclude that the threshold of an optimal robust policy in the Lagrangian relaxation of the N-patient MDP is non-increasing as m increases. □

In other words, as the maximum number of patients that can be proactively transferred at the same time in the N-patient MDP increases, an optimal robust policy in the Lagrangian relaxation of the N-patient MDP transfers more patients. Thus, one can expect that it is reasonable to be more aggressive with proactive transfers as the number of available ICU beds increases.

6. 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 sets of our hospital model (Figure 1) and our single-patient MDP (Figure 2). Using simulations, we then compare the performance using simulation of the optimal nominal and optimal robust policies on several metrics of interest: mortality, Length-Of-Stay (LOS) and average ICU occupancy. Note that because the state and action sets of the N-patient MDP of Section 5 are too large (exponential in the number of patients N), we focus our simulations on the Lagrangian relaxation of the N-patient MDP, which reduces to the single-patient MDP, as highlighted in the previous section. In the hospital model, the capacity constraint in the ICU is taken into account in that proactive transfers to the ICU can only happen when there are enough beds available in the ICU.

6.1. Data set.

Our retrospective dataset consists of 296,381 unique patient hospitalizations across 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 $\geq 8\%$. These scores have demonstrated their ability to accurately predict deterioration (Kipnis et al. 2016), and we use them use a proxy for the severity condition of the patients. 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). We remove 11,463 hospitalizations with missing gender or inpatient unit code, time inconsistencies (e.g., arrival after discharge, missing discharge time). We also drop patients involved in hospital transfers (5,781 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. Summary statistics (partition, mortality rate, average length-of-stay, etc.) for our cohort of patients are given in Appendix A.

6.2. Transition matrix and model of uncertainty.

We first calibrate the transition matrix $\boldsymbol{T}^0$ across the severity scores.

6.2.1. Nominal transition matrix. We use the AAM scores as our severity scores. The matrix $\boldsymbol{T}^0$ has dimension $n \times (n+3)$ where $n = 10$ is the number of severity scores. $\boldsymbol{T}^0$ is constructed as follows. Let $i \in [n]$ and $j \in [n] \bigcup \{CR, RL, D\}$. The coefficient T^0_{ij} 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^0_{ij}. To obtain the 95%-confidence intervals for the matrix $\boldsymbol{T}^0$, we use the method in Sison and Glaz (1995), as it computes sharper upper and lower bounds for the confidence intervals (May and Johnson 2000). We obtain :

$$[T^0_{ij} - \alpha_i, T^0_{ij} + 2 \cdot \alpha_i], \forall\, (i,j) \in [n] \times [n+3]. \tag{6.1}$$

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^0_{ij} is uniform across all $j \in [n+3]$. We refer to Appendix H for the values for $\alpha_1, ..., \alpha_n$, which are in the order of 10^{-4} to 10^{-3}.

6.2.2. Nominal factor matrix. To construct a factor matrix uncertainty set (4.1), we need to compute the coefficients $(u^\ell_i)_{(i,\ell)} \in \mathbb{R}^{n \times r}$, the nominal factors $\boldsymbol{W} = (\boldsymbol{w}_1, ..., \boldsymbol{w}_r) \in \mathbb{R}_+^{(n+3)\times r}$ such that $\boldsymbol{T}^0 \approx \boldsymbol{U}\boldsymbol{W}^\top$, and the confidence regions $\mathcal{W}^i$ for each factor $\boldsymbol{w}_i, i = 1, ..., r$. To do this, we solve the following Nonnegative Matrix Factorization (NMF) problem:

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

This is a non-convex optimization problem, but 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 we keep the best solution[2]. For $r = 8$, our solution $\hat{\boldsymbol{T}} = \boldsymbol{U}\hat{\boldsymbol{W}}^\top$ achieves the following errors: $\|\boldsymbol{T}^0 - \hat{\boldsymbol{T}}\|_1 = 0.0811, \|\boldsymbol{T}^0 - \hat{\boldsymbol{T}}\|_\infty = 0.0074, \|\boldsymbol{T}^0 - \hat{\boldsymbol{T}}\|_{\mathsf{relat},\mathsf{T}^0} = 0.3385$, where $\|\cdot\|_{\mathsf{relat},\mathsf{T}^0}$ stands for the maximum relative deviation from a parameter of $\boldsymbol{T}^0$:

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

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

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

	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^0_{ij}, for all $(i,j) \in [n] \times [n+3]$.

Recall that we utilize a factor matrix uncertainty set to model rank-constrained deviations from the nominal parameters, in contrast to full-rank deviations in rectangular uncertainty sets. Therefore, we expect the rank to be smaller than the number of states, $r < n$. We choose $r = 8$. This is the smallest integer for which we were able to find a nonnegative matrix factorization (of this rank) belonging to the confidence intervals. We give details about our simulations for $r = 7$ in Appendix J, for which we obtain similar insights as for $r = 8$. For $r = 8$, the maximum *absolute* deviation between T^0_{ij} 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 $\boldsymbol{T}^0$. This occurs for $T^0_{3,9}$, which represents a sudden, dramatic, and relatively rare health deterioration from state 3 to state 9.

Errors related to the confidence intervals. As our NMF approximation $\hat{\boldsymbol{T}}$ is supposed to be a low-rank approximation of $\boldsymbol{T}^0$, it is of interest to understand if it is consistent with the confidence intervals of our parameter estimates. We consider the relative errors of $\hat{\boldsymbol{T}}$ compared to $\boldsymbol{T}^0$, measured in terms of the confidence bounds $\alpha_1, ..., \alpha_n$. In particular, we compute the ratios

$$ratio_{(i,j)} = \frac{T^0_{ij} - \hat{T}_{ij}}{\alpha_i}, \forall\, (i,j) \in [n] \times [n+3]. \tag{6.2}$$

For $r = 8$, we find that all coefficients are in the confidence intervals as defined by (6.1), i.e., $ratio_{(i,j)} \in [-1, 2], \forall\, (i,j) \in [n] \times [n+3]$. The mean over (i,j) of the absolute values of the ratios (6.2) is 0.2345, with a median of 0.1579. Moreover, 95% of these absolute values are below 0.6729. Therefore, $\hat{\boldsymbol{T}}$ (our NMF solution of rank 8) is a plausible approximation for $\boldsymbol{T}^0$.

For completeness, we also compute the solutions to the NMF optimization problem for lower ranked matrices: $r = 5, 6, 7$. While the errors in the L_1- and $L_\infty-$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 $T^0_{9,9}$. 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 $r = 7$ also appears to be a reasonable approximation for $\boldsymbol{T}^0$. 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 $r = 8$. We also present detailed experiments for $r = 7$ in Appendix J.

6.2.3. Choice of uncertainty sets. We consider several uncertainty sets for our analysis. We consider the factor matrix uncertainty sets $\mathcal{U}_{\min}$ and $\mathcal{U}_{\mathsf{emp}}$ defined below, with rank-constrained parameter deviations from the nominal parameters. For the sake of comparison, we also consider the rectangular uncertainty set $\mathcal{U}_{\mathsf{sa}}$, with full-rank parameter deviations. We choose to relate our uncertainty sets with the confidence intervals estimated from the data, since we can interpret the uncertainty sets as the set of all plausible parameter realizations. This will also help in our comparisons with random parameter deviations in the confidence intervals. In particular, we consider the following uncertainty sets:

- $\mathcal{U}_{\min}$: We consider 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

$$\mathcal{U}_{\min} = \left\{ \boldsymbol{T} \;\middle|\; \boldsymbol{T} = \boldsymbol{U}\boldsymbol{W}^\top, \boldsymbol{W} \in \mathcal{W}_{\min} \right\},$$

where $\mathcal{W}_{\min} = \mathcal{W}^1_{\min} \times ... \times \mathcal{W}^r_{\min}$,

$$\mathcal{W}^\ell_{\min} = \left\{ \boldsymbol{w}_\ell \;\middle|\; \forall\, j \in [n+3], w_{\ell,j} - \hat{w}_{\ell,j} \in [-\alpha_{\min}, +2 \cdot \alpha_{\min}], \boldsymbol{w}_\ell \geq \boldsymbol{0}, \boldsymbol{w}_\ell^\top \boldsymbol{e}_{n+3} = 1 \right\}, \forall\, \ell \in [r].$$

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

while it is in principle possible to construct $\mathcal{U}_{\max}$, where the maximum deviation on each component component of the factor vectors must be within $[-\alpha_{\max}, 2 \cdot \alpha_{\max}]$, this would result in worst-case matrices where almost all coefficients are outside of the confidence intervals, contrary to $\mathcal{U}_{\min}$; see details in the next section.

- $\mathcal{U}_{\mathsf{emp}}$: 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 matrix $\boldsymbol{T}$. First, we sample q transition matrices $\boldsymbol{T}^1, ...\boldsymbol{T}^q$ uniformly in the 95% confidence intervals around $\boldsymbol{T}^0$, for $q = 10^4$. For each sampled matrix, we use Nonnegative Matrix Factorization to compute factor vectors $\boldsymbol{W}^1, ..., \boldsymbol{W}^q$ such that $\boldsymbol{T}^m \approx \boldsymbol{U}\boldsymbol{W}^{m\top}, m = 1, ..., q$. We 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 $\boldsymbol{W}^1, ..., \boldsymbol{W}^q$. We then define the uncertainty set $\mathcal{U}_{\mathsf{emp}} = \{\boldsymbol{T} \mid \boldsymbol{T} = \boldsymbol{U}\boldsymbol{W}^\top, \boldsymbol{W} \in \mathcal{W}_{\mathsf{emp}}\}$, where $\mathcal{W}_{\mathsf{emp}} = \mathcal{W}_{\mathsf{emp}}^1 \times ... \times \mathcal{W}_{\mathsf{emp}}^r$ represents the bootstrapped 95% confidence intervals for the factor vectors:

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

- $\mathcal{U}_{\mathsf{sa}}$: Finally, we also consider the following (s,a)-rectangular uncertainty set:

$$\mathcal{U}_{\mathsf{sa}} = \left\{ \boldsymbol{T} \;\middle|\; T_{ij} - T_{ij}^0 \in [-\alpha_i, +2 \cdot \alpha_i], \sum_{j=1}^{13} T_{ij} = 1, \forall\, i \in [n] \right\}.$$

In $\mathcal{U}_{\mathsf{sa}}$, the transitions from each state can be chosen unrelated to the transitions out of any other states. In a healthcare setting, there may be common characteristics that drive the dynamics of the health evolution of the patients at every severity conditions, such as demographics or comorbidities; this is overlooked in $\mathcal{U}_{\mathsf{sa}}$, contrary to the factor matrix uncertainty sets $\mathcal{U}_{\min}$ and $\mathcal{U}_{\mathsf{emp}}$, which explicitly model these relations.

6.3. Robustness analysis for the single-patient MDP.

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

6.3.1. 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^0_{ij}. 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 $\boldsymbol{p}_0 \in \mathbb{R}^{n+4}_+$ 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. While our theoretical results are agnostic to the choice of discount factor, we also verify that with alternative discount factors (e.g., $\lambda = 0.99$), we obtain similar insights. We choose the following rewards, satisfying Assumption 1 and Assumption 3:

$$\begin{aligned} r_W = 100,\, r_{RL} = \frac{1}{1-\lambda} \cdot 250,\, r_{PT-RL} = \frac{1}{1-\lambda} \cdot 190,\, r_{CR-RL} = \frac{1}{1-\lambda} \cdot 160, \\ r_D = \frac{1}{1-\lambda} \cdot 30,\, r_{CR-D} = \frac{1}{1-\lambda} \cdot 20,\, r_{PT-D} = \frac{1}{1-\lambda} \cdot 10, \end{aligned} \tag{6.3}$$

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

$$r_{RL} \geq r_{PT-RL} \geq r_{CR-RL}, , r_D \geq r_{CR-D} \geq r_{PT-D}. \tag{6.4}$$

Certainly, different choices of rewards may lead to different thresholds for the optimal nominal and the optimal robust policies. It is a notoriously complex problem to estimate the exact values for such quantities as r_{RL} and r_W, let alone r_{PT-D} and r_{PT-RL} (or more generally rewards in applications of reinforcement learning to healthcare, e.g., Yauney and Shah (2018), or Section II, "Representation for Reward Function" in Yu et al. (2019)). Appendix I 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. While the thresholds of the optimal and nominal robust policies for the single MDP are highly dependent on the rewards, our assumptions enjoy desirable robustness properties to translation and scaling of the rewards, as we prove in Appendix C. Overall, 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 (6.3).

6.3.2. Empirical results for the single-patient MDP. We verify that for our choice of rewards, Assumptions 1-2 are satisfied. Verifying Assumption 3 requires solving some linear programs, as the uncertainty sets $\mathcal{U}_{\min}$, $\mathcal{U}_{\text{emp}}$ and $\mathcal{U}_{\text{sa}}$ are defined by linear inequalities. 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 $\pi^{[1]}, ..., \pi^{[11]}$, and compare their nominal and worst-case rewards in the single-patient MDP for the different uncertainty sets. In Figure 3, we present the nominal and the worst-case rewards of threshold policies for an NMF approximation of rank 8. For any threshold $\tau = 1, ..., 11$, "Nom." stands for $R(\pi^{[\tau]}, \boldsymbol{T}^0)$, while "NMF-8" stands for $R(\pi^{[\tau]}, \hat{\boldsymbol{T}})$. 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 esti-

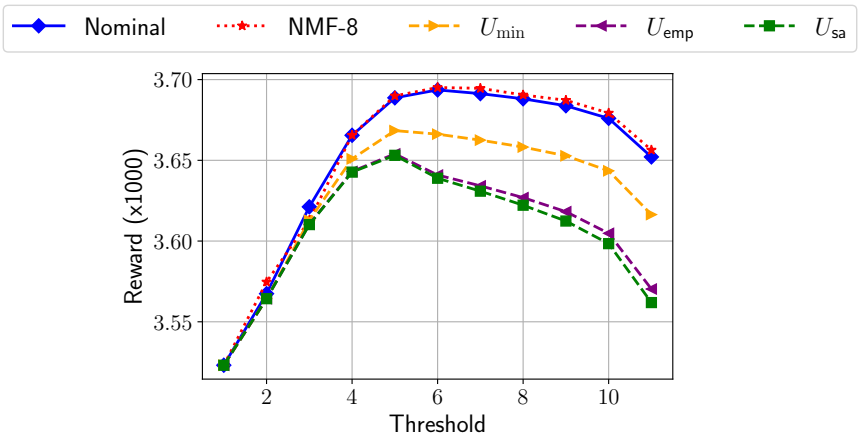


Figure 3 Empirical nominal and worst-case rewards for our single-patient MDP.

mated transition matrix, $R(\pi^{[\tau]}, \boldsymbol{T}^0)$, and that using the NMF approximation of the transition matrix, $R(\pi^{[\tau]}, \hat{\boldsymbol{T}})$, are practically indistinguishable. This provides additional support for using our NMF solution as an approximation for $\boldsymbol{T}^0$. 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 $\mathcal{U}_{\mathsf{emp}}$ do not belong to the 95% confidence intervals (6.1), even though $\hat{\boldsymbol{T}}$ belongs to (6.1). 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 (6.1), 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 $\mathcal{U}_{\mathsf{emp}}$. For instance, for the worst-case matrix in $\mathcal{U}_{\mathsf{emp}}$ 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 $T^0_{1,1}$ (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 $\mathcal{U}_{\mathsf{emp}}$ as plausible transition matrices for our hospital model.

6.4. 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 performance. 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 $\mathcal{U}_{\min}$,$\mathcal{U}_{\mathsf{emp}}$,$\mathcal{U}_{\mathsf{sa}}$), we compute $\boldsymbol{T}^{[\tau,\mathcal{U}]}$ a worst-case transition matrix for the single-patient MDP in $\mathcal{U}$:

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

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

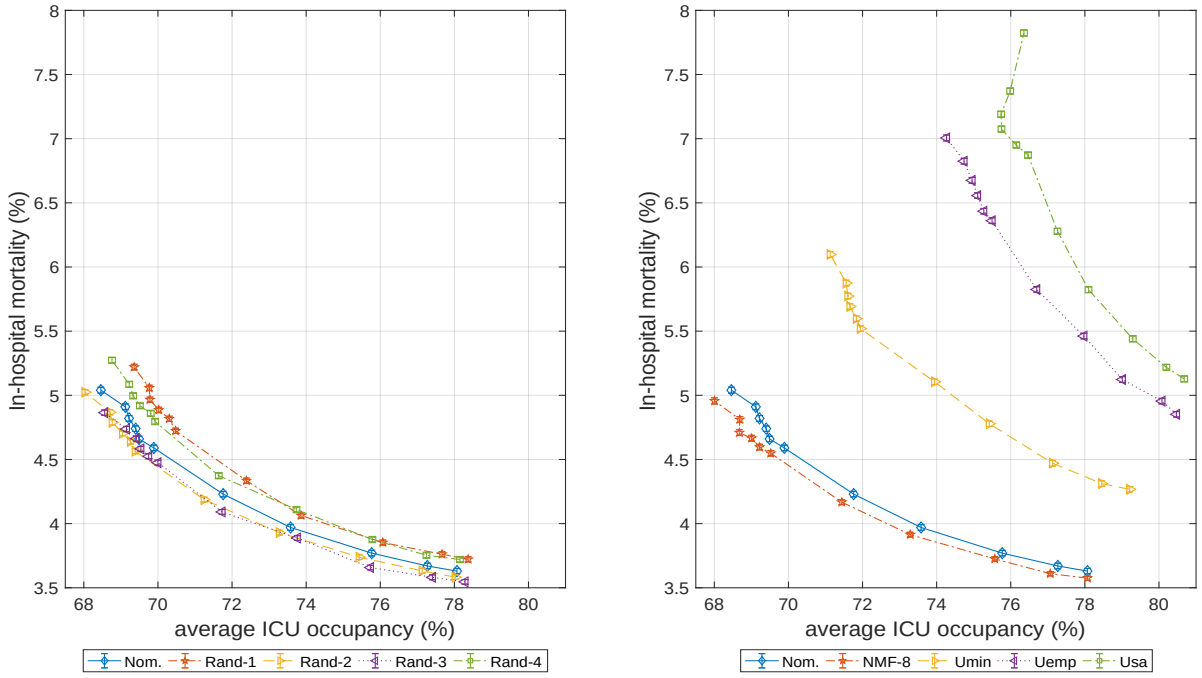
6.4.1. In-hospital mortality and Length-Of-Stay.

In-hospital mortality. In Figure 4a we study the variation of the hospital performance over the 95% confidence intervals for the nominal transition matrix $\boldsymbol{T}^0$. In particular, we sample $N = 20$ transition matrices in the confidence intervals (6.1) and plot the nominal performance (mortality rate versus average ICU occupancy) of all threshold policies as well as the performance 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 K.

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 performance for this particular matrix and threshold policy. As we saw in the single-patient MDP experiments, the hospital performance for $\boldsymbol{T}^0$ and $\hat{\boldsymbol{T}}$ are very close, again suggesting that the NMF approximation is reasonable. As before, we consider the three uncertainty sets: $\mathcal{U}_{\mathsf{min}}$, $\mathcal{U}_{\mathsf{emp}}$ and $\mathcal{U}_{\mathsf{sa}}$. Note that $\mathcal{U}_{\mathsf{min}}$ and $\mathcal{U}_{\mathsf{emp}}$ are centered around our NMF approximation $\hat{\boldsymbol{T}}$. Under uncertainty set $\mathcal{U}_{\mathsf{min}}$, the mortality rate can significantly increase, with relative increases from 18% to 23%. This substantial degradation occurs even though $\mathcal{U}_{\mathsf{min}}$ is our most-optimistic uncertainty set, with variations in the order of 10^{-4} from $\hat{\boldsymbol{T}}$. For worst-case matrices in $\mathcal{U}_{\mathsf{emp}}$ or $\mathcal{U}_{\mathsf{sa}}$, 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 $\boldsymbol{T}^0$. 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 performance.

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%.



(a) Random samples analysis. (b) Worst-case analysis.

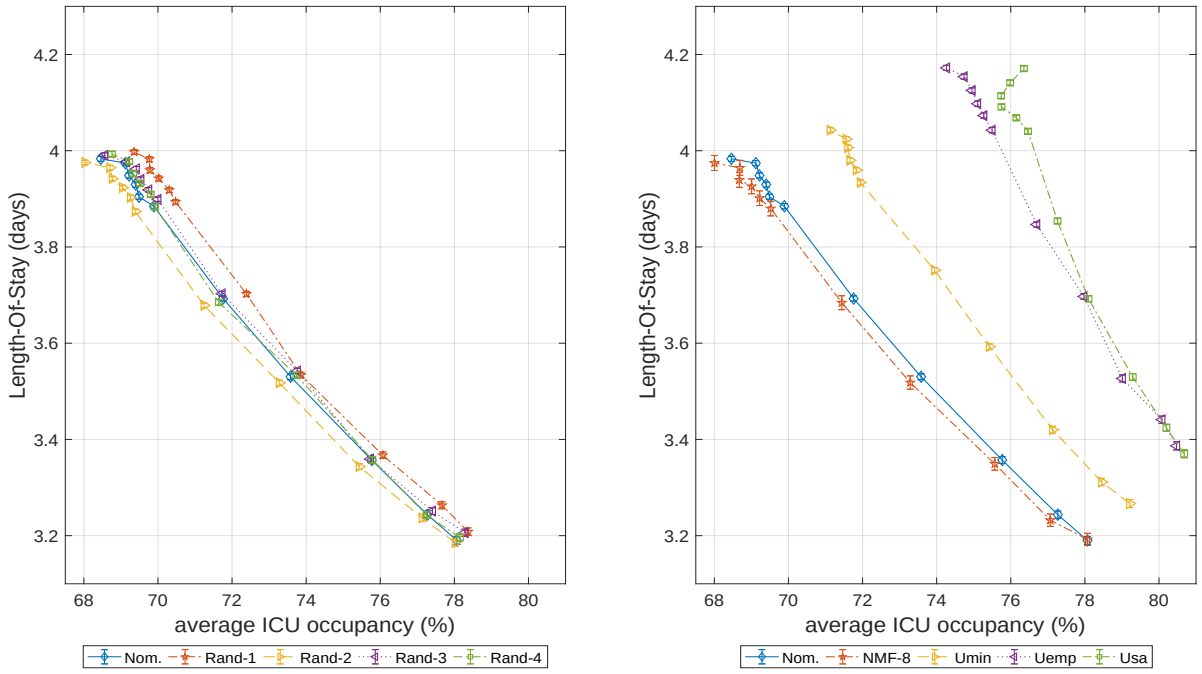
Figure 4 In-hospital mortality of the 11 threshold policies for the nominal estimated matrix, randomly sampled matrices in the 95% confidence intervals (left-hand side), and the worst-case matrices found by our single MDP model (right-hand side). Each point corresponds to a threshold policy: the policy 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 $\mathcal{U}_{\min}$,$\mathcal{U}_{\text{emp}}$ and $\mathcal{U}_{\text{sa}}$ 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 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 $\mathcal{U}_{\mathsf{sa}}$. In particular, we notice in Figure 4b that for $\mathcal{U}_{\mathsf{sa}}$, 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. Therefore, proactively transferring them may actually be Pareto improving. That said, we are somewhat cautious about the interpretations of $\mathcal{U}_{\mathsf{sa}}$. First, we do not observe this effect in the simulation with the nominal parameters $\boldsymbol{T}^0$, nor with $\mathcal{U}_{\min}$ and $\mathcal{U}_{\mathsf{emp}}$. Second, the worst-case transition matrices in $\mathcal{U}_{\mathsf{sa}}$ are extreme perturbations from $\boldsymbol{T}^0$. For instance, the coefficients $T^0_{1,RL}, ..., T^0_{n,RL}$ become $T^0_{1,RL} - \alpha_1, ..., T^0_{n,RL} - \alpha_n$, and the coefficients $T^0_{1,D}, ..., T^0_{n,D}$ become $T^0_{1,D} + 2 \cdot \alpha_1, ..., T^0_{n,D} + 2 \cdot \alpha_n$. In that sense, such coordinated, structured parameter misspecification appears unlikely. This is due to the ability to arbitrarily perturb the coefficients of $\boldsymbol{T}^0$, provided that the resulting rows still form a transition kernel, rather than accounting for potential relations across states that our factor matrix approach incorporates. Such extreme perturbations are unlikely to arise, which is why we focus our attention on the model of factor matrix uncertainty set.

Finally, we note that the worst-case matrices for our factor matrix uncertainty sets ($\mathcal{U}_{\min}$ and $\mathcal{U}_{\mathsf{emp}}$) are as far from the nominal estimation $\boldsymbol{T}^0$ as the worst-case matrices for $\mathcal{U}_{\mathsf{sa}}$, in terms of the 1-norm. This is because $\mathcal{U}_{\min}$ and $\mathcal{U}_{\mathsf{emp}}$ are centered around our nonnegative matrix factorization $\hat{\boldsymbol{T}}$ and not around $\boldsymbol{T}^0$. Therefore, both the (s, a)-rectangular uncertainty set $\mathcal{U}_{\mathsf{sa}}$ and the factor matrix uncertainty sets $\mathcal{U}_{\min}$ and $\mathcal{U}_{\mathsf{emp}}$ afford the same *budget of uncertainty*, in terms of numerical deviations from the nominal matrix $\boldsymbol{T}^0$. This means that the stark differences in empirical results and insights in our simulations with the hospital model are caused by the rank-constrained nature of the factor matrix uncertainty sets, and not by a difference in the radii of the uncertainty sets. From a modeling standpoint, underlying physiologic characteristics dictate the evolution of a patient's health. Thus, we expect the uncertainty to be reasonably captured by a deviation (from the nominal

estimation $\boldsymbol{T}^0$) that is not full-rank, and we expect the true worst-case performance of the threshold policies to be somewhere in between the performance in $\mathcal{U}_{\text{min}}$ and the performance in $\mathcal{U}_{\text{emp}}$.

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



(a) Random samples analysis. (b) Worst-case analysis.

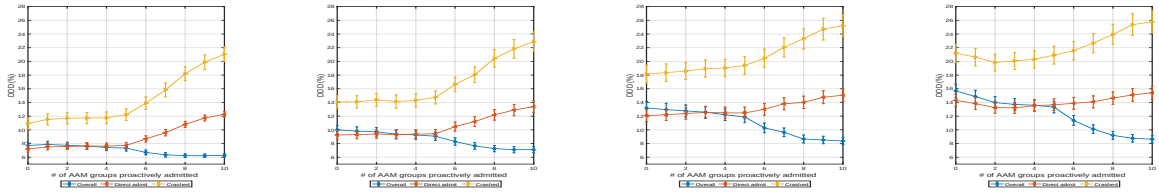
Figure 5 Length-Of-Stay of the 11 threshold policies for the nominal estimated matrix, randomly sampled matrices in the 95% confidence intervals (left-hand side), and the worst-case matrices found by our single MDP model (right-hand side). Each point corresponds to a threshold policy: the policy with 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 $\mathcal{U}_{\text{min}}$,$\mathcal{U}_{\text{emp}}$ and $\mathcal{U}_{\text{sa}}$ 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.

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 parameters deviations from the nominal matrix $\boldsymbol{T}^0$

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 $\mathcal{U}_{\mathsf{emp}}$ and $\mathcal{U}_{\mathsf{sa}}$. 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 $\mathcal{U}_{\mathsf{sa}}$, 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 in the previous paragraph for mortality, $\mathcal{U}_{\mathsf{sa}}$ allows more conservative parameter deviations, which may yield misleading empirical insights.

6.4.2. Impact of proactive transfers on demand-driven discharges. As the proactive transfer policies admit more patients to the ICU than reactive policies, they may increase ICU congestion and, consequently, the number of *demand-driven discharged* patients from the ICU. Such discharges are associated with worse outcomes (Chrusch et al. 2009b). In Figures 6a-6b, we explore the impact of proactive transfers onto the number of patients that are demand-driven discharged.

We can see that it is possible to proactively transfer up to the top 5 severity conditions without significantly impacting the proportions of demand-driven discharged patients. We also note that in this metric, the trends we see with mortality/length-of-stay are preserved. In other words, proactively transferring a small fraction of the riskiest patients (i.e., the highest AAM scores) can improve the average mortality rate/length-of-stay without significantly increasing the ICU occupancy or the number of demand-driven discharges. The findings are similar with the worst-case transition matrices (Figures 6b-6d), with the uncertainty set $\mathcal{U}_{\mathsf{sa}}$ leading to worse deterioration than our factor matrix uncertainty sets $\mathcal{U}_{\mathsf{emp}}$ and $\mathcal{U}_{\mathsf{sa}}$.



(a) Nominal. (b) $\mathcal{U} = \mathcal{U}_{\min}$. (c) $\mathcal{U} = \mathcal{U}_{\mathsf{emp}}$. (d) $\mathcal{U} = \mathcal{U}_{\mathsf{sa}}$.

Figure 6 Percentages of Demand-Driven Discharges (DDD) among ICU transfers in terms of the number of Advanced Alert Monitor (AAM) scores proactively transferred. We present the results for our nominal transition matrix and for the worst-case transition matrices in $\mathcal{U}_{\mathsf{sa}}, \mathcal{U}_{\mathsf{emp}}$ and $\mathcal{U}_{\min}$.

Because we do not allow proactive transfers when the ICU is full, demand-driven discharges only occur when there is an external arrival, when a patient crashes on the ward, or when a patient requires readmission to the ICU. Still, it would be concerning if the patients who are demand-driven discharged are among the most severe. We find that demand-driven discharged patients were discharged from the ICU after 75 to 85% of their ICU LOS, a duration which can often mean the patient has recovered enough to be safely transferred to a lower level of care (Lowery 1992).

6.4.3. Impact of potential waiting in the ward Our primary hospital model presented in Section 2 and used in the simulations thus far assumes that whenever a crashed, readmitted, or external patient arrives to a full ICU, it precipitates a demand-driven discharge patients from the ICU. In practice, it is possible that a patient requiring ICU admission when the ICU is full would need to *wait* in the ward until an ICU bed to become available.

We now numerically explore the impact of the possibility of waiting in the ward. In particular, we consider an alternative hospital model where demand-driven discharges never occur. The dynamics of the hospital is the same as in Section 2, except that if an ICU admission is required but the ICU is full, the new patient enters a *waiting queue*. Patients are served from the queue in a First-Come-First-Serve service discipline. Note that patients can enter the queue either as (i) direct admits, (ii) crashed patients, or (iii) readmits to the ICU. Our proactive transfer policies cannot send a patient to the queue. Every six hours, patients in the waiting queue either:

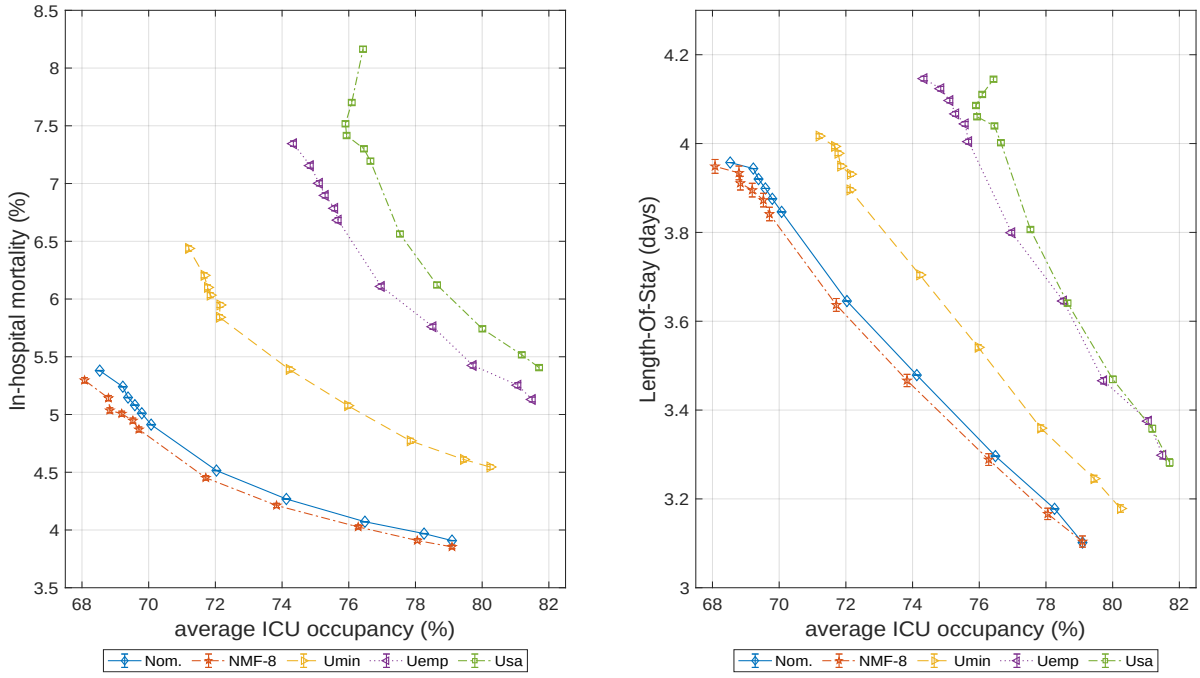
- Die in the queue: this happens with the same probability as that for the sickest patients (i.e., a severity score of 10), which amounts to a 6.84 % chance of dying every six hours. This is because patients requiring ICU admission have very severe conditions.
- Continue to wait in the queue. This would happen if there are no available ICU beds.
- Are transferred to the ICU. This will only occur when there is an available ICU bed (e.g., after a natural discharge). Priority is given to patients who have stayed longest in the queue. To focus on the impact of the queue, we assume patients from the queue who are admitted to the ICU have similar LOS/mortality risk as if they had not been in the queue.

We present our results below for this alternative model of hospital dynamics.

Impact of proactive transfers on hospital metrics. We present our results for the mortality rate and the LOS in the hospital in Figure 7, for this alternative model with a queue. The results are very similar to the primary hospital model without queue, even though the nominal metrics (mortality rate, ICU occupancy and LOS) are slightly worse. Despite these small quantitative differences, the qualitative insights are consistent. Threshold policies have the potential to improve in-hospital mortality rates and length-of-stay, at the price of an increase in ICU occupancy. Proactively transferring a small proportion of the patients (in our simulations, the top 10 % sickest patients) does not significantly increase the ICU occupancy. The results in worst-case metrics confirm these trends, with the rectangular uncertainty set $\mathcal{U}_{\mathsf{sa}}$ showing similar anomalous results as with the primary hospital model without queue.

Impact of proactive transfers on waiting queue metrics. We also consider the probability of entering the queue across different patient types (direct admits, crashed or readmitted), the average LOS in the ICU, the average length of the queue, and the average LOS in the queue.

In Figure 8a, we notice that the (nominal) probability a patient (direct admit, crashed or readmitted) enters the queue is increasing as more patients are proactively transferred to the ICU. This is because the ICU occupancy increases with the number of proactive transfers. Crucially, we see that the increase in probabilities does not increase substantially until proactive transfers become



(a) In-hospital Mortality (b) Length-Of-Stay

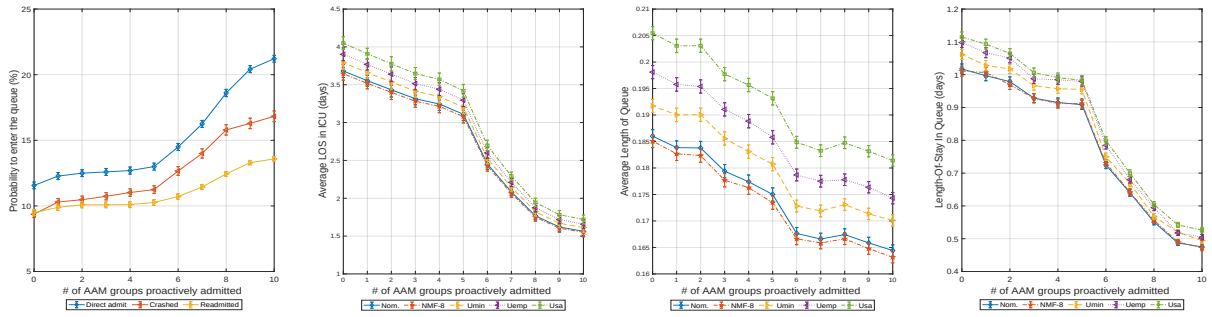
Figure 7 Average mortality and LOS in the hospital for the 11 threshold policies, in our hospital model with a queue.

quite aggressive (i.e. threshold higher than or equal to 5). Therefore, proactively transferring up to the worst five severity conditions does not result in significant changes in the number of patients in the queue. For the sake of brevity, we present our worst-case numerical experiments for this metric in Appendix L.

Additionally, in Figure 8b we note that proactively transferring patients also improves the average length-of-stay in the ICU. In particular, even though the average occupancy of the ICU is increasing when more patients are proactively transferred (see Figure 7), the patients who are admitted in the ICU tend to have shorter ICU LOS. This is because (i) proactively admitted patients have shorter ICU LOS than crashed patients from the same severity score, and (ii) as we proactively transfer more patients, patients enter the ICU from lower severity conditions and stay less time in the ICU than patients with more critical severity conditions. Therefore, even though the ICU

occupancy increases when more patients are proactively transferred, it is also the case that more patients can be admitted to the ICU (as patients in the ICU stay shorter periods of time).

As a consequence of shorter lengths of ICU visits, we observe that proactively transferring more patients has the beneficial effect of decreasing the average length of the queue and the average waiting time in the queue. (see Figures 8c and 8d). This may seem counter-intuitive given Figure 8a (where we see that more patients enter the queue as we proactively transfer more patients). However, following Figure 8b, we see that the ICU LOS is decreasing, and therefore, patients can be admitted to the ICU more often (than when fewer patients are proactively transferred). This is an important beneficial aspect of proactive transfer policies, as delays in patients admissions to the ICU are associated with worse mortality (Chalfin et al. 2007). The simulations with worst-case transition matrices show similar decrease in average queue length and waiting time as we proactively transfer more patients. Overall, this alternate hospital model with a waiting queue demonstrates



(a) Probability to enter the queue. (b) Average LOS in ICU. (c) Average length of the queue. (d) Average LOS in the queue

Figure 8 For our hospital with waiting queue, four different performance metrics in terms of the number of severity scores proactively transferred.

similar beneficial impact of proactive transfer policies as in the primary hospital model without queue. Note that in practice, we expect the discharge policy of the hospital to be a hybrid of these two models (some waiting and some demand-driven discharges). We have investigated these

two extreme scenarios and found that in both cases, proactively transferring the sickest patients may decrease the mortality rate and LOS of the hospital, without significantly increasing the ICU occupancy.

6.5. Practical Implications.

We now summarize here some practical implications and key take-aways of our numerical and theoretical analysis.

- **Impact of proactive transfer.** Proactively transferring patients to the ICU may improve the average mortality rate and LOS, at the price of increasing the ICU occupancy (Figure 4 and 5) and Demand-Driven Discharge (DDD) rates (Figures 6a-6d). Transferring only the sickest patients (here, the top 10 % of severity scores) does not lead to significant increases in ICU occupancy and DDD rates. The same conclusion holds when the hospital never demand-driven discharges patients from the ICU (Figures 7a-7b). In this case, proactively transferring patients may even improve the average time spent waiting in the queue (Figures 8a-8d).

- **Worst-case vs. random deviations.** The hospital metrics may significantly deteriorate for adversarial deviations in the transition parameters (Figures 4b and 5b). In stark contrast, the hospital metrics look fairly stable under standard sensitivity analysis, which considers random perturbations (Figures 4a and 5a). Therefore, classical random deviations analysis such as probabilistic multiway sensitivity analysis (Heitman et al. 2010) may to provide an overly optimistic estimations. Overall, this suggests that the naive approach of randomly perturbing the transition parameters to verify the effectiveness of our decisions may be misleading, as it does not take into account all potential events, some of which may have particularly adverse consequences even when the magnitude of the perturbation is small.

- **Unrelated vs. Related Uncertainty sets.** The deterioration of the hospital performance varies depending on the model of uncertainty; see Figures 4b and 5b. Compared to related models of uncertainty ($\mathcal{U}_{\min}$ and $\mathcal{U}_{\mathsf{emp}}$), an unrelated model of uncertainty ($\mathcal{U}_{sa}$) leads to both more extreme deterioration in performance as well as anomalous observations – e.g., that proactively transferring

some patients may decrease the average ICU occupancy. We note that the worst-case transition matrices for both types of uncertainty sets are similarly distanced from the nominal estimation $\boldsymbol{T}^0$ (as measured by the ℓ_1-norm). Therefore, the difference in performance and optimal robust policies can be attributed to the rank-constrained nature of $\mathcal{U}_{\min}$ and $\mathcal{U}_{\mathsf{emp}}$, compared to full-rank deviations in $\mathcal{U}_{\mathsf{sa}}$. Our numerical experiments shows that rectangularity indeed increases the level of conservatism of the model and may lead to erroneous insights on the impacts of transfer decisions. This lends additional support for the factor matrix uncertainty set being an appropriate model of uncertainty for healthcare applications.

Practical insights. Based on our theoretical and numerical analysis, our practical insights can be summarized as follows:

Threshold policies can be very effective to guide decisions for proactive transfers as they are 1) easy to implement and 2) have good theoretical and numerical performance. These properties hold when there is no parameter uncertainty as well as when data challenges introduce parameter uncertainty. Our results suggest that when there is parameter uncertainty (e.g., due to limited data and/or unobserved covariates), providers should be slightly more aggressive in their transfer policies as compared to when there is no uncertainty.

7. Conclusion 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 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 performance 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 which may also suffer from biases introduced by unobservable confounders. Since the severity scores are likely influenced by common underlying medical factors, we introduce a robust model that approximately accounts for potential relations 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, and that the optimal robust policy transfers more patients than the optimal nominal one. Our extensive simulations show that not accounting for parameter misspecification may lead to overly optimistic estimations of the hospital performance, 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 parameter 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 performance may be unlikely in practice, it is worth noting that the resulting parameter values for some worst-case performance 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 performance 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 mis-estimations of the metrics (average in-hospital mortality rate, 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 deterioration, within the confidence intervals given by some statistical estimators. It is to the discreton 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 levels 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 performance 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 (see 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

Adelman D, Mersereau AJ (2008) Relaxations of weakly coupled stochastic dynamic programs. *Operations Research* 56(3):712–727.

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`.

Barron Y (2016) Performance analysis of a reflected fluid production/inventory model. *Mathematical Methods of Operations Research* 83(1):1–31.

Barron Y (2018) A threshold policy in a Markov-modulated production system with server vacation: the case of continuous and batch supplies. *Advances in Applied Probability* 50(4):1246–1274.

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.

Caro F, Gupta AD (2015) Robust control of the multi-armed bandit problem. *Annals of Operations Research* 1–20.

Chalfin DB, Trzeciak S, Likourezos A, Baumann BM, Dellinger RP, study group DE, et al. (2007) Impact of delayed transfer of critically ill patients from the emergency department to the intensive care unit. *Critical care medicine* 35(6):1477–1483.

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 (2009a) 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`.

Chrusch C, Olafson K, McMillan P, Roberts D, Gray P (2009b) 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`.

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.

Gittins JC (1979) Bandit processes and dynamic allocation indices. *Journal of the Royal Statistical Society: Series B (Methodological)* 41(2):148–164.

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-Clément 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.

Heitman SJ, Hilsden RJ, Au F, Dowden S, Manns BJ (2010) Colorectal cancer screening for average-risk north americans: an economic evaluation. *PLoS medicine* 7(11):e1000370.

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.

Hsu YP (2018) Age of information: Whittle index for scheduling stochastic arrivals. *2018 IEEE International Symposium on Information Theory (ISIT)*, 2634–2638 (IEEE).

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 MJ, Lim AE (2016) Robust multiarmed bandit problems. *Management Science* 62(1):264–285.

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.

Lowery JC (1992) Simulation of a hospital's surgical suite and critical care area. *Proceedings of the 24th conference on Winter simulation*, 1071–1078.

Madani O, Hanks S, Condon A (1999) On the undecidability of probabilistic planning and infinite-horizon partially observable Markov decision problems. *AAAI/IAAI*, 541–548.

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.

May WL, Johnson WD (2000) Constructing two-sided simultaneous confidence intervals for multinomial proportions for small counts in a large number of cells. *Journal of statistical software* 5:1–24.

McClean 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 ressource 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.

Rasouli M, Saghafian S (2018) Robust partially observable Markov decision processes .

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.

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* .

Tripathi V, Modiano E (2019) A Whittle index approach to minimizing functions of age of information. *2019 57th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, 1160–1167 (IEEE).

Whittle P (1988) Restless bandits: Activity allocation in a changing world. *Journal of applied probability* 287–298.

Wiesemann W, Kuhn D, Rustem B (2013) Robust Markov decision processes. *Operations Research* 38(1):153–183.

Xu H, Mannor S (2010) Distributionally robust Markov decision processes. *Advances in Neural Information Processing Systems*, 2505–2513.

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.

Yu C, Liu J, Nemati S (2019) Reinforcement learning in healthcare: A survey. *arXiv preprint arXiv:1908.08796* .

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.

Summary statistics. The patient cohort is 53.80 % female, the average age is 67.34 years (standard deviation (std): 17.71). The average mortality is 3.2 % (9.5 % for patients who entered the ICU, 2.5 % for patients who never entered the ICU). The mean LOS is 90.5 hours (std: 135.2, mean of 149.1 hours for patients who entered the ICU and mean 81.0 for patients who never entered the ICU). 14.2 % of all hospitalizations were eventually admitted to the ICU. 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.

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 log-normally 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 $i \in \{1, ..., n\}$ following a non-homogeneous Poisson process. Every 6 hours, s/he then transitions to another risk score j with probability T^0_{ij}, 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 log-normally 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 log-normally 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 $\boldsymbol{T}$, for all severity score $i \in [n]$, the value function $\boldsymbol{V}$ of policy π satisfies $V_i \leq r_W + \lambda \cdot r_{RL}$. Let $i \in [n]$. By definition,

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.

V_i is the expected infinite-horizon reward starting from state i: $V_i = E^{\pi,\boldsymbol{T}}\left[\sum_{t=0}^{\infty}\lambda^t r_{i_t a_t}\,\middle|\, i_0 = i\right]$. Let us consider a trajectory O of the Markov chain on $\mathbb{S}$ associated with $(\pi,\boldsymbol{T})$. 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 $\dfrac{r_W\cdot(1-\lambda^t)}{1-\lambda}+\lambda^{t+1}\cdot\max\{r_{CR},r_{RL},r_D,r_{PT}\}$. Since the maximum instantaneous reward is r_{RL}, the reward associated with the trajectory O is smaller than $\dfrac{r_W\cdot(1-\lambda^t)}{1-\lambda}+\lambda^{t+1}\cdot r_{RL}$. Now

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

$$\le r_W+\lambda\cdot r_{RL}, \tag{B.2}$$

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 function $\boldsymbol{V}$ satisfies $V_i\le r_W+\lambda\cdot r_{RL},\forall\, i\in[n]$. □

Appendix C: Homogeneity and Translation for conditions (3.2) and (3.3).

LEMMA 2. *Let $(r_W,r_{CR},r_{RL},r_D,r_{PT})\in\mathbb{R}_+^5$ denote some rewards and $\boldsymbol{T}$ a transition matrix such that condition (3.2) and condition (3.3) hold.*

1. *Let $\alpha\ge 0$. For $\alpha\cdot(r_W,r_{CR},r_{RL},r_D,r_{PT})$ and $\boldsymbol{T}$, condition (3.2) and condition (3.3) still hold.*
2. *Let $\alpha\ge 0$. For $(r_W+\alpha,r_{CR}+\alpha,r_{RL}+\alpha,r_D+\alpha,r_{PT}+\alpha)$ and $\boldsymbol{T}$, condition (3.3) still holds.*

Proof. 1. This follows from $\alpha\ge 0$ and $\alpha\cdot r_{CR}\cdot T^0_{i,n+1}+\alpha\cdot r_{RL}\cdot T^0_{i,n+2}+\alpha\cdot r_D\cdot T^0_{i,n+3}=\alpha\cdot(r_{CR}\cdot T^0_{i,n+1}+r_{RL}\cdot T^0_{i,n+2}+r_D\cdot T^0_{i,n+3})$, and $\dfrac{\alpha\cdot r_W+\lambda\cdot\alpha\cdot r_{PT}}{\alpha\cdot r_W+\lambda\cdot\alpha\cdot r_{RL}}=\dfrac{r_W+\lambda\cdot r_{PT}}{r_W+\lambda\cdot r_{RL}}$.

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

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

Indeed, ϕ has a derivative in $\mathbb{R}_+$ and $\phi'(\alpha)=\frac{\lambda(1+\lambda)\cdot(r_{RL}-r_{PT})}{(r_W+\alpha+\lambda\cdot(r_{RL}+\alpha))^2}\geq 0$, since $r_{RL}\geq r_{PT}$. Therefore ϕ is a non-decreasing function, and for all $\alpha\geq 0$, the condition (3.3) holds. □

LEMMA 3. *Let $(r_W, r_{CR}, r_{RL}, r_D, r_{PT})\in\mathbb{R}^5_+$ denote some rewards and $\boldsymbol{T}$ a transition matrix such that condition* (3.6) *holds.*

1. *Let $\alpha\geq 0$. For $\alpha\cdot(r_W, r_{CR}, r_{RL}, r_D, r_{PT})$ and $\boldsymbol{T}$, condition* (3.6) *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 $\boldsymbol{T}$, condition* (3.6) *still holds.*

Proof. 1. Let α be a non-negative scalar. For the same reason as in Lemma 2, condition (3.6) still holds for $\alpha\cdot(r_W, r_{PT}, r_{RL}, r_{CR})$ and $\boldsymbol{T}$.

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

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

where $out(i)=r_{CR}\cdot T^0_{i,n+1}+r_{RL}\cdot T^0_{i,n+2}+r_D\cdot T^0_{i,n+3}$. 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 (3.6) 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 $F:\mathbb{R}^{n+4}\to\mathbb{R}^{n+4}$ denote the function that maps $\boldsymbol{V}\in\mathbb{R}^{n+4}$ to $F(\boldsymbol{V})$, where

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

From Puterman (1994), the function F is the Bellman operator associated with our single-patient MDP with transition kernel $\boldsymbol{T}^0$. If $\pi^\star$ is an optimal policy and $\boldsymbol{V}^\star$ is its value function, then $F(\boldsymbol{V}^\star)=\boldsymbol{V}^\star$, and $\pi^\star$ is such that

$$\pi^\star(i)=1 \iff r_W+\lambda\cdot r_{PT}\geq r_W+\lambda\cdot\sum_{j=1}^{n+3} T^0_{ij}V^\star_j, \forall\, i\in[n].$$

We will prove that $\pi^\star$ is a threshold policy by showing that $\left(r_W + \lambda \cdot \sum_{j=1}^{n+3} T_{ij}^0 V_j^\star\right)_{i\in[n]}$ is non-increasing. Indeed, this implies that if for $i \in [n]$ we have $\pi^\star(i) = 1$ then $\pi^\star(i') = 1$ for all $i \geq i'$. Let $i \in [n-1]$. We have

$$\begin{aligned}
r_W + \lambda \cdot \sum_{j=1}^{n+3} T_{ij}^0 V_j^\star &= r_W + \lambda \cdot \sum_{j=1}^{n} T_{ij}^0 V_j^\star + \lambda \cdot out(i) \\
&\geq r_W + \lambda \cdot \sum_{j=1}^{n} T_{ij}^0 \left(r_W + \lambda r_{PT}\right) + \lambda \cdot out(i) && \text{(D.1)} \\
&\geq r_W + \lambda \cdot \sum_{j=1}^{n} T_{i+1,j}^0 \left(r_W + \lambda r_{RL}\right) + \lambda \cdot out(i+1) && \text{(D.2)} \\
&\geq r_W + \lambda \cdot \sum_{j=1}^{n} T_{i+1,j}^0 V_j^\star + \lambda \cdot out(i+1) && \text{(D.3)} \\
&\geq r_W + \lambda \cdot \sum_{j=1}^{n+3} T_{i+1,j}^0 V_j^\star,
\end{aligned}$$

where Inequality (D.1) follows from the fact that $V_j^\star$ is larger (or equal) than $r_W + \lambda r_{PT}$ by definition of F and $\boldsymbol{V}^\star = F(\boldsymbol{V}^\star)$, (D.2) follows from Assumption 2, and (D.3) follows from Lemma 1. Therefore, we conclude that

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

Therefore, if for $i \in [n-1]$ we have

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

then for all $i' \geq n$ we have

$$r_W + \lambda \cdot r_{PT} \geq r_W + \lambda \cdot \sum_{j=1}^{n+3} T_{i'j}^0 V_j^t.$$

Therefore, we can conclude that the optimal nominal policy $\pi^\star$ is a threshold policy. □

Appendix E: Non-threshold optimal policies.

We provide an example of a single-patient MDP which does not satisfy Assumption 2 and where the optimal nominal policy is not threshold. The optimal policy in the MDP of Figure 9 is to proactively transfer a patient in state 1 and to not proactively transfer a patient in state 2.

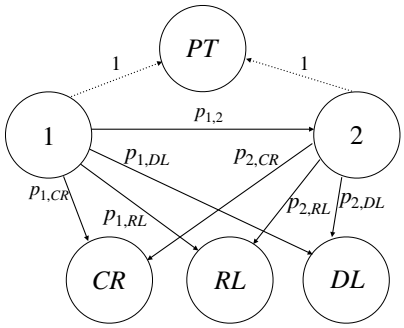


Figure 9 Example of an MDP where the optimal policy is not threshold. There is no self-transition in state 1 or 2. In state 1, the patient transitions to state $2, CR, RL$ or D (solid arcs), or is proactively transferred (dashed arc). In state 2, the patient has to exit the ward, either by proactive transfer , in which case s/he transitions to PT with probability 1 (dashed arc), either by transitioning CR, RL or D (solid arcs). The patient can not transition back to state 1. We provide values of the rewards and transitions for which the optimal policy is not threshold.

Condition (3.2) 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: $\pi^\star(1) = 0$. However,

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

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

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

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

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

and $\pi^\star(1) = 1$, 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: Proof of Proposition 1

Proof. The proof proceeds in two steps. Let $\pi^\star$ be an optimal nominal policy (not necessarily threshold) and $\boldsymbol{V}^\star$ be its value function.

1. Let π be any deterministic policy, $\boldsymbol{V}^\pi$ be its value function and let us write $F_\pi : \mathbb{R}^{n+4} \Rightarrow \mathbb{R}^{n+4}$ the map such that for any $i \in [n]$,

$$\begin{aligned}
\pi(i) = 1 &\Rightarrow F_\pi(\boldsymbol{V})_i = r_W + \lambda \cdot r_{PT}, \\
\pi(i) = 0 &\Rightarrow F_\pi(\boldsymbol{V})_i = r_W + \lambda \cdot \sum_{j=1}^{n+3} T_{ij}^0 V_j, \\
F_\pi(\boldsymbol{V})_{n+1} = r_{CR}, &F_\pi(\boldsymbol{V})_{n+2} = r_{RL}, F_\pi(\boldsymbol{V})_{n+3} = r_D, \\
F_\pi(\boldsymbol{V})_{n+4} &= r_{PT}.
\end{aligned}$$

Then for $\alpha > 0$ we have

$$\|F_\pi(\boldsymbol{V}^\star) - \boldsymbol{V}^\star\|_\infty \leq \alpha \Rightarrow \|\boldsymbol{V}^\pi - \boldsymbol{V}^\star\|_\infty \leq \alpha/(1-\lambda). \tag{F.1}$$

This in turn implies that $R(\pi) \leq R(\pi^\star) \leq R(\pi) + \alpha/(1-\lambda)$, since $R(\pi) = \boldsymbol{p}_0^\top \boldsymbol{V}^\pi$ and $R(\pi^\star) = \boldsymbol{p}_0^\top \boldsymbol{V}^\star$. The implication (F.1) holds because by definition $\boldsymbol{V}^\pi = F_\pi(\boldsymbol{V}^\pi)$ and if $\|F_\pi(\boldsymbol{V}^\star) - \boldsymbol{V}^\star\|_\infty \leq \alpha$ we have

$$\begin{aligned}
\|\boldsymbol{V}^\pi - \boldsymbol{V}^\star\|_\infty &= \|\boldsymbol{V}^\pi - F_\pi(\boldsymbol{V}^\star) + F_\pi(\boldsymbol{V}^\star) - \boldsymbol{V}^\star\|_\infty \\
&\leq \|\boldsymbol{V}^\pi - F_\pi(\boldsymbol{V}^\star)\|_\infty + \|F_\pi(\boldsymbol{V}^\star) - \boldsymbol{V}^\star\|_\infty \\
&\leq \|F_\pi(\boldsymbol{V}^\pi) - F_\pi(\boldsymbol{V}^\star)\|_\infty + \alpha \\
&\leq \lambda \|\boldsymbol{V}^\pi - \boldsymbol{V}^\star\|_\infty + \alpha.
\end{aligned}$$

2. Therefore, to show Proposition 1, we will construct a threshold policy π such that

$$\|F_\pi(\boldsymbol{V}^\star) - \boldsymbol{V}^\star\|_\infty \leq 2(n-1)\epsilon\lambda.$$

Assume that $\pi^\star$ is not threshold. Let $i_{\min}$ be the smallest severity condition such that $\pi^\star(i_{\min}) = 1$ (if there is no such $i_{\min}$, the optimal policy would not proactively transfer any patients and would be threshold). We can construct a threshold policy π such that $\|F_\pi(\boldsymbol{V}^\star) - \boldsymbol{V}^\star\|_\infty \leq 2(n-1)\epsilon\lambda$ as follows: $\pi(i) = 0$ for $i < i_{\min}$, and $\pi(i) = 1$ for $i \geq i_{\min}$.

- For $i < i_{\min}$, we have both $\pi^\star(i) = 0$ (by definition of $i_{\min}$) and $\pi(i) = 0$ (by definition of π). Therefore, for $i < i_{\min}$, we have

$$F_\pi(\boldsymbol{V}^\star)_i = F(\boldsymbol{V}^\star)_i, \forall\, i < i_{\min}. \tag{F.2}$$

- Let $i \geq i_{\min}$. Using the same method as in Appendix D for the proof of Theorem 1 but with conditions (3.4)-(3.5) instead of conditions (3.2)-(3.3), we can show that for any $i \in [n-1]$, we have

$$r_W + \lambda \cdot \sum_{j=1}^{n+3} T^0_{ij} V^\star_j \geq r_W + \lambda \cdot \sum_{j=1}^{n+3} T^0_{i+1,j} V^\star_j - 2\lambda\epsilon. \tag{F.3}$$

Then for all $n \geq i \geq i_{\min}$, we can iterate (F.3), and this yields

$$2\lambda\epsilon \cdot (i - i_{\min}) + r_W + \lambda \cdot \sum_{j=1}^{n+3} T^0_{i_{\min} j} V^\star_j \geq r_W + \lambda \cdot \sum_{j=1}^{n+3} T^0_{i,j} V^\star_j.$$

Note that $i - i_{\min} \leq n - 1$. Therefore, for all $i \geq i_{\min}$, we have

$$2(n-1)\lambda\epsilon + r_W + \lambda \cdot \sum_{j=1}^{n+3} T^0_{i_{\min} j} V^\star_j \geq r_W + \lambda \cdot \sum_{j=1}^{n+3} T^0_{i,j} V^\star_j.$$

Since $\pi^\star(i_{\min}) = 1$, we know that

$$V^\star_{i_{\min}} = r_W + \lambda \cdot r_{PT},$$
$$r_W + \lambda \cdot r_{PT} \geq r_W + \lambda \cdot \sum_{j=1}^{n+3} T^0_{i_{\min} j} V^\star_j.$$

This shows that for $i \geq i_{\min}$, we have

$$2(n-1)\lambda\epsilon + r_W + \lambda \cdot r_{PT} \geq r_W + \lambda \cdot \sum_{j=1}^{n+3} T^0_{i,j} V^\star_j. \tag{F.4}$$

Because $\pi^\star$ is an optimal policy, we always have $V^\star_i \geq r_W + \lambda \cdot r_{PT}$. Overall, this shows that for $i \geq i_{\min}$ we have

$$2(n-1)\lambda\epsilon + r_W + \lambda \cdot r_{PT} \geq r_W + \lambda \cdot \sum_{j=1}^{n+3} T^0_{i,j} V^\star_j \geq r_W + \lambda \cdot r_{PT}. \tag{F.5}$$

For $i \geq i_{\min}$, note that $\pi(i) = 1$ so that $F_\pi(V^\star)_i = r_W + \lambda \cdot r_{PT}$. Inequality (F.5) shows that for all $i \in [i_{\min}, n]$, we have

$$2(n-1)\lambda\epsilon + F_\pi(V^\star)_i \geq V^\star_i \geq F_\pi(V^\star)_i, \tag{F.6}$$

because for $i \in [i_{\min}, n]$, either we have $\pi^\star(i) = 1$ and then $V^\star_i = r_W + \lambda \cdot r_{PT} = F_\pi(\boldsymbol{V}^\star)_i$, either $\pi^\star(i) = 0$ and then $V^\star_i = r_W + \lambda \cdot \sum_{j=1}^{n+3} T^0_{i,j} V^\star_j$ and we can use (F.5).

- Overall, combining (F.2) and (F.6), we see that we have constructed a threshold policy π such that π such that $\|F_\pi(\boldsymbol{V}^\star) - \boldsymbol{V}^\star\|_\infty \leq 2(n-1)\lambda\epsilon$, which in turn implies that

$$R(\pi) \leq R(\pi^\star) \leq R(\pi) + \frac{2(n-1)\lambda\epsilon}{1-\lambda}.$$

□

Appendix G: Proofs of Section 4.3

The proof of Proposition 2 relies on three lemmas. We start by the following analysis of the sensitivity of Assumption 3 as regards the value of r_{PT}.

LEMMA 4. *Assume that Assumption 3 holds for $r_{PT} = 0$. Then Assumption 3 holds for all $r_{PT} \in [0, r_{RL}]$.*

Proof. Recall that Assumption 3 states that Assumption 2 holds for all $\boldsymbol{T} \in \mathcal{U}$, while Assumption 2 brings down to two conditions: condition (3.2) and condition (3.3).

Note that condition (3.2) does not depend of r_{PT}. If condition (3.3) is satisfied for $r_{PT} = 0$ and all $\boldsymbol{T} \in \mathcal{U}$, it is satisfied for any $r_{PT} \in [0, r_{RL}]$, since the left-hand side is increasing with r_{PT} while the right-hand side does not depend of r_{PT}. Therefore, if we assume that Assumption 3 holds for $r_{PT} = 0$, we know that it holds for any $r_{PT} \in [0, r_{RL}]$. □

We also prove the following lemma, which relates variations in r_{PT} to variation in the *nominal* value function.

LEMMA 5. *Let us fix the transition kernel to $\boldsymbol{T} \in \mathcal{U}$. Let $V^\star$ be the optimal nominal value function for a choice of r_{PT}. Now let us consider a reward $r'_{PT} = r_{PT} + \epsilon$. Let $\boldsymbol{V}^{\star'}$ be the new nominal value function. Then*

$$V_j^{\star'} \leq V_j^\star + \epsilon, \, \forall \, j \in [n]. \tag{G.1}$$

Note that (G.1) simply means that increasing the reward for PT by ϵ does not increase the optimal value function by more than ϵ.

Proof. We can prove (G.1) in the same way that we prove Lemma 1. In particular, by definition

$$V_j^{\star'} = E^{\pi, \boldsymbol{T}} \left[\sum_{t=0}^{+\infty} r_{i_t, a_t} \,|\, i_0 = j. \right]$$

We can consider a trajectory O of the Markov chain associated with $\pi^{\star'}, \boldsymbol{T}^0$. Along this trajectory, if the state PT is visited, at time $t \geq 1$ a reward $\lambda^t r'_{PT} = \lambda^t (r_{PT} + \epsilon)$ is obtained and the trajectory ends. Otherwise, the trajectory goes on until it reaches another absorbing state. Therefore, along this trajectory, the total accumulated reward when the reward for PT, r'_{PT}, is within $\lambda\epsilon$ of the total accumulated reward for the same trajectory when the reward for PT is r_{PT}. The reason for the factor λ is because the state PT has to be visited after at least one period, starting from a severity condition i. Taking the expectation over all possible trajectories, we obtain (G.1). □

We now prove Lemma 6, a variant of Lemma 5 in the case of *worst-case* value functions. Note that the proof of Lemma 6 is not as straightforward as for Lemma 5, because the worst-case kernels may be different for different values of r_{PT}.

LEMMA 6. *Let $V^\star$ the optimal robust value function for a choice of r_{PT}. Now let us consider a reward $r'_{PT} = r_{PT} + \epsilon$. Let $\boldsymbol{V}^{\star'}$ the new robust value function for r'_{PT}. Then*

$$V_j^{\star'} \leq V_j^\star + \epsilon, \, \forall \, j \in [n]. \tag{G.2}$$

Proof. For any pair $(\pi, \boldsymbol{T})$, we will write $\boldsymbol{V}^{\pi,\boldsymbol{T}}$ the value function when the decision-maker chooses π and the transition kernel is $\boldsymbol{T}$.

Let $\pi^{\star'}, \boldsymbol{T}^{\star'}$ be the pair of policy-kernel solving the robust MDP problem for the reward r'_{PT}, and let $\pi^\star, \boldsymbol{T}^\star$ be the pair of policy-kernel solving the robust MDP problem for the reward r_{PT}. We have, for $j \in [n]$,

$$\begin{aligned} V_j^{\star'} &= V_j^{\pi^{\star'}, \boldsymbol{T}^{\star'}} && \text{(G.3)} \\ &\leq V_j^{\pi^{\star'}, \boldsymbol{T}^\star} && \text{(G.4)} \\ &\leq V_j^{\pi^{\star'}, \boldsymbol{T}^\star} + \epsilon && \text{(G.5)} \\ &\leq V_j^{\pi^\star, \boldsymbol{T}^\star} + \epsilon, && \text{(G.6)} \end{aligned}$$

where (G.3) follows from the definition of $\left(\pi^{\star'}, \boldsymbol{T}^{\star'}\right)$, (G.4) follows from (4.3) in the robust maximum principle, (G.5) follows from Lemma 5, and (G.6) follows from (4.4) in the robust maximum principle. □

We are now ready to prove Proposition 2.

Proof of Proposition 2. Our proof proceeds in three steps.

• **Optimal policies are threshold.** Following Lemma 4, we know that Assumption 3 is satisfied for $r_{PT} \in [0, r_{RL}]$. Therefore, following Theorem 1, an optimal robust policy can be chosen threshold for every $r_{PT} \in [0, r_{RL}]$.

• $\mathcal{I}(r_{PT})$ **for** $r_{PT} = 0$ **and** $r_{PT} = r_{RL}$**.** Recall the Bellman equation: for a given choice of reward r_{PT}, let $\boldsymbol{V}^\star$ the value function of an optimal robust policy. Then the optimal action for severity condition i is chosen as the argmax in

$$\max\{r_W + \lambda \min_{\boldsymbol{T} \in \mathcal{U}} \boldsymbol{T}_i^\top \boldsymbol{V}^\star, r_W + \lambda r_{PT}\}. \tag{G.7}$$

When $r_{PT} = 0$, it is straightforward that the maximum of (G.7) is always the first term, i.e., the decision-maker always chooses to proactively transfer the patient. Therefore, when $r_{PT} = 0$, the optimal robust action for any severity condition is to not proactively transfer, i.e. $\mathcal{I}(0) = \emptyset$.

Now, we also know from Lemma 1 that $V_j^\star \leq r_W + \lambda r_{RL}$, for any $j \in [n]$ (as long as r_{RL} is the maximum of the instantaneous rewards). Therefore, when $r_{PT} = r_{RL}$, an optimal robust action at each severity condition i is to proactively transfer, i.e. $\mathcal{I}(r_{RL}) = [n]$.

• **Monotonicity of $\mathcal{I}(r_{PT})$.** We now prove that $\mathcal{I}(r_{PT})$ is monotonically increasing. Using Lemma 4, we can always choose an optimal policy that is threshold. Now we want to show that the threshold of the optimal policy is monotonically *non-increasing* (so that $\mathcal{I}(r_{PT})$ is increasing).

Let $i \in [n]$ and let us consider a choice of the reward r_{PT} such that

$$r_W + \lambda \min_{\boldsymbol{T} \in \mathcal{U}} \boldsymbol{T}_i^\top \boldsymbol{V}^\star \leq r_W + \lambda r_{PT}. \tag{G.8}$$

Note that $V^\star$ depends on r_{PT}. Now let us consider a reward $r'_{PT} = r_{PT} + \epsilon$. Let $\boldsymbol{V}^{\star'}$ the new robust value function. Our goal is to prove

$$r_W + \lambda \min_{\boldsymbol{T} \in \mathcal{U}} \boldsymbol{T}_i^\top \boldsymbol{V}^{\star'} < r_W + \lambda r_{PT} + \epsilon \tag{G.9}$$

so that the optimal robust policy for r'_{PT} still proactively transfers a patient in a severity condition i, and $r_{PT} \mapsto \mathcal{I}(r_{PT})$ is monotonically increasing. Now we obtain

$$\begin{aligned} r_W + \lambda \min_{\boldsymbol{T} \in \mathcal{U}} \boldsymbol{T}_i^\top \boldsymbol{V}^{\star'} &\leq r_W + \lambda \min_{\boldsymbol{T} \in \mathcal{U}} \boldsymbol{T}_i^\top \boldsymbol{V}^\star + \epsilon && \text{(G.10)} \\ &\leq r_W + \lambda r_{PT} + \lambda \epsilon && \text{(G.11)} \\ &\leq r_W + \lambda r'_{PT}, && \text{(G.12)} \end{aligned}$$

where (G.10) follows from (G.2) in Lemma 6 and the fact that $\boldsymbol{T}$ is a transition matrix, (G.11) follows from (G.8), and (G.12) follows from the definition of r'_{PT}.

□

Appendix H: 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 $\boldsymbol{T}^0$. This method yields

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

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 severe 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 $\boldsymbol{T}^0$ and $\hat{\boldsymbol{T}}$ are more important for some severity scores than others. Therefore, we compute the maximum absolute and relative deviations between each row of $\hat{\boldsymbol{T}}$ and each row of $\boldsymbol{T}^0$. 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, $T^0_{1,7} = 5.40 \cdot 10^{-5}$ 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 I: 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 (6.3)). We consider the impact of a change in the reward parameters presented in Section 6.3.1. We have seen in Section 6.3.2 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 $\mathcal{U}_{\mathsf{emp}}$). We choose to study the variations in r_{RL} and r_{PT-RL}, as these rewards have reversed influences on the thresholds of the optimal policies: 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 the variations in the hospital performance 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 Worst-case mortality and average ICU occupancy of the optimal robust policies for the uncertainty sets $\mathcal{U}_{\min}$,$\mathcal{U}_{\mathsf{emp}}$ and $\mathcal{U}_{\mathsf{sa}}$, for variations in the rewards r_{PT-RL} and r_{RL}.

r_{PT-RL}	Mort. (%)			ICU occupancy (%)			r_{RL}	Mort. (%)			ICU occupancy (%)		
	$\mathcal{U}_{\min}$	$\mathcal{U}_{\mathsf{emp}}$	$\mathcal{U}_{\mathsf{sa}}$	$\mathcal{U}_{\min}$	$\mathcal{U}_{\mathsf{emp}}$	$\mathcal{U}_{\mathsf{sa}}$		$\mathcal{U}_{\min}$	$\mathcal{U}_{\mathsf{emp}}$	$\mathcal{U}_{\mathsf{sa}}$	$\mathcal{U}_{\min}$	$\mathcal{U}_{\mathsf{emp}}$	$\mathcal{U}_{\mathsf{sa}}$
180	5.73	6.47	7.02	71.88	75.34	77.25	240	4.78	5.46	6.02	75.4	77.97	78.96
185	5.57	6.35	7.02	72.34	75.62	77.34	245	5.11	5.82	6.41	73.95	76.70	77.81
190	5.11	5.83	6.40	73.95	76.70	77.80	250	5.11	5.83	6.40	73.95	76.70	77.80
195	5.11	5.46	6.02	73.94	77.98	78.96	255	5.57	6.36	6.41	72.33	75.63	77.81
200	4.78	5.12	5.64	75.39	79.03	80.02	260	5.57	6.36	7.02	72.33	75.63	77.24

In Table 3, we notice that the hospital performance 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 $r_{PT-RL} = 260$. In contrast, when the optimal robust threshold is the same, *variations in rewards value do not impact the hospital worst-case performance.* For instance, the optimal robust policies are the same

(equal to $\pi^{[5]}$ across all three uncertainty sets) when $r_{RL} = 250$ and when $r_{RL} = 245$. The hospital worst-case performance (mortality, LOS, ICU occupancy) are also the same, when $r_{RL} = 250$ and when $r_{RL} = 245$. However, these two rows of worst-case performance 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 performance 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 performance, since variations in the rewards parameters on the single-patient MDP side still yield worst-case hospital performance that are very similar.

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

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

Errors of the NMF approximations. For $r = 7$, we compute a new $\hat{\boldsymbol{T}}$ solution of the NMF optimization program. Our solution $\hat{\boldsymbol{T}} = \boldsymbol{U}\hat{\boldsymbol{W}}^\top$ achieves the following errors: $\|\boldsymbol{T}^0 - \hat{\boldsymbol{T}}\|_1 = 0.1932$, $\|\boldsymbol{T}^0 - \hat{\boldsymbol{T}}\|_\infty = 0.0224$, $\|\boldsymbol{T}^0 - \hat{\boldsymbol{T}}\|_{\mathsf{relat},\mathsf{T}^0} = 0.4093$. In more details, $\hat{\boldsymbol{T}}$ 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 4 Statistics of the absolute and relative deviations of $\hat{\boldsymbol{T}}$ from $\boldsymbol{T}^0$ for a rank $r = 7$.

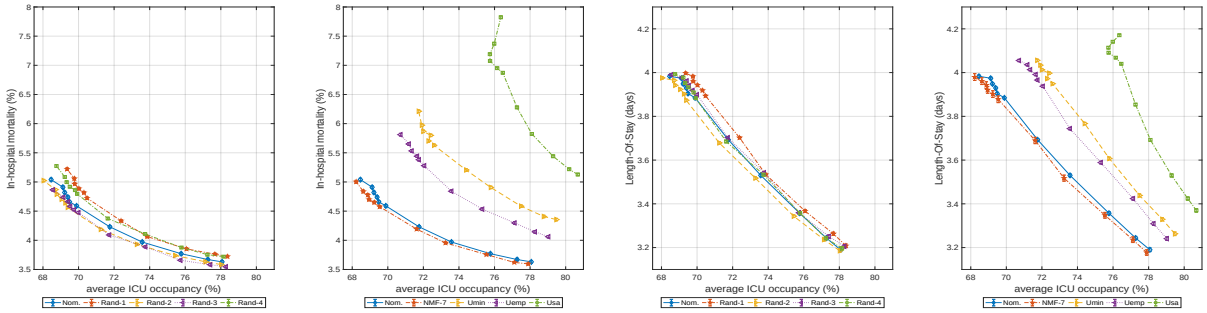
As we can see in Table 4, the absolute deviations remains small. Additionally, the relative differences between the coefficients are moderate, with half being less than 8.56%. That said, the maximum relative different is 40.93%. This occurs with $\hat{T}_{4,6} = 0.0035$, while $T^0_{4,6} = 0.0060$; so 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 performance of the 11 threshold policies, for our uncertainty sets $\mathcal{U}_{\min}$ and $\mathcal{U}_{\mathsf{emp}}$, when the rank is $r = 7$. For references we still show the performance for the nominal transition kernel $\boldsymbol{T}^0$ (nominal performance), for the uncertainty set $\mathcal{U}_{\mathsf{sa}}$ and for our NMF solution of rank $r = 7$.

We first note that the hospital performance with our NMF approximation of rank $r = 7$ are very close to the hospital performance for $\boldsymbol{T}^0$, which provides support that $\hat{\boldsymbol{T}}$ is a plausible transition matrix. We notice that the performance of the threshold policies can still significantly deteriorate,

even for small variations from the nominal matrix $\boldsymbol{T}^0$. In particular, there is a 20% increase in the average mortality, for some worst-cases matrices in $\mathcal{U}_{\min}$ and $\mathcal{U}_{\mathsf{emp}}$. Interestingly, the uncertainty set $\mathcal{U}_{\min}$ yields worst-case mortality rates that are higher than for worst-cases matrices in $\mathcal{U}_{\mathsf{emp}}$, contrary to what we noticed in Section 6 for rank $r = 8$. However, these two uncertainty sets still yield the same insights, which are that the performance 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 performance. Moreover, these insights are still different from the worst-cases performance in $\mathcal{U}_{\mathsf{sa}}$, since the results for $\mathcal{U}_{\mathsf{sa}}$ are independent of the rank chosen for our NMF approximation. In particular, for worst-cases in $\mathcal{U}_{\mathsf{sa}}$, 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 6 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 (mortality). (b) Worst-case analysis (mortality). (c) Random samples analysis (LOS). (d) Worst-case analysis (LOS).

Figure 10 In-hospital mortality and length-of-stay of the 11 threshold policies for the nominal estimated matrix, randomly sampled matrices in the 95% confidence intervals and the worst-case matrices found by our single MDP model (right-hand side).

Appendix K: Hospital performance for random deviations around the nominal kernel.

We sample at random 20 matrices in the confidence intervals (6.1). In order to do so, we first sample a matrix of deviations $\boldsymbol{D} \in \mathbb{R}^{10\times 13}$, with $D_{ij} \in [-\alpha_i, +2\cdot\alpha_i], \forall (i,j) \in [10]\times[13]$. Note that the

matrix $\boldsymbol{T}^0 + \boldsymbol{D}$ 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 $\boldsymbol{T}^0 + \boldsymbol{D}$ onto the simplex and we obtain a matrix a new matrix $\tilde{\boldsymbol{T}}$. If the corresponding matrix $\tilde{\boldsymbol{T}}$ is inside the confidence-intervals, we compute the hospital performance of the 11 threshold policies. Otherwise, we reject $\tilde{\boldsymbol{T}}$ and sample a new deviation matrix $\boldsymbol{D}$.

Using this method, we compute the performance of the threshold policies for 20 matrices chosen randomly inside the confidence intervals (6.1). Out of these 20 simulations, 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 performance ranged from 3.00% to 3.84% (from threshold 0 to threshold 11). For each policy, the maximum relative deviation from the nominal performance 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 performance ranged from 0.34% to 1.04% (for threshold 3 and threshold 11). For each policy, the maximum relative deviation from the nominal performance 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 performance ranged from 0.37% to 0.57% (for threshold 0 and threshold 11). For each policy, the maximum relative deviation from the nominal performance ranged from 0.99% to 1.57% (for threshold 11 and threshold 0).

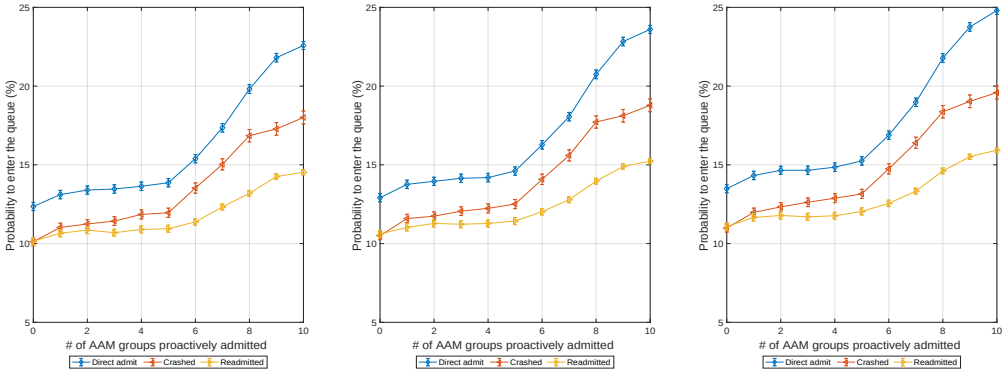
Appendix L: Additional figures for Section 6.4.3.

L.1. Worst-case simulations for the probability to enter the queue

We present in Figure 11 the worst-case probabilities that a patient of a certain type (direct admit, crashed and readmitted) will enter the waiting queue. The nominal probabilities are given in Figure 8a. We notice that in the worst-case, the probabilities can moderately deteriorate; still, the trends remain the same as in the nominal case. Namely, the probability that a certain type of patient enters the queue remains fairly stable until the threshold increases above 5 (proactively transferring the patients with the top 10 % riskiest severity conditions).

L.2. Worst-case simulations for the proportion of patient types in the queue

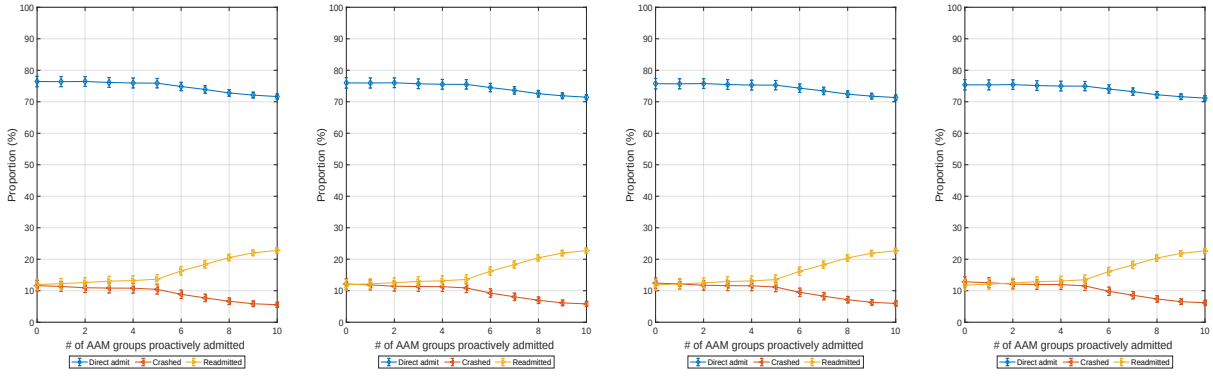
We present here our worst-case simulations for the proportion of patient types in the waiting queue. We note that the worst-case results are very similar to the nominal case, which is as expected since we are computing *proportions* (and not the absolute number of patient types in the queue). The only change compared to the nominal case is that the proportion of crashed patients slightly increases



(a) $\mathcal{U} = \mathcal{U}_{\min}$. (b) $\mathcal{U} = \mathcal{U}_{\text{emp}}$. (c) $\mathcal{U} = \mathcal{U}_{\text{sa}}$.

Figure 11 For different uncertainty sets, worst-case probability to enter the queue for different patient types (direct admits, crashed and readmitted), for different threshold policies.

as the worst-case transition matrices chosen in the uncertainty sets are increasing the likelihood of crash. The first six threshold policies only moderately increase the worst-case proportions.



(a) Nominal. (b) $\mathcal{U} = \mathcal{U}_{\min}$. (c) $\mathcal{U} = \mathcal{U}_{\text{emp}}$. (d) $\mathcal{U} = \mathcal{U}_{\text{sa}}$.

Figure 12 For the nominal matrix and for the worst-case for different uncertainty sets, proportions of patient types (direct admits, crashed and readmitted) in the waiting queue for different threshold policies.