Extubation Decision Making with Predictive Information for Mechanically Ventilated Patients in ICU

Guang Cheng
Institute of Operations Research and Analytics, National University of Singapore, Singapore 117602, gcheng@u.nus.edu

Jingui Xie
TUM School of Management, Technical University of Munich, Heilbronn, Germany 74076, jingui.xie@tum.de

Zhichao Zheng
Lee Kong Chian School of Business, Singapore Management University, Singapore 178899, danielzheng@smu.edu.sg

Haidong Luo
Department of Cardiac, Thoracic & Vascular Surgery, National University Hospital, Singapore 119074, hai_dong_luo@nuhs.edu.sg

Oon Cheong Ooi
Department of Cardiac, Thoracic & Vascular Surgery, National University Hospital, Singapore 119074, oon_cheong_ooi@nuhs.edu.sg

Weaning patients from mechanical ventilators is a critical decision in intensive care units (ICUs), significantly affecting patient outcomes and the throughput of ICU, especially during the COVID-19 pandemic. In this study, we aim to improve the current extubation protocols by incorporating predictive information on patient health conditions. We develop a discrete-time, finite-horizon Markov decision process (MDP) with predictions on future information to support the extubation decision. We characterize the structure of the optimal policy and provide important insights into how predictive information can lead to different decision protocols. We prove that adding predive information is always beneficial, even if the physicians overtrust the predictions as long as the predictive model is moderately accurate. Using a comprehensive data set from an ICU in a tertiary hospital in Singapore, we compare the performance of different policies and demonstrate that incorporating predictive information can reduce ICU length of stay (LOS) by up to 9.4% and, simultaneously, decrease the extubation failure rate by up to 18.9%. The benefits are more significant for patients with poor initial conditions at ICU admission. Furthermore, simply optimizing LOS using a classical MDP model without incorporating predictive information leads to an increased extubation failure rate by up to 6%. Using publicly available data on critical COVID-19 patients, we show that applying the extubation protocols using predictive information can improve ICU throughput by up to 9.0%.

Key words: Intensive care unit; mechanical ventilation; extubation; predictive information

1. Introduction

Healthcare-related costs have risen dramatically in recent years, and some have been identified as inefficient and redundant (Olsen et al. 2010). Meanwhile, growing healthcare demand drives service providers to practice fast-track management, which requires rapid and accurate decision-making at every stage of patient service, including the services provided by intensive care units (ICUs) (Wilmore and Kehlet 2001, Zhu et al. 2012). On the other hand, with increasing data availability, advanced machine learning techniques, and increasing computational power, highly accurate

predictive models are becoming more accessible. Although considerable attention has been paid to developing predictive models in the healthcare context, how to leverage predictive information to improve the quality of medical decisions remains understudied. In this paper, we aim to integrate predictive information with prescriptive analytics to evaluate how predictive information improves the quality of decision-making in healthcare service delivery. In particular, we investigate *how and to what extent the predictive information can improve the outcomes of the extubation decision—i.e., weaning patients from mechanical ventilators–in ICUs.*

In ICUs, most patients require invasive mechanical ventilation (MV), which provides support for breathing. Patients are not allowed to be discharged from ICUs if they are intubated (i.e., receiving MV support). Since critical care is costly for both patients and hospitals (Barrett et al. 2014, Halpern and Pastores 2015), the decision to extubate is critical for these patients, and time to extubation is usually regarded as the first primary service outcome for surgical care in hospitals (Wong et al. 2016). On the one hand, early extubation increases the risk of extubation failure (defined as the need for re-intubation or other breathing support measures) if the patient cannot breathe spontaneously. Extubation failure is significantly associated with longer ICU length of stay (LOS) and higher mortality rate (Epstein et al. 1997, Thille et al. 2013). On the other hand, prolonged ventilation wastes limited ICU resources and capacity, leading to the units being congested. During the COVID-19 pandemic, ventilators become a scarce resource, and physicians must ration ventilators among critically ill COVID-19 patients (Cohen et al. 2020). Moreover, due to endotracheal tube placement, invasive ventilation is painful for patients, who can develop ventilator-associated complications, including airway injury, alveolar damage, and ventilator-associated pneumonia (Hess 2011), which worsen patients' respiratory conditions.

Currently, healthcare decisions typically follow protocols that consider current physiological data or historical clinical events based on clinical trials and experience. However, this could be improvident in medical decision-making, especially for critical care. For instance, a patient may deteriorate in hours without intervention, and if conditions worsen, it may be too late to act. Early detection of imminent deterioration and preventive intervention can help save lives in critical care settings (Blum 2018). Therefore, in practice, physicians may make personal predictions about patients' future conditions based on their own experience. However, the accuracy of subjective prediction is limited and varies significantly between individuals. Massive amounts of data are underutilized, and important signals hidden in the data could easily be overlooked. Using appropriately validated and accurate predictive models, we shed light on the design of medical decision protocols that incorporate predictive information; these may outperform current decisions.

Thus motivated, we aim to investigate the optimal time to extubate patients in MV. In recent years, advancements in machine learning and econometric analysis have rendered predictions of

patients' future conditions and treatment effects more accessible and accurate. In this paper, we develop models that take advantage of predictive information to prescribe when to extubate patients. We formulate a Markov decision process (MDP) model to characterize the optimal stopping problem with predictive information, which is subsequently applied to the extubation problem. The proposed model can be extended to many crucial decisions in ICUs, such as stopping continuous renal replacement therapy (also known as dialysis) for patients who develop acute kidney injury after major operations. The entirety of ICU service can be regarded as a medical intervention, and discharge can be seen as an optimal stopping problem.

Main contributions of our work are summarized as follows:

- We show that the classical MDP model can be leveraged to embed predictions on future information for the extubation decision and characterize the structure of the optimal policy. Using this proposed modeling framework, we theoretically examine the value gained by incorporating predictive information. We show that predictive information is always valuable if the misclassification matrix is known. When the misclassification matrix is unknown, we derive the sufficient and necessary condition under which the predictive information can still be beneficial even if the noisy prediction is considered to be perfectly accurate—i.e., the optimal policy with perfect prediction is applied in the environment with noisy prediction—in a special case with two patient classes. A necessary condition for the latter case to be possible is that the predictive model is barely accurate, which can be guaranteed if the model's area under the receiver operating characteristic curve (AUC) is greater than or equal to 0.5. This result provides stronger evidence for the common belief that the model must have an AUC above 0.5 to be useful in practice.
- We implement our models for the extubation problem in ICUs, a critical component of fast-track management, to evaluate the effectiveness of incorporating predictive information in medical decision-making. We calibrate the models using a comprehensive set of medical data collected from an ICU and demonstrate that predictive information effectively reduces both ICU LOS and extubation failure rate (EFR), which is closely related to in-hospital mortality. We further analyze the impact of prediction accuracy and show that even a moderately accurate forecast can substantially improve decision quality. These results provide support for more precise and personalized fast-track management of ICUs. We further estimate potential benefits by applying the proposed model during the ongoing global pandemic of coronavirus disease 2019 (COVID-19).

The rest of the paper is organized as follows. We review the related literature in Section 2. Section 3 introduces the clinical settings and data. In Section 4, we introduce our modeling framework and derive the structural results of the optimal policy for extubation decision with predictions. We further discuss the clinical settings and data, based on which, we estimate the parameters in the

proposed model. In Section 5, we present the performance of our model to evaluate the effectiveness of predictive information in extubation decision, and Section 6 concludes.

2. Literature Review

Our work is related to five streams of literature: (1) ICU extubation decision-making; (3) integration of predictive and prescriptive analytics; (4) MDP and its applications in healthcare; and (5) optimal stopping and its applications.

2.1. ICU Extubation Problem

Being one of the most critical decisions in ICUs, the extubation problem received a lot of attention in the medical literature (Thille et al. 2013, Randolph et al. 2002, Xie et al. 2019, Chung et al. 2020), most of which focus on understanding the risk factors associated with extubation failures and the impact of different extubation protocols on patient outcomes. There is a growing body of studies on applying machine learning models in the extubation decision problem in recent years. Most of these studies tried to predict extubation outcomes using machine learning methods. For instance, Wang et al. (2010) used the adaptive neuro-fuzzy inference system to predict blood gas exchange for the management of ventilated patients in the ICU, and Kuo et al. (2015) applied an artificial neural network to predict extubation outcomes for mechanically ventilated patients. Tsai et al. (2019) combined machine learning techniques with Bayesian decision analysis. They predicted extubation outcomes in the first step, then used Bayes' rule to update the belief about patients' health condition. However, as Kwong et al. (2019) pointed out, only a limited number of studies have started to assess the efficacy and effectiveness of machine learning techniques for mechanical ventilation in the ICU. Along this direction, Prasad et al. (2017) and Jagannatha et al. (2018) applied reinforcement learning to learn the sedation prescription and extubation decision for ventilated patients. They assumed Markov transition of patient health condition and formulated the reinforcement learning model as an MDP. They used the fitted Q-iteration algorithm (Riedmiller 2005) to learn the reward function from the data and then chose the policy with the maximum reward. However, the policy thus learned is the probability of choosing a specific dosage level for sedation and whether to extubate the patient, which is difficult to interpret and challenging to implement in practice. Moreover, Prasad et al. (2017) acknowledged the issue of selection bias in the data, and the reinforcement learning model fails to account for this in the policy learning.

Unlike this branch of literature, our work aims to leverage the predictive information to make more precised and personalized extubation decisions, which is not considered in reinforcement learning models. We also want to derive a policy that is interpretable and implementable in practice. Hence, we leverage the classical MDP model, which has been widely used in medical research and is known for its interpretability among physicians. Moreover, we try to handle the selection bias for

parameter estimation using the advanced econometrics methods. The policy we derive is a personalized extubation recommendation based on both the current health condition and the predicted treatment effect of continued intubation for each patient. In this sense, our work broadly relates to numerous studies on applying operations research and machine learning techniques in personalized medical decision-making (Mišić et al. 2010, Akan et al. 2012, Chan et al. 2013, Bertsimas et al. 2017, Ayer et al. 2019, Hu et al. 2018). In line with this stream of literature, we apply the machine learning models to predict the treatment effect, which is equivalent to predicting patients' future health conditions under treatment. However, we are interested in investigating how to embed such predictive information into sequential decision-making to provide personalized recommendations for treatment continuation decisions.

2.2. Integration of Predictive and Prescriptive Analytics

How to integrate predictive information into prescriptive analytics has recently drawn much attention in the operations research and management science community. Bertsimas and Kallus (2019) demonstrated that traditional models led to inadequate solutions in the big data era and proposed a framework that combined traditional optimization models with predictive information to prescribe more efficient solutions to single-stage decision problems. They argued that the parameters conventionally computed using the sample mean were insufficient. These parameters could be more precisely predicted using machine learning methods, and thus they used the results from a predictive model as input parameters in optimization models. Bertsimas et al. (2016) developed a statistical model to predict the outcomes under different chemotherapy regimens for patients with cancer and used the predictive results to parameterize an optimization model to recommend chemotherapy regimens. Spencer et al. (2014) performed a pilot study on an admission control problem with forecast future arrivals. Formulating this problem as a queuing model, they showed that future information could significantly reduce the system delay. However, the improvement was at the expense of a lengthy look-ahead window, which required more sophisticated and accurate predictive models. To assess whether less future information can still achieve a significant improvement, Xu (2015) discussed the necessity of a nontrivial forecast window for arrivals in queuing models. The results showed that the performance would be no better than an online policy if the forecast windows were not long enough. Based on these results, Xu and Chan (2016) assumed that the arrivals in the future w time slots could be forecast, and proposed a proactive policy for admission control in the emergency department. Simulation results showed that the proposed proactive policy outperformed the online policy, even when the predictions contained noise. Dai and Shi (2018a) developed a framework that supported on inpatient overflow decision in hospitals. They formulated the multi-class parallel-queue system as an MDP, in which the system state

included the count of waiting patients and patients who would be discharged during the same day. They assumed that the to-be-discharged information was prior knowledge and solved the problem using an approximate dynamic programming approach due to the curse of dimensionality. Shi et al. (2018) formulated an MDP that prescribed ICU discharge decisions. In their model, the state variable was the number of patients within each patient class, which corresponded to a particular risk of readmission. The evolution of readmission risk was assumed to be dependent only on prior ICU stays. They proposed a heuristic policy that minimized total ICU LOS. In their case study, a model that predicted personalized readmission risk was embedded, with each patient viewed as a patient class. An increasing number of studies incorporated predictive information into prescriptive analytics with applications in various domains (Chen and Farias 2013, Huang et al. 2018). Our research, while similar to the above literature in terms of leveraging future predictive information in decision-making, differs with respect to problem formulation and methodology.

2.3. MDP and Its Applications in Healthcare

Our model builds on the theory of MDP (Bellman 1957). Markov chain (MC) models have long been used by the medical community to model disease progression, which is the backbone of many healthcare decision-making problems. MDP models enable decision-makers to choose the action that makes the correct trade-offs between immediate rewards and uncertain future gains, while observing the current situation and yielding the best possible solution. MDP has been widely adopted in the healthcare domain when facing dynamic and stochastic decision-making problems (Schaefer et al. 2005). For instance, a substantial body of research has investigated MDP applications in liver transplantation. Alagoz et al. (2004) formulated an MDP model to determine the optimal timing for living-donor liver transplantation. Similarly, MDP was used to obtain the optimal policies for choosing between living-donor and cadaveric livers (Alagoz et al. 2007a), and deciding whether to accept or decline a cadaveric liver for a patient in need of liver transplantation (Alagoz et al. 2007b). MDP has also been used in HIV therapy decisions (Shechter et al. 2008), initiating cholesterol lowering medication for patients with diabetes (Denton et al. 2009), biopsy decisions (Chhatwal et al. 2010), and post-mammography diagnosis decisions (Ayvaci et al. 2012).

It is worthwhile mentioning that our model can alternatively be formulated as a partially observable Markov decision process (POMDP), which is a generalization of MDP that allows the states to not be completely observable. In the healthcare community, POMDP has recently gained much attention and has been successfully used to design personalized screening policies for breast cancer (Ayer et al. 2012, 2016) prostate cancer (Zhang et al. 2012), and colorectal cancer (Erenay et al. 2014). Our model differs from classical MDP and POMDP models by incorporating predictions about the future state into the state space. In practice, predictions usually come from machine

learning models built and validated using historical data. In this sense, our model is similar to higher-order Markov chains in which system transition depends on historical states (Raftery 1985, Ching et al. 2008, 2013). However, higher-order Markov chains typically require that exponential numbers of parameters be characterized; otherwise restrictive assumptions must be imposed on the dependence structure to reduce the number of parameters. As a result, higher-order Markov decision processes (HMDP) are not easy to optimize and have very limited application in practice (Ching et al. 2004). Our approach allows more flexibility in capturing historical information—which is not restricted to state information, as in HMDP—and leveraging advanced machine learning methods to project historical information onto a single prediction state. Therefore, our model can efficiently reduce both the number of parameters and the computational complexity that HMDP suffers from.

2.4. Optimal Stopping and Its Applications

Our model also relates to the literature on the optimal stopping problem (Chow et al. 1971) that arises in many applications, such as finance (Haugh and Kogan 2004, Detemple 2005, Boyarchenko and Levendorskii 2007, Sturt 2021), healthcare (Alagoz et al. 2004), and marketing (Feng and Gallego 1995, Erat and Kavadias 2008). State-of-the-art work on the optimal stopping problem entails approximating value functions and identifying greedy policies by approximate dynamic programming, simulation, and duality (Goldberg and Chen 2018). For instance, Desai et al. (2012) used the pathwise optimization method to obtain upper and lower bounds on the value function of an optimal stopping problem. Li and Linetsky (2013) solved a class of optimal stopping problems by eigenfunction expansion. More recently, Ciocan and Mišić (2020) proposed a binary tree-based greedy policy for the optimal stopping problem, and characterized the depth needed for a tree policy to approximate any optimal policy to a given precision. They formulated the problem of learning such policies from the data as a sample average approximation problem. The advantage of their approach is that it is data-driven and interpretable, and thus easy to implement in practice. The proposed policy was shown to be superior to other methods by applying it to an option pricing problem. In a different vein, Iancu et al. (2020) considered a robust optimal stopping problem in continuous time that the system state cannot be observed except for a limited number of monitoring times, which the decision-maker must choose in conjunction with a suitable stopping policy. Unlike the conventional models that characterize the system evolution through specified stochastic processes, they adopted a robust optimization framework and described the system states using an uncertainty set that captures the possible system values at future monitoring times. The system evolution is modeled by updating the bounding functions—which put lower and upper bounds on the future system values in the uncertainty set—after each observation. The decision-maker's objective is to maximize the worst-case total reward. Our work adopts the conventional

approach that the system evolution is characterized by a Markov process, but differs from the typical optimal stopping problem by integrating predictions regarding future information into the state space.

3. Clinical Settings and Data

Our study is based on the data collected from the cardiothoracic ICU of a tertiary hospital in Singapore. The data set contains 5,566 ICU admissions from March 2010 to October 2016. For each ICU admission, we compile patient-level admission data, such as age, gender, race and time of admission. During the ICU stay, comprehensive physiological data, such as body temperature, heart rate, and blood pressure, were documented by a digital tracking system. Laboratory test outcomes, medications, procedures, and nursing care notes were also integrated into the data set. Clinical variables were usually updated every 15 minutes to 4 hours, except for laboratory tests, which were often updated daily. The data set provides detailed information about MV. Nursing staff documented the time of intubation and extubation, type of endotracheal tube, and ventilator mode at MV initiation. Patient conditions were recorded every 15 minutes to 4 hours on a ventilator. Based on this rich data, we can identify patients who received MV and observe the ventilation duration and the outcomes of extubation, monitor a patient's respiratory condition—which is associated with the risk of extubation failure—and construct a model that predicts a patient's future condition. Following previous medical studies (Demling et al. 1988, Girault et al. 2011) and the practice in our partner ICU, we define extubation failure as the need for re-intubation or noninvasive ventilation (NIV) within 7 days after extubation.

The data contains 4,026 cases with MV initiation. We apply the same set of exclusion criteria as those used in both the medical literature and the reinforcement learning models (Epstein et al. 1997, Fontela et al. 2005, Chen et al. 2019, Prasad et al. 2017). We first eliminate readmitted cases within 30 days. Note that we keep the first admission in this step. Although patients cannot be identified as new admissions or readmissions in the first month, we do not have cases with MV in the first month of the data. Patients with excessively long LOS typically have rare conditions and their treatment plans could be far from standard protocols. Hence, patients with over 30-day ICU LOS are excluded from our study. Similarly, patients with more than one week of ventilation were much more complicated and should be handled on a case-by-case basis, depending on the physician's discretion. Thus, we do not consider these cases. We also find a few patients who have simultaneous records for intubation and extubation, which are likely incorrect and thus excluded. Cases with an NIV record during MV are excluded since their exact extubation times cannot be determined. Patients who died within 7 days after extubation are excluded since it is difficult to identify their extubation outcomes. We do not consider terminal extubations (i.e., patients who

die before extubation). According to the physicians, terminal extubation is quite rare and typically happens in extreme cases with severe comorbidities. Such cases should be treated with very special care, and the general extubation protocol may not be applicable. Hence, terminal extubation cases are usually removed in the medical literature of extubation analysis. In our data set, the number of terminal extubation cases is rather small, and other exclusion criteria have already excluded them. In our partner hospital, patients are not allowed to be discharged within 6 hours after extubation. There were a few violations in the data, and we excluded these cases because the cause of violations cannot be determined. We also remove cases without records of respiratory rate and tidal volume since these are necessary variables used in assessing patients' health conditions under MV. We will elaborate in detail later on how we assess and classify patient health conditions. Cases of unplanned extubation (extubation by mistake) are excluded from our analysis because they are not the result of regular extubation decisions. Finally, 3,067 cases are enrolled in our study. Table EC.1 illustrates our data selection process in details.

Table 1 provides summary statistics of our study sample at MV initiation. First, we have demographic information documented at ICU admission. The average age was 60.1 years old, and most of the patients were male (77.0%) and Chinese (65.7%). We extracted most of the relevant clinical variables for our study based on the literature and discussions with the physicians. Detailed descriptions of clinical variables are provided in Table EC.2 in Appendix EC.6. These variables were updated frequently from every 15 minutes to daily, which allows us to train a real-time predictive model for the patient's future condition. Finally, in our study sample, patients were ventilated for 11.7 hours on average. There were 481 (16.0%) cases of extubation failures. The mean ICU LOS was 97.6 hours, and the mortality rate was 0.9%.

4. Model Formulation and Calibration

In this section, we describe the ICU extubation problem (i.e., weaning patients from ventilation extubation) and formulate the problem as an MDP with perfect or imperfect predictive information (MDP-P). For comparison, we also provide a basic formulation that does not consider predictions on future information (label as MDP-B model) in Appendix EC.1. We derive structural results for the optimal policy and compare the optimal value functions in different settings. Then, we describe the procedure to estimate the parameters from our data.

The extubation problem is illustrated in Figure 1. Suppose a patient is initiated with mechanical ventilation. The decision-maker (DM) (e.g., a physician) monitors the patient's condition regularly and has to determine whether to extubate the patient at every checkpoint (decision epoch). Once the ventilator is removed, the decision process ends. Otherwise, the DM makes the extubation decision afterward.

Table 1 **Summary statistics of the study sample.**

Variable	Mean	SD	Variable	Count	Percentage
Demographics					
Age (year)	60.1	11.5	Gender: male	2,361	77.0
Weight (kg)	66.1	13.7	Race		
			Chinese	2,016	65.7
			Malay	468	15.3
			Indian	262	8.5
			Others	321	10.5
Clinical information					
HR (beats/min)	87.6	13.1	Cardiac rhythm		
RR (times/min)	14.7	4.1	Paced	1,083	35.3
GCS	3.8	2.7	Sinus rhythm	2,678	87.3
Creatinine (μmol/L)	108.8	110.7	Sinus tachycardia	345	11.2
Haematocrit (%)	34.6	6.7	Nursing assessment		
Haemoglobin (g/dL)	11.7	2.3	Cardiac	2,005	65.4
Potassium (mEq/L)	4.1	0.6	Respiratory	1,011	33.0
Sodium (mEq/L)	137.9	4.1	Vascular	1,789	58.3
Platelets ($\times 10^9$/L)	216.5	86.5	Musculoskeletal	948	30.9
WBC ($\times 10^9$/L)	10.4	4.9	Neurological	1,874	61.1
Mean arterial BP (mmHg)	85.8	17.5	Nutrition	1,812	59.1
Systolic BP (mmHg)	126.1	22.6			
Diastolic BP (mmHg)	64.0	14.9			
Temperature, °C	34.9	3.1			
Arterial pO2 (mmHg)	169.9	113.0			
Arterial pCO2 (mmHg)	41.6	7.4			
Arterial pH	7.4	0.1			
Arterial SaO2 (%)	92.5	10.8			
FiO2 (%)	75.2	11.4			
Tidal volume (L)	0.5	0.1			
Outcomes					
Ventilation duration (hr)	11.5	14.8	Extubation failure	481	16.0
ICU LOS (hr)	97.6	73.1	Mortality	22	0.9

SD: standard deviation; HR: heart rate; RR: respiratory rate; GCS: Glasgow Coma Scale; WBC: white blood cell; BP: blood pressure.

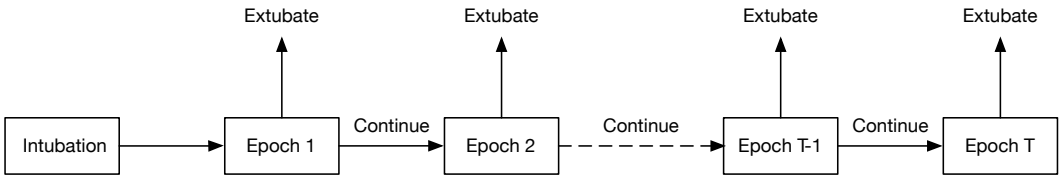


Figure 1 **An illustration of the extubation problem.**

4.1. Model Formulation

In this part, we formulate a mathematical model based on the theory of MDP to characterize the extubation problem with predictive information and derive the optimal policy, and denote this

model as MDP-P. We assume that the DM has access to forecast information that can be noisy. In what follows, we first discuss the basic concepts and settings of MDP-P.

Time horizon. We assume that extubation decisions are made within finite and discrete-time epochs indexed by $t \in \mathcal{T} = \{1, \ldots, T\}$, where T represents a predefined upper limit on the ventilation period. The extubation decision for patients who are ventilated for more than 1 week is rather complicated and substantially different from the routine decision process. In our study ICU, physicians aim to extubate patients within 1 week, except for those expected to be ventilated for a more extended period. In our data, only 67 (1.6%) patients were ventilated for more than 1 week. Hence, we set the maximum ventilation duration as 1 week (i.e., 168 hours) in this study. In our study ICU, medical staff will review the condition of a patient on a ventilator every 6 hours to decide whether to extubate the patient. Therefore, the decision interval is set to 6 hours in this study, and consequently, there are a total of 28 epochs—i.e., $T = 28$, with the last epoch a terminal epoch where all the patients will be extubated.

States. To capture the patient's breathing condition, we classify patients into one of C classes based on the observed clinical variables, with each class $c \in \mathcal{C} = \{1, 2, \ldots, C\}$ potentially corresponding to a particular respiratory condition. We use class 1 to represent the worst condition, and class C corresponds to the best condition. We use c_t to denote the patient's class in epoch t. Since predictive information is accessible, the DM can observe not only the current class c_t but also a prediction about the patient's class in next epoch c_{t+1} denoted as $\hat{c}_{t+1}$. Note that the current patient class c_t can be fully observed. Thus the state in epoch t is $(c_t, \hat{c}_{t+1})$ and the state space is $\mathcal{C} \times \mathcal{C}$.

For the extubation problem, clinicians have developed early extubation protocols in recent years based on the rapid shallow breathing index (RSBI) which is an established predictor for extubation failure (Fitch et al. 2014, Mahle et al. 2016, Yang and Tobin 1991). RSBI is defined as the ratio of respiratory rate to tidal volume, i.e., $\text{RSBI} = \text{RR}/V_T$, where RR is the respiratory rate (times/min) and V_T (L) is the tidal volume. A lower RSBI value indicates better breathing function. Ventilators in ICUs, which frequently record the respiratory rate and tidal volume of intubated patients, enable real-time monitoring of patients' respiratory condition. Yang and Tobin (1991) found that an RSBI of 105 is a good cut-off point for predicting the extubation outcome. Meade et al. (2001) suggested that patients with an RSBI less than 65 can be extubated. In another study, Chao and Scheinhorn (2007) assessed an extubation threshold of RSBI less than 80 and concluded that it is too conservative. The authors pointed out that a higher RSBI threshold could be adopted to allow more patients to benefit from successful early extubation. More recently, McConville and Kress (2012) proposed a set of early extubation strategies, and they reaffirmed the RSBI threshold of 105 proposed by Yang and Tobin (1991).

In this study, we follow the medical literature by adopting RSBI as the classification criterion for stratifying patients into different risk groups for extubation. Respiratory rate and tidal volume are documented by ventilators and are frequently updated every 15 minutes to 1 hour in our data set. We classify RSBI into five classes $\{[105,+\infty),[65,105),[45,65),[35,45),(0,35)\}$ according to the threshold values found in the medical literature (Meade et al. 2001, McConville and Kress 2012), and suggestions from the physicians in our study ICU. Thus, $C=5$ in our case. We also conduct sensitivity analysis with alternative and more granular patient classification schemes in Appendix EC.7.2. The findings and insights are consistent with results reported here in the base case.

Transitions. If the DM decides to continue in epoch t, i.e. $a_t=0$, the patient in class i will evolve to class j in epoch $t+1$ with probability $P(i,j)$, where $i,j\in\mathcal{C}$. Otherwise, if the treatment is terminated, i.e. $a_t=1$, the decision process terminates. We will use ventilation and treatment interchangeably in the rest of the paper. In the context of machine learning, the confusion matrix is commonly used to assess the performance of predictive models in binary classification problems. Hence, to capture the prediction state and prediction errors, we adopt a misclassification matrix, denoted as Q, which is an analog of the confusion matrix for the multi-class classification problem. Each component of the misclassification matrix $Q(i,j):=Pr(\hat{c}_t=j\mid c_t=i)$ represents the probability that a class i patient is predicted to be class j. The diagonal elements of Q are probabilities that the corresponding class of patients can be correctly predicted. These are essentially sensitivity and specificity for a binary classification problem. Q is the identity matrix I if the prediction is perfectly accurate. Based on the law of total probability, given the current class c_t, the probability that the predictive model will yield a prediction $\hat{c}_{t+1}$ on the patient class in epoch $t+1$ can be computed as $\tilde{Q}(c_t,\hat{c}_{t+1})=\sum_{c\in\mathcal{C}}P(c_t,c)Q(c,\hat{c}_{t+1})$. Following the Bayes' rule, given the current system state $(c_t,\hat{c}_{t+1})$, the probability that the true class in epoch $t+1$ is c_{t+1} can be written as

$$\pi\left(c_{t+1}\mid c_t,\hat{c}_{t+1}\right)=\frac{P(c_t,c_{t+1})Q(c_{t+1},\hat{c}_{t+1})}{\tilde{Q}(c_t,\hat{c}_{t+1})}. \tag{1}$$

Here π can be viewed as a belief vector of the DM on the future state given current class c_t and predicted class $\hat{c}_{t+1}$. Therefore, the transition probability from state $(c_t,\hat{c}_{t+1})$ to $(c_{t+1},\hat{c}_{t+2})$ can be written as

$$Pr\left(c_{t+1},\hat{c}_{t+2}\mid c_t,\hat{c}_{t+1}\right)=\pi\left(c_{t+1}\mid c_t,\hat{c}_{t+1}\right)\tilde{Q}(c_{t+1},\hat{c}_{t+2}). \tag{2}$$

Actions. In each decision epoch, the DM has to decide whether to extubate the patient. The action in epoch $t\in\mathcal{T}$ is $a_t\in\mathcal{A}=\{0,1\}$, where 0 indicates continuation and 1 indicates extubation. As mentioned before, we assume the treatment should be stopped despite the patient's health condition in the last epoch, i.e., $a_T=1$.

Costs. In each epoch, the corresponding cost depends on the patient's current class and the action taken by the DM, while the predicted class does not affect the immediate cost. In particular,

we set our objective as minimizing the expected ICU LOS, which is also one of the primary service outcomes for surgical care provided by hospitals (Wong et al. 2016). This objective is also aligned with the goals of fast track management, which requires physicians to extubate patients more aggressively to shorten patient stays in ICU (Sato et al. 2009, Fitch et al. 2014). We will show later, in the detailed definition of the cost parameters, that the key clinical outcome for the extubation decision—the extubation failure rate (EFR)—can easily be calibrated into the cost parameters to capture the trade-off between service and clinical outcomes. In our data set, there is no significant association between the remaining LOS after extubation and ventilation duration (correlation coefficient $= 0.01$, p-value $= 0.92$). Hence, we assume that the remaining LOS is independent of the length of the previous patient stay, which is consistent with the assumption in the literature that ICU LOS is geometrically distributed (Chan et al. 2012, Dai and Shi 2018a,b). Note that the extubation decision has no impact on ICU stays before intubation. Therefore, we only consider the ICU stay from intubation to be the objective in this study. In the rest of this paper, we use LOS to represent ICU LOS after MV initiation. Based on this setting, the immediate cost H incurred by continuing ventilation is the decision interval, which is equal to 6 hours. The terminal cost $G(c)$ is then the expected remaining LOS after extubation. We assume that $G(c)$ is decreasing[1] in c, which means the remaining LOS is shorter when a patient is extubated in a better condition. In Appendix EC.3, we show that our structural results hold when the cost to continue is multiplied by a discount factor.

Based on the above setups, the optimality equation of the MDP-P model can be written as

$$J_t^P(c_t,\hat{c}_{t+1}) = \min\left\{G(c_t),\ H(c_t) + \sum_{c_{t+1}\in\mathcal{C}}\sum_{\hat{c}_{t+2}\in\mathcal{C}} Pr\left(c_{t+1},\hat{c}_{t+2}\,\middle|\, c_t,\hat{c}_{t+1}\right) J_{t+1}^P(c_{t+1},\hat{c}_{t+2})\right\}, \quad \forall t\in\mathcal{T}\backslash\{T\}, c_t,\hat{c}_{t+1}\in\mathcal{C}, \tag{3}$$

where the transition probability $Pr(c_{t+1},\hat{c}_{t+2} \mid c_t,\hat{c}_{t+1})$ is computed as in Equation (2), and $J_t^P(c_t,\hat{c}_{t+1})$ represents the minimum expected total cost incurred by a class c_t patient in epoch t when his or her breathing condition in epoch $t+1$ is predicted as $\hat{c}_{t+1}$. The boundary value is $J_T^P(c_T,\hat{c}_{T+1}) = G(c_T)$, $\forall c_T\in\mathcal{C}$. Note that in epoch T, although we have a predicted class $\hat{c}_{T+1}$ beyond the time horizon T, it does not affect the value function since the process terminates in epoch T, and the terminal cost only depends on patient class in epoch T.

To derive the structural results for MDP-P, we require some assumptions. In what follows, we first introduce two related concepts. The first is the monotone likelihood ratio (MLR) dominance. A probability vector $\boldsymbol{x}$ is said to dominate $\boldsymbol{y}$ in MLR ordering if $\boldsymbol{x}(j)\boldsymbol{y}(i) \geq \boldsymbol{x}(i)\boldsymbol{y}(j)$, $i<j$, denoted as $\boldsymbol{x} \geq_{lr} \boldsymbol{y}$. MLR dominance is widely used in POMDPs since it is preserved under conditional

[1] In this paper, we use increasing (decreasing) in the weak sense, i.e., nondecreasing (nonincreasing).

expectation (e.g., Bayesian update). Next, we introduce the definition of totally positive of order 2 (TP2). A matrix P is TP2 if all of its second-order minors are nonnegative. Equivalently, P is TP2 if and only if $P_j \geq_{lr} P_i$ for $i < j$ (Krishnamurthy 2016). We are now ready to introduce the required assumptions for the MDP-P model.

ASSUMPTION 1. *Class transition matrix P is TP2.*

Assumption 1 ensures that a healthier patient will evolve to a better condition in expectation. For $C = 2$, Assumption 1 holds if $P(1,1) \geq P(2,1)$. For $C > 2$, if the transition matrix P has a tridiagonal structure (i.e., $P(i,j) = 0$, for $j < i-1$ or $j > i+1$), then Assumption 1 holds if $P(i,i)P(i+1,i+1) \geq P(i,i+1)P(i+1,i)$.

ASSUMPTION 2. *Misclassification matrix Q is TP2.*

Assumption 2 implies that a patient with a better health state is more likely to be predicted in a more favorable state in the next period than another patient who is in a worse health state.

ASSUMPTION 3. *$G(c) - \sum_{c' \in \mathcal{C}} \pi(c' \mid c, \hat{c}) G(c) - H$ is decreasing in c for any $\hat{c} \in \mathcal{C}$.*

Assumption 3 states that a healthier patient gets less benefit from one more period of ventilation regardless of the prediction for the future health state. Equipped with these three assumptions, we can show the properties of the optimal value function for the MDP-P model below, followed by the structure of the optimal policy. Proofs of all technical results are relegated to Appendix EC.2.

LEMMA 1. *Under Assumptions 1 and 2, the optimal value function for the MDP-P model $J_t^P(c_t, \hat{c}_{t+1})$ is nonincreasing in c_t and $\hat{c}_{t+1}$ for $t \in \mathcal{T}$.*

Lemma 1 first indicates that optimal treatment cost decreases with patient class. Meanwhile, if a patient is predicted to be in a better health state in epoch $t+1$, the optimal treatment cost is expected to be lower. We characterize the structure of the optimal policy for the MDP-P model in the following theorem.

THEOREM 1. *Under Assumptions 1–3, for the MDP-P model, there exists an optimal switching curve Γ_t in epoch $t \in \mathcal{T} \backslash \{T\}$ that partitions the state space into two subsets $\mathcal{R}_t^0$ and $\mathcal{R}_t^1$, such that the optimal action is*

$$a_t^*(c_t, \hat{c}_{t+1}) = \begin{cases} 0 & if\ (c_t, \hat{c}_{t+1}) \in \mathcal{R}_t^0, \\ 1 & if\ (c_t, \hat{c}_{t+1}) \in \mathcal{R}_t^1. \end{cases}$$

In addition, the optimal threshold policy has the following properties:

(1) $a_t^*(c_t, \hat{c}_{t+1})$ is nondecreasing in c_t but nonincreasing in $\hat{c}_{t+1}$.

(2) $\mathcal{R}_{t+1}^0 \subseteq \mathcal{R}_t^0$ and $\mathcal{R}_t^1 \subseteq \mathcal{R}_{t+1}^1$ for $t = 1, 2, \ldots, T-2$.

Theorem 1 confirms that the optimal policy for MDP-P is a switching curve policy. We can observe that the propensity to extubate the patient is increasing with the patient's current class and decreasing in the predicted class. As stated in Appendix EC.1, the optimal policy in MDP-B is a time-dependent threshold policy which only depends on the current class. With predictive information, patients who are extubated according to the optimal policy in MDP-B may continue to be ventilated if their condition can improve in the next epoch. Similarly, patients who keep ventilated according to the optimal policy in MDP-B may be extubated if the condition is predicted to deteriorate in the next epoch. This implies that the predictive information enables more precise decisions through personalized treatment effect prediction, which is confirmed as an important factor in other medical decision-making problems (Chan et al. 2012, Kim et al. 2014). Moreover, the extubation region $\mathcal{R}_t^1$ grows with time t (equivalently, the continuing region $\mathcal{R}_t^0$ shrinks with time t), which indicates that the stopping criterion should be relaxed with the increase in treatment duration t. Note that there is potential ethical issue arises from such guidelines, although most of which can be addressed by well-estimated terminal costs. We provide detailed discussion of related issues in Section 5.4.

Remark 1. We can also formulate the above optimal stopping problem with predictive information using the POMDP model. Under POMDP framework, the predictive information is modeled as an observation in each epoch, and the patient's condition in the next epoch is modeled as a partially observable state. The DM forms a belief about the patient's actual condition in the next epoch using the misclassification matrix, which is called information matrix in POMDP terminology, and makes the stopping decision based on the belief. If the decision is to continue the treatment, the patient's condition evolves into the next epoch, and a new observation—a prediction of the patient's condition in the next next epoch—is revealed by the predictive model and the process continues. We provide more details on POMDP formulation in Appendix EC.4 and show that all of the structural results regarding the optimal policies can be proved under the same set of assumptions. Note that in general, a POMDP model cannot be equivalently formulated as an MDP model. Our problem is a special case, since in our model the partially observed state—i.e., the patient's condition in the next epoch—is revealed in the next epoch. In contrast, in many other POMDP models, the partially observed state(s) may depend on the action and may not be fully observed within a fixed period.

4.2. The Benefit of Inaccurate Prediction

In this section, we theoretically explore the value of imperfect predictive information. We first compare the optimal cost of the MDP-B model with that of the MDP-P model. In practice, if the prediction accuracy is acceptable, the DM may use the predictive information directly as if

it were accurate. Hence, we further discuss a model in which the policy is derived using identity matrix I as the misclassification matrix, but the true misclassification matrix Q is unknown. In a special case with two patient classes, we derive sufficient and necessary conditions for the predictive information to be beneficial.

For a patient in class c_t who has been treated for the past $t-1$ periods, the expected optimal remaining treatment cost in the MDP-P model, denoted as $J_t^P(c_t)$, can be calculated as

$$J_t^P(c_t) = \sum_{\hat{c}_{t+1} \in \mathcal{C}} \tilde{Q}(c_t, \hat{c}_{t+1}) J_t^P(c_t, \hat{c}_{t+1}). \tag{4}$$

We have the following result:

THEOREM 2. *$J_t^P(c_t) \le J_t^B(c_t)$, $\forall t \in \mathcal{T}, c_t \in \mathcal{C}$, where $J_t^B(c_t)$, defined in Equation* (EC.1)*, represents the minimum expected total cost incurred by a class c_t patient at epoch t for the base model without prediction.*

Theorem 2 states that any predictive information (even inaccurate) is valuable. This result is based on the fact that the true misclassification matrix Q is known to the DM. In practice, however, the true misclassification matrix may not be easily obtainable. The misclassification matrix is usually estimated by out-of-sample tests from data. The accuracy of such estimation greatly hinges on the quality of the data, which may be limited. In real life, the DM could make decisions purely based on the prediction if the accuracy of the predictive model is acceptable. To capture this feature, we next analyze the cost of the MDP-P model under the optimal policy based on assuming I as the misclassification matrix in a special case with two patient classes.

Suppose there are two patient classes, where class 1 represents the worst condition, and class 2 is a better condition. Therefore, the transition matrix can be characterized by its diagonal elements, i.e., $p_1 := P(1,1)$ and $p_2 := P(2,2)$. We assume $0 < p_1 < 1$ and $0 < p_2 < 1$ to avoid trivial situations. Similarly, the misclassification matrix can also be characterized by its diagonal elements, i.e., $q_1 := Q(1,1)$ and $q_2 := Q(2,2)$, which are the sensitivity and specificity respectively in such a binary classification problem. Note that in this case, the optimal decision for class 2 patients is always extubation since there is no benefit from further ventilation. Hence, we only need to consider the cost of class 1 patients.

THEOREM 3. *If the DM applies an imperfect predictive model as if it were perfect when $C = 2$, the optimal cost is less than that without prediction (MDP-B) if and only if $\frac{H}{G(1)-G(2)}$ falls in the interval $\left[\frac{\gamma(1-p_1)(1-\lambda^{T-1})-(1-p_1)(1-\lambda)(p_1^{T-1}-\lambda^{T-1})}{(1-p_1^{T-1})(1-\lambda)-(1-p_1)(1-\gamma)(1-\lambda^{T-1})}, \frac{1-\gamma-\lambda}{1-\gamma}\right]$, where $\gamma = p_1 q_1 + (1-p_1)(1-q_2)$ and $\lambda = p_1(1-q_1)$. A necessary and sufficient condition for the above interval to be nonempty is $q_1 + q_2 \ge 1$. Moreover, the gap between the costs under the two policies is concavely increasing in sensitivity q_1 and linearly increasing in specificity q_2.*

Theorem 3 confirms the intuition that a predictive model with reasonably good sensitivity and specificity is worth using in practice, even if the noisy prediction is treated as perfect. We mathematically validate that the exact requirement is that the sum of sensitivity and specificity should be greater than or equal to 1. This condition on the prediction accuracy can be guaranteed if the predictive model's AUC—which measures the model's discriminative power—is higher than 0.5, which is a mild requirement. Theorem 3 says that for any reasonable predictive model, there exist nonempty intervals such that as long as the cost ratio $H/[G(1) - G(2)]$ falls in this interval, it is better to use predictive information even if its true accuracy is ignored and taken as perfect. Although predictive information might be detrimental when it is not properly deployed, the performance can be improved linearly by raising the specificity of the predictive model q_2 and concavely by increasing the sensitivity q_1.

Recall that H is the treatment cost, and $G(1) - G(2)$ represents the benefit of extubation at a better class. Intuitively, if the treatment cost is too high relative to the potential benefit of extubation at a better class, the optimal decision would always be extubation regardless of the prediction. On the other hand, if the treatment cost is very low, the predictive information is also useless because it is always beneficial to continue the treatment and extubate only at a better class. Furthermore, it can be shown that the upper limit of the interval is decreasing in p_1 but increasing in q_1 and q_2, while the lower limit is decreasing in q_2. In words, if patients are more likely to stay in class 1 under ventilation, the treatment cost must be low enough for the predictive information to be beneficial. With the increasing accuracy of the predictive model, a wider range of cost ratios can be tolerated in order to make use of predictive information.

4.3. Parameter Estimation

In this part, we estimate the unknown parameters in the proposed model. In particular, we first estimate terminal costs by a two-step approach. Transition probabilities are estimated by the sample mean. We then train a simple prediction model and estimate the misclassification matrix via out-of-sample test.

Terminal costs. The terminal cost $G(c)$ incurred by extubation is the expected remaining LOS after extubation (RLOS) conditioning on patient class, i.e., $G(c) := \mathbf{E}[\text{RLOS} \mid c]$. Typically, we expect $G(c)$ is positively correlated with the risk of extubation failure $r(c)$ since patients with failed extubation need further support to breathe. A failed attempt may even leave the patient in a worse condition, lengthening the remaining stays. In our data, patients who experienced extubation failure had significantly longer RLOS (144.8 hours vs. 66.5 hours, p-value $= 0.000$). This conjecture is confirmed by our estimation result. Hence, by including both the treatment cost and the terminal cost in the objective, we in fact take EFR into consideration when minimizing

the overall cost. One can also use a higher weight on the terminal cost so that the EFR becomes a more dominant consideration in the optimization. We conduct a sensitivity analysis of different weights on the terminal costs in the objective and report the results in Appendix EC.7.3. As shown later in the numerical results, the policies obtained from our models with predictive information not only minimize LOS but also reduce EFR significantly, even in the base setup without using higher weights on terminal costs.

When estimating the terminal cost, the major challenge is to deal with selection bias. Within each patient class, extubated patients were likely to be healthier than those who were not extubated, as the extubation decisions were made by physicians. Therefore, the sample average approach would lead to a biased estimator for the terminal cost. To address this concern, we conduct a two-step procedure leveraging our data to mitigate the estimation bias. We discuss the details of the estimation procedure below.

Leveraging on other information in our data set, we can rewrite the expression for $G(c)$ as

$$G(c) = \mathbf{E}\left[\mathbf{E}\left[\mathrm{RLOS} \mid c, \boldsymbol{x}\right] \mid c\right], \tag{5}$$

where $\boldsymbol{x}$ is a vector of patient characteristics and clinical information immediately before extubation, as shown in Table 1, except for patient class. We also include the sequential organ failure assessment (SOFA) score, which is a well-established severity score used in ICUs (Vincent et al. 1996), to control the underlying health condition. Our first step is to estimate $\mathbf{E}[\mathrm{RLOS} \mid c, \boldsymbol{x}]$, and the second step is to obtain the distribution of patient characteristics $\boldsymbol{x}$ within each patient class using a propensity score approach to recover $\mathbf{E}[\mathrm{RLOS} \mid c]$. The first target fits into linear regression naturally. Specifically, the fitted value from the regression model

$$\mathrm{RLOS} = \beta_0 + \boldsymbol{\beta}_1' \boldsymbol{x} + \boldsymbol{\beta}_2' \boldsymbol{c} + u \tag{6}$$

is an estimator of $\mathbf{E}[\mathrm{RLOS} \mid c, \boldsymbol{x}]$, where (with a little abuse of notation) $\boldsymbol{c}$ denotes the patient class dummy; β_0, $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$ are corresponding coefficients; and u is the random error term. Our approach is consistent with many empirical studies that use untransformed LOS as the dependent variable (e.g., Moran et al. 2008, Niskanen et al. 2009, KC 2019). Verburg et al. (2014) showed that linear regression with untransformed LOS truncated at 30 days gives the highest R-square compared to other forms of LOS and regression models. In the literature, it is also a common practice to use the logarithmically transformed LOS as the dependent variable (KC and Terwiesch 2012, Hu et al. 2018), in which the goal is usually to identify the risk factors for longer LOS. In our setting, however, we cannot apply the logarithmic transformation since we aim to estimate the expectation of RLOS, and expectation does not hold under the logarithmic transformation. Nevertheless, we

also estimate the model with logarithmically transformed RLOS and use the exponential function to transform the fitted log(RLOS) back to the scale of RLOS. This approach leads to a higher root mean squared error (RMSE) compared with Equation (6) (61.9 vs. 60.0).

In the second step, we need to estimate $G(c)$ based on Equation (5). The challenge here is to obtain the distribution of the covariates within each patient class. As discussed before, the conditions in the terminal epoch cannot represent the distribution of those variables in each class due to selection bias. We adopt a stratification-based propensity weighting approach (Rosenbaum and Rubin 1983) to mitigate bias. The main idea is to use the distribution of propensity to treat—in this case, to be extubated—as an estimator for the population distribution of the covariates. This is possible since our data contains all of the real-time tracking information on patients from ICU admission. Following this approach, we first estimate the propensity score for each patient in each epoch via a logistic regression, which has an AUC of 0.762. Next, within each class c, sample instances are stratified into K_c subgroups of equal size according to the propensity scores. Since there are fewer data points in the worst class (i.e., class 1), we choose five strata for class 1 and ten for the other classes. Cochran (1968) showed that five groups were sufficient to remove 90% of the bias incurred by imbalanced covariates. After that, we compute the average $\widehat{RLOS}$ estimated from Equation (6) for extubated patients in each subgroup k and class c, denoted as $\overline{\widehat{RLOS}}_{ck}$. Finally, we use the weighted average of the means for all subgroups within the patient class as an estimator for the terminal cost:

$$\hat{G}(c) = \sum_{k=1}^{K_c} \overline{\widehat{RLOS}}_{ck} \frac{N_c^k}{N_c},$$

where N_c^k is the number of samples in class c subgroup k; N_c is the total number of sample points in class c; and $\hat{G}(c)$ denotes estimated terminal cost.

We also estimate the EFR for each patient class following a similar two-step procedure but with a linear probability model (LPM) in the first step. We choose LPM rather than logistic regression for the same reason as using linear regression to estimate $G(c)$. In our data, the LPM produces an almost identical AUC compared with the logistic regression (0.775 vs. 0.776). Note that the estimated EFR is not used in the model to optimize the extubation protocol but rather in the numerical analysis to evaluate the performance of different extubation policies.

Estimation results are shown in Table 2. As expected, both the risk of extubation failure $r(c)$ and RLOS(c) are decreasing in patient class c. More detailed information on the regression models used in this part can be found in Appendix EC.8.

Class transition matrix. To estimate transition probabilities, we first separate the ventilation period into epochs at an interval of 6 hours for each patient. We next identify patient classes in each epoch using the corresponding RSBI values and the aforementioned classification criterion.

Table 2 Estimated RLOS and EFR for each patient class.

Terminal class	No. of observations	RLOS (hr)	EFR (%)
1	29	137.8	40.9
2	150	98.2	29.6
3	419	91.1	21.0
4	688	88.6	17.7
5	1,781	79.0	13.8

RLOS: remaining length of stay; EFR: extubation failure rate.

If there are multiple RSBI records within one epoch for a patient, we take the averaged value to characterize the patient's condition in that epoch. Then the transition probability $P(i,j)$ from class i to j can be estimated as $P(i,j)=n_{ij}/n_i$, where n_i denotes the number of samples points in class i and n_{ij} denotes the number of sample points in class i that transit to j in the next epoch. One-step transition probabilities for each patient class can be estimated from the data as follows:

$$\hat{P}=\begin{pmatrix} 0.35 & 0.21 & 0.20 & 0.11 & 0.13 \\ 0.09 & 0.40 & 0.28 & 0.13 & 0.10 \\ 0.02 & 0.14 & 0.48 & 0.24 & 0.12 \\ 0.01 & 0.05 & 0.23 & 0.36 & 0.35 \\ 0.01 & 0.02 & 0.07 & 0.18 & 0.72 \end{pmatrix}.$$

We notice that the selection bias issue can also affect the estimation of the transition matrix. Thus, we further apply the propensity score weighting technique to adjust the bias in our sensitivity analysis, details of which can be found in Appendix EC.7.1. The main results are consistent with those we derived using the above approach.

Misclassification matrix. In this part, we train a practical predictive model using the random forest algorithm in our data set and investigate the benefit of incorporating the prediction produced by this model. Time-series data in each epoch of each patient are used as sample points, and the patient class in the next period is the target. We have 4,376 samples, of which 75% are used for training, and the other 25% are used for validation and estimation of the misclassification matrix Q. Note that the focus of this paper is not to train a model with the strongest predictive power. Instead, our objective is to demonstrate how a reasonable predictive model can affect decision quality if its prediction is embedded in the extubation protocol. We use all of the variables in Table 1 in our predictive model to predict patients' RSBI class in the coming period. We also categorize continuous variables according to their normal ranges and add them as additional features. For example, heart rate (HR) is categorized as $HR<60$, $60\le HR\le 100$, and $HR>100$. The out-of-sample misclassification matrix $\hat{Q}$ of the predictive model is given below:

$$\hat{Q}=\begin{pmatrix} 0.60 & 0.08 & 0.17 & 0.15 & 0.00 \\ 0.10 & 0.60 & 0.16 & 0.09 & 0.05 \\ 0.01 & 0.09 & 0.61 & 0.21 & 0.08 \\ 0.00 & 0.02 & 0.31 & 0.53 & 0.14 \\ 0.00 & 0.01 & 0.11 & 0.28 & 0.60 \end{pmatrix}.$$

Note that the estimated matrices $\hat{P}$ and $\hat{Q}$ are not TP2. But we will show later that the optimal policy is almost a switching-curve policy. Furthermore, we will demonstrate how to leverage on Theorem 1 to derive near-optimal policies that are more interpretable and easier to implement. The idea is to enforce Assumptions 1 and 2 in parameter estimation using constrained maximum likelihood estimation (MLE). More details are presented in Appendix EC.9.

5. Numerical Results

In this part, we use the parameters estimated above to derive optimal policies for extubation in different settings. We first derive the optimal policies for extubation without prediction and with perfect prediction, then compare the objectives (i.e., LOS) and the EFR for each patient class under different policies. Our results validate the value of predictive information on the next decision epoch. We also show that if more future information can be predicted, the performance improvement is more significant. Next, we use our clinical data and machine learning tools to train a practical predictive model. Even though the prediction accuracy is moderate, we can still observe notable performance improvement compared with the case without any prediction. Using the derived result, we estimate the impact of applying our model to critical COVID-19 patients. Finally, to further study the impact of prediction accuracy on system performance, we conduct a range of numerical comparisons with different prediction errors.

5.1. Perfect Prediction and the Value of Future Information

In this part, we first derive the benefit of perfect prediction (i.e., $Q = I$), which provides an upper bound for the value of predictive information. Given the estimated parameters, we can numerically derive the optimal extubation policy without prediction (label as OP-B; Figure 2), as well as the optimal extubation policy with perfect predictions (label as OP-PP; Figure 3) by setting $Q = I$. The results show that for the base model without prediction, it is optimal to extubate the patient whenever he or she evolves to class 2 or a better class—equivalent to a cut-off value of 105 in RSBI—which is consistent with the recommendation of Yang and Tobin (1991) and McConville and Kress (2012). However, when we know that a class 2 patient will evolve to a better class if ventilation is continued for one more period, it will be optimal to continue ventilation for this patient instead of extubation, as shown in Figure 3, and similarly for a class 3 patient. We can observe from Figure 3 that the OP-PP policy in each decision epoch depends on both current class information and future class information. The black dashed lines in Figure 3 circle the subgroup of patients whose recommendations are based on extubation decisions under the OP-PP policy that differs from those under the OP-B policy. Note that the optimal decision in the last epoch is more aggressive than other epochs, which is due to the terminal effect. The policy can be easily adjusted,

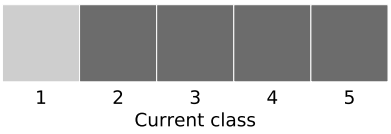
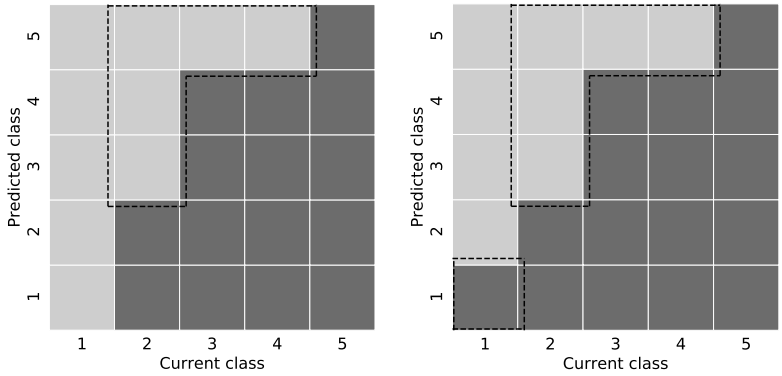


Figure 2 **Optimal extubation policy without predictive information. The dark gray grid corresponds to extubate and the light gray grid to continue ventilation. The policy is stationary across time epochs.**



(a) $t = 1, 2, \ldots, 26$. (b) $t = 27$.

Figure 3 **Optimal extubation policy with perfect predictive information. The dark gray grid corresponds to extubate and the light gray grid to continue ventilation.**

and discretionary action can be taken by physicians when a patient is still intubated near the end of the decision horizon.

Under the OP-B policy, the expected total cost $J_1^B(c)$ can be computed for each patient with initial class c using Equation (EC.1). Moreover, the EFR for a patient with initial class c under the OP-B policy, denoted as $R_1^B(c)$, can be computed using the following recursive formula:

$$R_t^B(c_t) = a_t^*(c_t) r(c_t) + [1 - a_t^*(c_t)] \sum_{c_{t+1} \in \mathcal{C}} P(c_t, c_{t+1}) R_{t+1}^B(c_{t+1}),\ t \in \mathcal{T} \backslash \{T\},$$

where $a_t^*(\cdot)$ follows the OP-B policy and $r(\cdot)$ is the estimated EFR given in Table 2. The boundary condition is $R_T^B(c) = r(c)$ for all c. The expected total cost under the optimal policy with predictive information (for any general misclassification matrix) can be obtained from Equations (3) and (4). The corresponding EFR $R_t^P(c_t)$ for a patient with class c_t at epoch t can be calculated as follows:

$$R_t^P(c_t) = \sum_{\hat{c}_{t+1} \in \mathcal{C}} \tilde{Q}(c_t, \hat{c}_{t+1}) R_t^P(c_t, \hat{c}_{t+1}),$$

where $R_t^P(c_t, \hat{c}_{t+1}) = [1 - a_t^*(c_t, \hat{c}_{t+1})] P(c_{t+1}, \hat{c}_{t+2} \mid c_t, \hat{c}_{t+1}) R_{t+1}^P(c_{t+1}, \hat{c}_{t+2}) + a_t^*(c_t, \hat{c}_{t+1}) r(c_t)$, and $a_t^*(\cdot, \cdot)$ follows the optimal policy with predictive information. For comparison, we also estimate the current cost and failure rate for each initial class in our data set. These results are summarized in Table 3.

Table 3 Performance of different extubation policies for each initial patient class.

Class	Data		OP-B		OP-PP	
	LOS (hr)	EFR (%)	LOS (hr)	EFR (%)	LOS (hr)	EFR (%)
1	115.3	21.4	105.8	21.8	104.5	19.4
2	103.9	29.1	104.2	29.6	101.7	23.6
3	99.5	20.2	97.1	21.0	96.4	20.1
4	93.9	16.7	94.6	17.7	93.3	16.3
5	86.7	13.9	85.0	13.8	85.0	13.8

OP-B: optimal extubation policy without predictive information; OP-PP: optimal extubation policy with perfect prediction (i.e., $Q = I$); LOS: length of stay; EFR: extubation failure rate.

From Table 3, we can observe that OP-B can reduce the LOS for ventilated patients compared with current practice except for class 2 and 4 patients. This confirms our earlier concern that when physicians made extubation decisions in practice, they might have taken additional information into account, which could partially indicate a patient's future health condition. Moreover, such improvement on LOS is sometimes obtained at a price of increased EFR, which is somewhat expected from the optimization perspective. For example, the class 1 patient's EFR increases under the OP-B policy compared with the data. Interestingly, when predictive information is taken into account, we can see that not only is LOS further decreased from the OP-B policy, but also that EFR is reduced compared with the data, especially for class 2 patients. This is largely due to the situation in which a subgroup of patients who evolve to better conditions can be identified and then ventilated longer until they are in better conditions. Thus, EFR decreases significantly in this group of patients. Simultaneously, patients who may suffer from prolonged intubation can be detected by accurate prediction, and early extubation can be carried out for this subgroup of patients. This demonstrates the essence of precision medicine and personalized medical decision-making.

Note that patients with initial class 5 do not benefit from predictive information, which is reasonable. The decision for those patients is rather straightforward, as the majority can be extubated smoothly after a short period of ventilation and typically have a low risk of extubation failure. This is consistent with physicians' experience in our partner hospital—i.e., patients who started with healthier breathing conditions were typically ventilated due to the requirements of surgery and anesthesia, so most of them could be safely extubated soon after admission to the ICU. We will omit the result for class 5 patients in the subsequent analysis.

To put the savings in ICU LOS into perspective, consider a hospital with 3,000 ICU admissions requiring MV support per year (our study ICU is just one of five specialty ICUs in the hospital). If we manage to save 2 hours per patient stay on average for these patients, we can save 250 days in a year, which is equivalent to an ICU bed if utilization is around 70%. If the hospital maintains the number of beds, the reduction in LOS can reduce the congestion level. Given that a high congestion

level can result in worse outcomes (KC and Terwiesch 2009, Kuntz et al. 2014), the reduction in LOS would also improve the quality of care. In terms of operations, the reduction in LOS can shorten the waiting time for subsequent patients. Since delays can propagate in a queuing system (Luo et al.), a 2-hour reduction can lead to a more significant improvement in the throughput for both the ICU and the entire hospital (Kim et al. 2020). From another perspective, ICU stay involves numerous treatments and intervention decisions, and although the extubation decision is one of the most critical decisions, its impact on the entire ICU LOS could still be limited. In our data, patients on average stay in the ICU for 66.5 hours after successful extubation. If we subtract this interval from the LOS, the reduction due to better extubation decisions would be much more substantial.

Intuitively, if a predictive model can predict patient conditions beyond the 6-hour limit, LOS could be further reduced. We next analyze the effect of extending the prediction window from one step to two or three steps—i.e., from 6 hours to 12 or 18 hours. Table 4 shows that LOS and EFR can be further reduced when more future information is predictable.

Table 4 Performance of the extubation policy with perfect prediction for each initial patient class when the prediction window is longer.

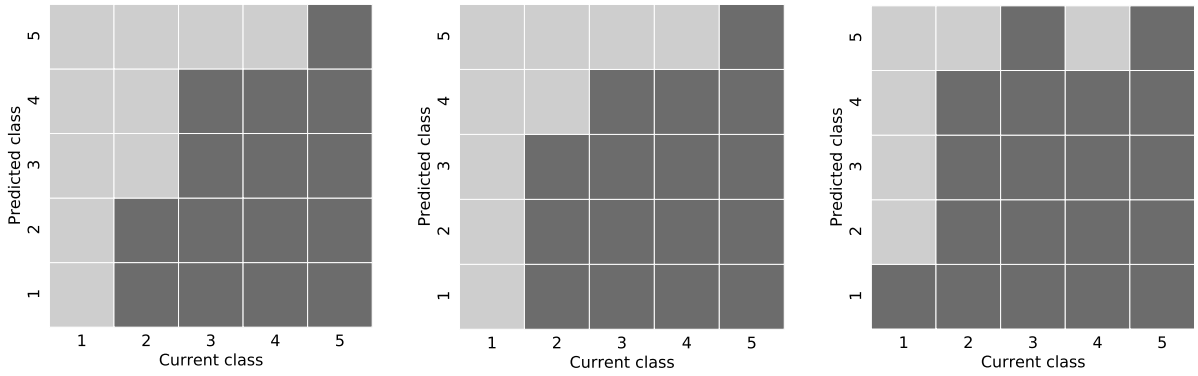
Class	12-hour window		18-hour window	
	LOS (hr)	EFR (%)	LOS (hr)	EFR (%)
1	104.4	19.1	104.4	19.1
2	101.4	23.1	101.3	22.6
3	96.4	19.7	96.4	17.7
4	93.3	16.3	93.3	16.3

LOS: length of stay; EFR: extubation failure rate.

5.2. A Practical Predictive Model and Imperfect Prediction

In the above analysis, we compare the model with perfect prediction to MDP-B, which provides an upper bound on the value of predictive information. However, prediction can hardly be 100% accurate in reality. With the estimated misclassification matrix $\hat{Q}$ in Section 4.3, we derive the optimal policy for the model with an imperfect prediction, as shown in Figure 4. We observe that the optimal policy is not a threshold policy in decision epoch 27. In particular, in decision epoch 27, a class 3 patient predicted to be class 5 (i.e., the best condition) in the next epoch (which is the last epoch) will be extubated despite the predicted improvement in health condition. This is because the accuracy of a class 3 patient predicted to be class 5 in the next epoch is much smaller compared with the accuracy of a class 4 patient predicted to be class 5 in the next epoch (0.48 vs. 0.75), which leads to the observed "jump" in the optimal policy that does not "believe"

in the prediction that a class 3 patient is going to be class 5 in the next epoch and recommends extubation. Note that this does not violate our analytical results because neither the estimated misclassification matrix nor the transition probability matrix satisfies the TP2 condition in our data set (i.e., Assumptions 1 and 2). Consequently, a threshold policy is not guaranteed. We will address this issue in detail in Section EC.9 and propose a new approach to obtain a practically more implementable policy leveraging on the analytical results derived in Section 4.1.



(a) $t = 1, 2, \ldots, 24$. (b) $t = 25, 26$. (c) $t = 27$.

Figure 4 Optimal extubation policy with imperfect predictive information. The dark gray grid corresponds to extubate and the light gray grid to continue ventilation.

Table 5 shows the performance of the optimal extubation policy with imperfect predictions (label as OP-IP). We can observe that the benefit of predictive information on reducing LOS decreases when the prediction is not perfect. However, it can still reduce LOS and EFR when compared with the model without predictive information. Next, we analyze the setting in Section 4.2, in which the DM is not aware of the misclassification matrix $\hat{Q}$ and simply apply the optimal policy assuming the prediction is perfect (i.e., the OP-PP policy). The performance is then evaluated in the true environment that the prediction is not imperfect and follows the misclassification matrix $\hat{Q}$. Although the OP-PP policy is not optimal in this case, we find that its performance is extremely closed to the optimal one under the OP-IP policy (no difference in our reporting accuracy). This is because our prediction model is fairly accurate and the loss due to ignoring its true misclassification matrix is negligible, which confirms our results in Theorem 3,

5.3. Sensitivity Analysis of Prediction Accuracy

In different settings, prediction accuracy depends on the availability of data as well as the power of a predictive model. In this part, we conduct sensitivity analysis to show how prediction accuracy will affect the outcomes. We fix all of the diagonal elements of the misclassification matrix $Q(i, i)$ to values from 0.1 to 0.9 with a step size of 0.1, and randomly generate the remaining entries

Table 5 Performance of the optimal extubation policy with imperfect prediction.

Class	OP-IP		Ignoring $\hat{Q}$ and Using OP-PP	
	LOS (hr)	EFR (%)	LOS (hr)	EFR (%)
1	105.3	19.3	105.3	19.3
2	102.9	23.0	102.9	23.0
3	97.1	20.3	97.1	20.3
4	94.3	16.9	94.3	16.9

OP-IP: optimal extubation policy with imperfect prediction; OP-PP: optimal extubation policy with perfect prediction (i.e., $Q = I$); LOS: length of stay; EFR: extubation failure rate.

for the misclassification matrix Q. Recall that the i-th diagonal element of Q is the probability that a patient's future state can be correctly predicted as class i if it is class i. Thus, these diagonal elements $Q(i,i)$ can be viewed as a precision parameter of the predictive model. Under each precision level, we generate 1,000 random misclassification matrices (random off-diagonal elements) and run our model to compute the average LOS and EFR. The results are summarized in Table 6.

Table 6 Extubation performance under the optimal extubation policy with imperfect prediction given different prediction precisions.

Precision	Initial class 1		Initial class 2		Initial class 3		Initial class 4	
	LOS (hr)	EFR (%)	LOS (hr)	EFR (%)	LOS (hr)	EFR (%)	LOS (hr)	EFR (%)
OP-B	105.8	21.8	104.2	29.6	97.1	21.0	94.6	17.7
0.0	105.6	20.8	103.6	26.5	97.1	21.0	94.6	17.6
0.1	105.7	20.9	103.9	27.0	97.1	21.0	94.6	17.6
0.2	105.7	21.1	104.0	27.6	97.1	21.0	94.6	17.7
0.3	105.7	21.1	103.9	27.4	97.1	21.0	94.6	17.7
0.4	105.7	20.8	103.8	26.7	97.1	21.0	94.6	17.6
0.5	105.6	20.5	103.6	25.9	97.1	21.0	94.5	17.4
0.6	105.4	20.1	103.3	24.9	97.1	20.8	94.3	17.0
0.7	105.2	19.7	102.9	24.1	97.0	20.5	94.1	16.8
0.8	105.0	19.5	102.5	23.8	96.8	20.3	93.8	16.6
0.9	104.8	19.4	102.1	23.7	96.6	20.2	93.6	16.5
1.0	104.6	19.3	101.7	23.6	96.3	20.0	93.4	16.3

LOS: length of stay; EFR: extubation failure rate.

From Table 6, we can observe that the model with imperfect predictive information outperforms the model without prediction even if the prediction precision is moderate. This is a bit surprising since a predictive model with low accuracy is likely to produce frequent wrong predictions that yield more noise than value for decision-making (Kappen et al. 2018). It appears that the misclassification matrix in our model allows a rational DM to speculate about the probability of each class that a patient's condition could evolve to based on both predictive information and the original transition

matrix, and the Bayesian update allows the value of even noisy prediction to be extracted to improve decision quality.

Nevertheless, it is possible that the DM does not know the misclassification matrix, or it is only used as an indicator of whether to believe the predictions. Therefore, the DM may make decisions purely based on the predictions as if they were correct. For comparison, we run another set of numerical tests in the setting we discussed in Section 4.2. The numerical results are summarized in Table 7. It can be observed that when the DM believes the prediction despite its error, the model without predictive information could perform better when the prediction accuracy is low. However, the situation is reversed under a moderate prediction accuracy—i.e., the optimal policy with the identical misclassification matrix can outperform the OP-B policy in the imperfect prediction environment. Note that for class 3 patients, the requirement for accuracy is relatively higher compared with other classes, i.e., the accuracy should be greater than 0.7.

Table 7 Extubation performance under the optimal extubation policy with perfect prediction (OP-PP).

Precision	Initial class 1		Initial class 2		Initial class 3		Initial class 4	
	LOS (hr)	EFR (%)	LOS (hr)	EFR (%)	LOS (hr)	EFR (%)	LOS (hr)	EFR (%)
OP-B	105.8	21.8	104.2	29.6	97.1	21.0	94.6	17.7
0.0	107.5	20.2	106.8	24.7	99.1	20.9	96.2	18.0
0.1	107.1	20.1	106.1	24.6	98.7	20.8	95.8	17.8
0.2	106.7	20.0	105.5	24.4	98.4	20.7	95.5	17.6
0.3	106.4	19.9	104.9	24.3	98.1	20.6	95.2	17.5
0.4	106.1	19.8	104.3	24.2	97.8	20.5	94.9	17.3
0.5	105.8	19.7	103.8	24.1	97.5	20.5	94.6	17.1
0.6	105.5	19.7	103.4	24.0	97.2	20.4	94.3	16.9
0.7	105.2	19.6	102.9	23.9	97.0	20.3	94.1	16.8
0.8	105.0	19.5	102.5	23.8	96.8	20.3	93.8	16.6
0.9	104.8	19.4	102.1	23.7	96.6	20.2	93.6	16.5
1.0	104.6	19.3	101.7	23.6	96.3	20.0	93.4	16.3

LOS: length of stay; EFR: extubation failure rate.

5.4. Ethical Consideration

Ethical consideration is very important in medical settings. The optimal decision in Theorem 1 suggests to extubate a patient if the patient's condition is predicted to deteriorate. There will be a concern if a patient who is predicted to deteriorate under treatment deteriorates more without the treatment. This is an ethical problem but does not alter the model analysis much as long as the terminal costs are well estimated. One can further extend the terminal costs to depend on predicted class in the next epoch to incorporate this consideration.

Nevertheless, to investigate such an issue in our data, we conduct the following test. We apply the random forest model developed in Section 4.3 to make predictions for all the patients at all

the decision epochs. We first select patients who are predicted to deteriorate and were actually extubated. We then use propensity score matching to obtain a subsample of patients predicted to deteriorate but were not extubated. For both groups of patients, we calculate their SOFA scores in the next epoch and perform a one-tailed t test with the hypothesis that the SOFA scores of the extubated patients were smaller. Recall that higher SOFA scores indicate worse health conditions. Only patients in classes 4 and 5 permit such analysis due to the lack of observations for patients in classes 2 and 3. The results in Table 8 show that for patients who are predicted to deteriorate, the clinical conditions of those who were extubated were on average much better than that of intubated patients, and the difference was statistically significant.

Table 8 SOFA scores of extubated and intubated patients whose conditions are predicted to deteriorate.

Patient class	Extubated patients	Intubated patients	p-value
4	6.4 ($\pm$2.9)	11.0 ($\pm$3.4)	< 0.001
5	6.8 ($\pm$3.4)	9.5 ($\pm$5.6)	0.046

Based on the literature and discussion with physicians, this is likely because the deterioration of breathing conditions mostly results from prolonged or unnecessary ventilation. Unlike other medical interventions, mechanical ventilation itself does not cure any disease but only provides support on breathing while the patient's breathing function is under-recovery (through other treatments). Prolonged ventilation can cause ventilator-associated complications such as pneumothorax, airway injury, alveolar damage, and ventilator-associated pneumonia (Hess 2011). Therefore, stopping the ventilation if the patient is predicted to deteriorate is not as controversial as other medical interventions, at least in our clinical context.

Furthermore, given the scarcity of ICU resources, such a decision is in line with an important ethical goal of public health that prioritizes patients most likely to benefit (White and Angus 2020), because freed up resources can be used to save lives. For instance, in the COVID-19 pandemic, ventilators are in significantly short supply. As a result, ventilators must be rationed among critical patients with care, and guidelines have been developed to extubate patients who are not likely to improve under further intubation (Truog et al. 2020, White and Lo 2020).

5.5. Application in COVID-19 Critical Care

Under the current COVID-19 pandemic, there is a surging demand for ventilators (Ranney et al. 2020, Truog et al. 2020). Consequently, the extubation decision is a critical problem during the crisis, and the global healthcare system is under enormous pressure to develop frameworks for rationing ventilators (White and Lo 2020). There is even intensive debate regarding how to protect

healthcare workers from potential future legal liability for making such critical decisions (Cohen et al. 2020). We believe that the framework we develop in this paper, together with advanced machine learning predictive models, can be applied in such a scenario to improve the utilization of ventilators by supporting physicians' extubation decisions and thus making a meaningful impact.

In this subsection, we use our earlier results to estimate potential savings by implementing our model during the COVID-19 pandemic. The estimation procedure is outlined as follows:

1. Estimate the number of patients who need critical care services by $N_c = cN$, where c is the proportion of COVID-19 patients admitted to ICUs and N denotes the total number of COVID-19 patients.
2. Estimate the total demand for mechanical ventilators as $N_v = vN_c$, where v is the proportion of ICU patients who need mechanical ventilation.
3. Estimate the total savings in LOS by $S_{LOS} = N_v L\rho$, where L is the average ICU LOS of infected patients and ρ is the percentage improvement in LOS by implementing our model. Since admitted critical COVID-19 patients typically suffer from poor respiratory conditions, we assume they are initially in class 1, as measured by RSBI (Xu et al. 2020). In our data, the average LOS for class 1 patients is 120.4 hours, and thus a 10.8-hour reduction in ICU LOS corresponds to a 9.0% improvement. We note that these numbers are estimated from our single-center data with a special group of patients. Thus the estimates may be biased, and better estimates can be derived when more data on COVID-19 patients are available.
4. Estimate additional ICU admissions by dividing the total saving in LOS by the average LOS of critical COVID-19 patients.

The total number of confirmed cases is drawn from the John Hopkins Coronavirus Resource Center[2] on June 29, 2020. In the U.S., 4.9% of COVID-19 patients needed critical care (CDC COVID-19 Response Team 2020). Bhatraju et al. (2020) reported that 75% of COVID-19 patients who were admitted to ICU needed mechanical ventilation, and the average ICU LOS was nine days in the U.S. In China, a multi-center study showed that 5.0% of COVID-19 patients were admitted to the ICU, and 45% required mechanical ventilation (Guan et al. 2020). In China, the average ICU LOS was 8 days (Zhou et al. 2020). In the U.K., a report from the Intensive Care National Audit & Research Centre (ICNARC)[3] showed that a total of 12,881 COVID-19 patients were admitted to ICU by June 26, 2020, and 60% of whom received mechanical ventilation. The total number of cases in the U.K. on June 26, 2020 was 310,836; therefore the ICU demand ratio is 4.1%. The average ICU LOS was 12 days for survivors and 9 days for non-survivors, and the

[2] https://coronavirus.jhu.edu

[3] https://www.icnarc.org/DataServices/Attachments/Download/0a0738a6-dcb7-ea11-9127-00505601089b

percentage of survivors was 59.1%. Therefore, we estimate the average ICU LOS as 11 days in the U.K. Parameters in the U.K. and China are also used to estimate potential savings in Europe and Asia, respectively. Parameters for the three countries are summarized in Table 9, and the estimated results are presented in Table 10.

Table 9 Parameters for COVID-19 patients.

Country	Percentage of critical care	Percentage of ventilator demand	Average ICU LOS
U.S.	4.9%	75%	9 days
China	5.0%	45%	8 days
U.K.	4.1%	60%	11 days

ICU: intensive care unit; LOS: length of stay.

Table 10 Potential savings during COVID-19 pandemic from deploying the proposed model.

Country	Total cases	ICU admissions	Demands for ventilator	Saved LOS (days)	Additional admissions
US.	2,548,996	124,901	93,676	75,625	8,403
China	84,757	4,238	1,907	1,369	171
UK.	312,640	12,818	7,691	3,449	314
Europe	2,415,328	99,028	59,417	26,649	2,423
Asia	2,215,078	110,754	49,839	35,765	4,471

ICU: intensive care unit; LOS: length of stay.

From Table 10, we can observe that the proposed model can potentially improve ICU throughput and admit more patients who need ICU services. For instance, the U.S. could admit 8,403 more critical COVID-19 patients, and Europe could admit 2,423 more. It is also worth mentioning that our proposed model can reduce the rate of extubation failure, which is positively associated with the mortality rate. By leveraging predictive information in ICU decision-making, we can significantly improve system efficiency and care quality.

6. Conclusion

In this paper, we propose a framework that exploits predictive information in a sequential decision-making environment for a medical treatment stopping problem. Following the literature, the conventional decision problem without predictive information is modeled using MDP as a benchmark. We extend the state space of the benchmark MDP model to incorporate a prediction of the future state, and system dynamics are extended to capture the quality of prediction. For the new MDP model, we derive the structural properties of the optimal policies and demonstrate the impact of predictive information. We demonstrate that the optimal policies are threshold policies, which can then be easily incorporated into existing medical decision protocols. In addition, we propose an

innovative parameter estimation procedure that leverages the analytical results to obtain near-optimal threshold policies that are more interpretable and easy to implement in practice.

To validate the proposed framework, we apply our models to the extubation problem in ICUs and calibrate the models using a comprehensive data set from an ICU. The numerical results show that predictive information allows for more precise identification of potential patients for early extubation, which in turn reduces the side effects of traditional aggressive extubation strategies that apply the same criteria to every patient based on their current condition. Consequently, new protocols with predictive information could reduce both ICU LOS and EFR, especially for those patients whose initial condition is poor. In contrast, optimal policies from the benchmark model without using predictive information can even lead to worse outcomes compared with the observed outcomes in the data, since physicians in practice may have incorporated some predictive information in their decision-making based on their experience and judgment.

To further validate our model, we train a practical predictive model using our data and a random forest algorithm to anticipate a patient's condition in the next decision epoch and measure the quality of the prediction using an out-of-sample misclassification matrix. Although the accuracy of the simple predictive model tested in our analysis is not very high, the results show that the optimal policy that incorporates those predictions still outperforms the policy that does not use predictive information. To analyze the impact of prediction precision on outcomes, we conduct numerical studies using a randomly generated misclassification matrix. As a comparison, we also investigate the case in which prediction errors are either not considered or ignored by applying the policy derived from the model with perfect predictions to inaccurate predictions. In both situations, we observe that predictive information can still improve decision quality, even with a moderately accurate predictive model. Moreover, we also estimate the potential impact of our proposed framework in the current COVID-19 pandemic and validate the value of incorporating predictive information in medical decision-making.

Our research can be extended in several directions. First, our proposed framework focuses on integrating predictive information and does not take into account many operational considerations, such as capacity rationing issues under stochastic demands. Future research could incorporate more operational aspects in the analysis. Second, this paper studies the optimal stopping problem for medical treatment. In practice, it is also critical to decide when to initiate treatment and how to control medication dosage and intensity, as well as other modalities during treatment. For example, one might consider the problem of optimal treatment initiation when treatment termination is determined by an optimal stopping model and analyze the value of predictive information. Third, although we validate the value of predictive information, it enlarges numerically with longer prediction windows. With a longer prediction window, theoretical analysis is not trivial. We leave these and other problems to future research.

References

Akan M, Alagoz O, Ata B, Erenay FS, Said A (2012) A broader view of designing the liver allocation system. *Operations Research* 60(4):757–770.

Alagoz O, Maillart LM, Schaefer AJ, Roberts MS (2004) The optimal timing of living-donor liver transplantation. *Management Science* 50(10):1420–1430.

Alagoz O, Maillart LM, Schaefer AJ, Roberts MS (2007a) Choosing among living-donor and cadaveric livers. *Management Science* 53(11):1702–1715.

Alagoz O, Maillart LM, Schaefer AJ, Roberts MS (2007b) Determining the acceptance of cadaveric livers using an implicit model of the waiting list. *Operations Research* 55(1):24–36.

Ayer T, Alagoz O, Stout NK (2012) OR Forum–A POMDP approach to personalize mammography screening decisions. *Operations Research* 60(5):1019–1034.

Ayer T, Alagoz O, Stout NK, Burnside ES (2016) Heterogeneity in women's adherence and its role in optimal breast cancer screening policies. *Management Science* 62(5):1339–1362.

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

Ayvaci MU, Alagoz O, Burnside ES (2012) The effect of budgetary restrictions on breast cancer diagnostic decisions. *Manufacturing & Service Operations Management* 14(4):600.

Barrett M, Smith M, Elixhauser A, Honigman L, Pines J (2014) Utilization of intensive care services, 2011. *HCUP statistical brief* 185.

Bellman R (1957) A Markovian decision process. *Journal of Mathematics and Mechanics* 679–684.

Bertsimas D, Kallus N (2019) From predictive to prescriptive analytics. *Management Science* .

Bertsimas D, Kallus N, Weinstein AM, Zhuo YD (2017) Personalized diabetes management using electronic medical records. *Diabetes Care* 40(2):210–217.

Bertsimas D, O'Hair A, Relyea S, Silberholz J (2016) An analytics approach to designing combination chemotherapy regimens for cancer. *Management Science* 62(5):1511–1531.

Bhatraju PK, Ghassemieh BJ, Nichols M, Kim R, Jerome KR, Nalla AK, Greninger AL, Pipavath S, Wurfel MM, Evans L, et al. (2020) Covid-19 in critically ill patients in the Seattle region—case series. *New England Journal of Medicine* .

Blum B (2018) Saving lives in the ICU through artificial intelligence. URL `https://www.israel21c.org/saving-lives-in-the-icu-through-artificial-intelligence/`.

Boloori A, Saghafian S, Chakkera HA, Cook CB (2020) Data-driven management of post-transplant medications: an ambiguous partially observable Markov decision process approach. *Manufacturing & Service Operations Management* .

Boyarchenko S, Levendorskii S (2007) *Irreversible decisions under uncertainty: optimal stopping made easy*, volume 27 (Springer Science & Business Media).

CDC COVID-19 Response Team (2020) Severe outcomes among patients with coronavirus disease 2019 (COVID-19)—United States, February 12–March 16, 2020. *MMWR Morb Mortal Wkly Rep* 69(12):343–346, URL http://dx.doi.org/http://dx.doi.org/10.15585/mmwr.mm6912e2.

Chan CW, Farias VF, Bambos N, Escobar GJ (2012) Optimizing intensive care unit discharge decisions with patient readmissions. *Operations Research* 60(6):1323–1341.

Chan CW, Green LV, Lu Y, Leahy N, Yurt R (2013) Prioritizing burn-injured patients during a disaster. *Manufacturing & Service Operations Management* 15(2):170–190.

Chao DC, Scheinhorn DJ (2007) Determining the best threshold of rapid shallow breathing index in a therapist-implemented patient-specific weaning protocol. *Respiratory care* 52(2):159–165.

Chen T, Xu J, Ying H, Chen X, Feng R, Fang X, Gao H, Wu J (2019) Prediction of extubation failure for intensive care unit patients using light gradient boosting machine. *IEEE Access* 7:150960–150968.

Chen Y, Farias VF (2013) Simple policies for dynamic pricing with imperfect forecasts. *Operations Research* 61(3):612–624.

Chhatwal J, Alagoz O, Burnside ES (2010) Optimal breast biopsy decision-making based on mammographic features and demographic factors. *Operations Research* 58(6):1577–1591.

Ching WK, Huang X, Ng MK, Siu TK (2013) Higher-order Markov chains. *Markov Chains*, 141–176 (Springer).

Ching WK, Ng MK, Fung ES (2008) Higher-order multivariate Markov chains and their applications. *Linear Algebra and its Applications* 428(2-3):492–507.

Ching WK, Ng MK, So MM (2004) Customer migration, campaign budgeting, revenue estimation: the elasticity of Markov decision process on customer lifetime value. *Advanced Modeling and Optimization* 6(2):65–80.

Chow Y, Robbins HA, Siegmund D (1971) Great expectations: the theory of optimal stopping.

Chung WC, Sheu CC, Hung JY, Hsu TJ, Yang SH, Tsai JR (2020) Novel mechanical ventilator weaning predictive model. *The Kaohsiung journal of medical sciences* 36(10):841–849.

Ciocan DF, Mišić VV (2020) Interpretable optimal stopping. *Management Science* .

Cochran WG (1968) The effectiveness of adjustment by subclassification in removing bias in observational studies. *Biometrics* 295–313.

Cohen IG, Crespo AM, White DB (2020) Potential legal liability for withdrawing or withholding ventilators during COVID-19: assessing the risks and identifying needed reforms. *JAMA* ISSN 0098-7484, URL http://dx.doi.org/10.1001/jama.2020.5442.

Dai J, Shi P (2018a) Inpatient overflow: an approximate dynamic programming approach. *Manufacturing & Service Operations Management* .

Dai J, Shi P (2018b) Recent modeling and analytical advances in hospital inpatient flow management. *Available at SSRN* URL `https://ssrn.com/abstract=3310853`.

Demling RH, Read T, Lind LJ, Flanagan HL (1988) Incidence and morbidity of extubation failure in surgical intensive care patients. *Critical care medicine* 16(6):573–577.

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

Desai VV, Farias VF, Moallemi CC (2012) Pathwise optimization for optimal stopping problems. *Management Science* 58(12):2292–2308.

Detemple J (2005) *American-style derivatives: Valuation and computation* (CRC Press).

Epstein SK, Ciubotaru RL, Wong JB (1997) Effect of failed extubation on the outcome of mechanical ventilation. *Chest* 112(1):186–192.

Erat S, Kavadias S (2008) Sequential testing of product designs: implications for learning. *Management Science* 54(5):956–968.

Erenay FS, Alagoz O, Said A (2014) Optimizing colonoscopy screening for colorectal cancer prevention and surveillance. *Manufacturing & Service Operations Management* 16(3):381–400.

Feng Y, Gallego G (1995) Optimal starting times for end-of-season sales and optimal stopping times for promotional fares. *Management science* 41(8):1371–1391.

Fitch ZW, Debesa O, Ohkuma R, Duquaine D, Steppan J, Schneider EB, Whitman GJ (2014) A protocol-driven approach to early extubation after heart surgery. *The Journal of thoracic and cardiovascular surgery* 147(4):1344–1350.

Fontela PS, Piva JP, Garcia PC, Bered PL, Zilles K (2005) Risk factors for extubation failure in mechanically ventilated pediatric patients. *Pediatric Critical Care Medicine* 6(2):166–170.

Girault C, Bubenheim M, Abroug F, Diehl JL, Elatrous S, Beuret P, Richecoeur J, L'Her E, Hilbert G, Capellier G, et al. (2011) Noninvasive ventilation and weaning in patients with chronic hypercapnic respiratory failure: a randomized multicenter trial. *American journal of respiratory and critical care medicine* 184(6):672–679.

Goldberg DA, Chen Y (2018) Beating the curse of dimensionality in options pricing and optimal stopping. *arXiv preprint arXiv:1807.02227* .

Guan Wj, Ni Zy, Hu Y, Liang Wh, Ou Cq, He Jx, Liu L, Shan H, Lei Cl, Hui DS, et al. (2020) Clinical characteristics of coronavirus disease 2019 in China. *New England Journal of Medicine* .

Halpern NA, Pastores SM (2015) Critical care medicine beds, use, occupancy and costs in the United States: a methodological review. *Critical care medicine* 43(11):2452.

Haugh MB, Kogan L (2004) Pricing American options: a duality approach. *Operations Research* 52(2):258–270.

Hess DR (2011) Approaches to conventional mechanical ventilation of the patient with acute respiratory distress syndrome. *Respiratory care* 56(10):1555–1572.

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.

Huang T, Bergman D, Gopal R (2018) Predictive and prescriptive analytics for location selection of add-on retail products. *Production and Operations Management* .

Iancu DA, Trichakis N, Yoon DY (2020) Monitoring with limited information. URL `https://web.stanford.edu/~daniancu/Papers/Working/Monitoring/monitoring_limited_info.pdf`.

Jagannatha A, Thomas P, Yu H (2018) Towards high confidence off-policy reinforcement learning for clinical applications. *CausalML Workshop, ICML*.

Kappen TH, van Klei WA, van Wolfswinkel L, Kalkman CJ, Vergouwe Y, Moons KG (2018) Evaluating the impact of prediction models: lessons learned, challenges, and recommendations. *Diagnostic and Prognostic Research* 2(1):11.

KC DS (2019) Heuristic thinking in patient care. *Management Science* .

KC DS, Terwiesch C (2009) Impact of workload on service time and patient safety: an econometric analysis of hospital operations. *Management Science* 55(9):1486–1498.

KC DS, Terwiesch C (2012) An econometric analysis of patient flows in the cardiac intensive care unit. *Manufacturing & Service Operations Management* 14(1):50–65.

Kim SH, Chan CW, Olivares M, Escobar G (2014) ICU admission control: an empirical study of capacity allocation and its implication for patient outcomes. *Management Science* 61(1):19–38.

Kim SH, Pinker E, Rimar J (2020) An empirical study of the effect of ICU capacity strain on patient discharge. *Available at SSRN* URL `https://ssrn.com/abstract=2644600`.

Krishnamurthy V (2016) *Partially observed Markov decision processes: from filtering to controlled sensing* (Cambridge University Press), URL `http://dx.doi.org/10.1017/CBO9781316471104`.

Kuntz L, Mennicken R, Scholtes S (2014) Stress on the ward: evidence of safety tipping points in hospitals. *Management Science* 61(4):754–771.

Kuo HJ, Chiu HW, Lee CN, Chen TT, Chang CC, Bien MY (2015) Improvement in the prediction of ventilator weaning outcomes by an artificial neural network in a medical ICU. *Respiratory care* 60(11):1560–1569.

Kwong MT, Colopy GW, Weber AM, Ercole A, Bergmann JH (2019) The efficacy and effectiveness of machine learning for weaning in mechanically ventilated patients at the intensive care unit: a systematic review. *Bio-Design and Manufacturing* 2(1):31–40.

Li L, Linetsky V (2013) Optimal stopping and early exercise: an eigenfunction expansion approach. *Operations Research* 61(3):625–643.

Luo D, Bayati M, Plambeck EL, Aratow M (????) Low-acuity patients delay high-acuity patients in an emergency department. *Available at SSRN* URL `https://ssrn.com/abstract=3095039`.

Mahle WT, Nicolson SC, Hollenbeck-Pringle D, Gaies MG, Witte MK, Lee EK, Goldsworthy M, Stark PC, Burns KM, Scheurer MA, Cooper DS, Thiagarajan R, Sivarajan VB, Colan SD, Schamberger MS, Shekerdemian LS (2016) Utilizing a collaborative learning model to promote early extubation following infant heart surgery. *Pediatric Critical Care Medicine* 17(10):939–947.

McConville JF, Kress JP (2012) Weaning patients from the ventilator. *New England Journal of Medicine* 367(23):2233–2239.

Meade M, Guyatt G, Cook D, Griffith L, Sinuff T, Kergl C, Mancebo J, Esteban A, Epstein S (2001) Predicting success in weaning from mechanical ventilation. *Chest* 120(6):400S–424S.

Mišić VV, Aleman DM, Sharpe MB (2010) Neighborhood search approaches to non-coplanar beam orientation optimization for total marrow irradiation using IMRT. *European Journal of Operational Research* 205(3):522–527.

Moran JL, Bristow P, Solomon PJ, George C, Hart GK (2008) Mortality and length-of-stay outcomes, 1993–2003, in the binational Australian and New Zealand intensive care adult patient database. *Critical care medicine* 36(1):46–61.

Müller A, Stoyan D (2002) *Comparison methods for stochastic models and risks*, volume 389 (Wiley New York).

Niskanen M, Reinikainen M, Pettilä V (2009) Case-mix-adjusted length of stay and mortality in 23 Finnish ICUs. *Intensive care medicine* 35(6):1060–1067.

Olsen L, Saunders RS, Yong PL, et al. (2010) *The healthcare imperative: lowering costs and improving outcomes: workshop series summary* (Washington, DC: National Academies Press).

Prasad N, Cheng LF, Chivers C, Draugelis M, Engelhardt BE (2017) A reinforcement learning approach to weaning of mechanical ventilation in intensive care units. *arXiv preprint arXiv:1704.06300* .

Raftery AE (1985) A model for high-order Markov chains. *Journal of the Royal Statistical Society: Series B (Methodological)* 47(3):528–539.

Randolph AG, Wypij D, Venkataraman ST, Hanson JH, Gedeit RG, Meert KL, Luckett PM, Forbes P, Lilley M, Thompson J, et al. (2002) Effect of mechanical ventilator weaning protocols on respiratory outcomes in infants and children: a randomized controlled trial. *Jama* 288(20):2561–2568.

Ranney ML, Griffeth V, Jha AK (2020) Critical supply shortages–the need for ventilators and personal protective equipment during the Covid-19 pandemic. *New England Journal of Medicine* .

Riedmiller M (2005) Neural fitted Q iteration–first experiences with a data efficient neural reinforcement learning method. *European Conference on Machine Learning*, 317–328 (Springer).

Rosenbaum PR, Rubin DB (1983) The central role of the propensity score in observational studies for causal effects. *Biometrika* 70(1):41–55.

Sato M, Suenaga E, Koga S, Matsuyama S, Kawasaki H, Maki F (2009) Early tracheal extubation after on-pump coronary artery bypass grafting. *Ann Thorac Cardiovasc Surg* 15(4):239–42.

Schaefer AJ, Bailey MD, Shechter SM, Roberts MS (2005) Modeling medical treatment using Markov decision processes. *Operations Research and Health Care*, 593–612 (Boston, MA: Springer).

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

Shi P, Helm J, Deglise-Hawkinson J, Pan J (2018) Timing it right: balancing inpatient congestion versus readmission risk at discharge. *Available at SSRN* URL `https://ssrn.com/abstract=3202975`.

Spencer J, Sudan M, Xu K (2014) Queuing with future information. *The Annals of Applied Probability* 24(5):2091–2142.

Sturt B (2021) A nonparametric algorithm for optimal stopping based on robust optimization. *arXiv preprint arXiv:2103.03300* .

Thille AW, Richard JCM, Brochard L (2013) The decision to extubate in the intensive care unit. *American Journal of Respiratory and Critical Care Medicine* 187(12):1294–1302.

Truog RD, Mitchell C, Daley GQ (2020) The toughest triage-—allocating ventilators in a pandemic. *New England Journal of Medicine* .

Tsai TL, Huang MH, Lee CY, Lai WW (2019) Data science for extubation prediction and value of information in surgical intensive care unit. *Journal of clinical medicine* 8(10):1709.

Verburg IW, de Keizer NF, de Jonge E, Peek N (2014) Comparison of regression methods for modeling intensive care length of stay. *PloS one* 9(10).

Vincent JL, Moreno R, Takala J, Willatts S, De Mendonça A, Bruining H, Reinhart C, Suter P, Thijs LG (1996) The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure.

Wang A, Mahfouf M, Mills GH, Panoutsos G, Linkens DA, Goode K, Kwok HF, Denaï M (2010) Intelligent model-based advisory system for the management of ventilated intensive care patients. Part II: advisory system design and evaluation. *Computer methods and programs in biomedicine* 99(2):208–217.

White DB, Angus DC (2020) A proposed lottery system to allocate scarce COVID-19 medications: promoting fairness and generating knowledge. *JAMA* .

White DB, Lo B (2020) A framework for rationing ventilators and critical care beds during the COVID-19 pandemic. *JAMA* .

Wilmore DW, Kehlet H (2001) Management of patients in fast track surgery. *British Medical Journal* 322(7284):473–476.

Wong WT, Lai VK, Chee YE, Lee A (2016) Fast-track cardiac care for adult cardiac surgical patients. *Cochrane Database of Systematic Reviews* (9).

Xie J, Cheng G, Zheng Z, Luo H, Ooi OC (2019) To extubate or not to extubate: risk factors for extubation failure and deterioration with further mechanical ventilation. *Journal of cardiac surgery* 34(10):1004.

Xu K (2015) Necessity of future information in admission control. *Operations Research* 63(5):1213–1226.

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 Z, Shi L, Wang Y, Zhang J, Huang L, Zhang C, Liu S, Zhao P, Liu H, Zhu L, et al. (2020) Pathological findings of COVID-19 associated with acute respiratory distress syndrome. *The Lancet respiratory medicine* 8(4):420–422.

Yang KL, Tobin MJ (1991) A prospective study of indexes predicting the outcome of trials of weaning from mechanical ventilation. *New England Journal of Medicine* 324(21):1445–1450.

Zhang J, Denton BT, Balasubramanian H, Shah ND, Inman BA (2012) Optimization of prostate biopsy referral decisions. *Manufacturing & Service Operations Management* 14(4):529–547.

Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, Xiang J, Wang Y, Song B, Gu X, et al. (2020) Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *The Lancet* .

Zhu F, Lee A, Chee YE (2012) Fast-track cardiac care for adult cardiac surgical patients. *Cochrane Database of Systematic Reviews* (10).

Appendix: Proofs and Additional Results

EC.1. Base Model without Predictive Information

In the base scenario, we consider the extubation problem without predictive information in a conventional way. We formulate this decision process as a finite-state, finite-horizon, discrete-time MDP model, denoted as the MDP-B model. Most of the components are the same as in MDP-P, except for state and transition probabilities. In the basic setting, the system state is just patients' current class; thus, the transition probabilities are fully characterized by matrix P.

The optimality equation for MDP-B can be written as

$$J_t^B(c_t) = \min\Big\{G(c_t),\ H(c_t) + \sum_{c_{t+1}\in\mathcal{C}} P(c_t, c_{t+1}) J_{t+1}^B(c_{t+1})\Big\}, \quad \forall c_t \in \mathcal{C}, t \in \mathcal{T}\backslash\{T\}, \tag{EC.1}$$

where $J_t^B(c_t)$ represents the minimum expected total cost incurred by a class c_t patient at epoch t, and the boundary value is

$$J_T^B(c_T) = G(c_T), \quad \forall c_T \in \mathcal{C}. \tag{EC.2}$$

To solve the optimality equation (EC.1) and obtain the optimal policy for the MDP-B model, we first make the following assumptions, which are reasonable in a healthcare setting. Our first assumption is related to the definition of first-order stochastic dominance. For probability vectors $\boldsymbol{x}$ and $\boldsymbol{y}$, $\boldsymbol{x}$ is said to first-order stochastically dominate $\boldsymbol{y}$, denoted as $\boldsymbol{x} \geq_{st} \boldsymbol{y}$, if $\sum_{i\geq k} \boldsymbol{x}(i) \geq \sum_{i\geq k} \boldsymbol{y}(i)$ for any k. In decision theory, it implies that $\boldsymbol{x}$ is preferable to $\boldsymbol{y}$.

Assumption EC.1. *$P_j \geq_{st} P_i$ for all $i, j \in \mathcal{C}$ such that $i < j$.*

Assumption EC.1 defines the structure of the transition probability matrix P in the sense of first-order stochastic dominance. It implies that a healthier patient in class j will evolve to a better state in expectation than a less healthy patient in class i, where $i, j \in \mathcal{C}$ and $i < j$. The next assumption is a joint condition of the transition probability matrix and cost functions. For a class c patient, $\sum_{c'\in\mathcal{C}} P(c, c')G(c')$ is the expected terminal cost after one more period of treatment. Hence, the difference $G(c) - \sum_{c'\in\mathcal{C}} P(c, c')G(c') - H$ represents the net benefit from one more period of treatment. The following assumption states that the net benefit from one more period of treatment diminishes with the patient's health condition.

Assumption EC.2. *$G(c) - \sum_{c'\in\mathcal{C}} P(c, c')G(c') - H$ is nonincreasing in c.*

In other words, a healthier patient gets less benefit from one more period of treatment.

Before characterizing the structure of the optimal policy for the MDP-B model, we first investigate the property of the optimal value function in the following lemma.

LEMMA EC.1. *Under Assumption EC.1, the optimal value function $J_t^B(c_t)$ for the MDP-B model is nonincreasing in c_t for any $t \in \mathcal{T}$.*

Lemma EC.1 asserts that a healthier patient has lower expected remaining treatment cost, given that the patient has been treated for t periods. The structure of the optimal policy for the MDP-B model is then characterized in the following theorem.

THEOREM EC.1. *Under Assumptions EC.1–EC.2, there exists an optimal threshold c_t^* in decision epoch $t \in \mathcal{T}\backslash\{T\}$ for the MDP-B model, such that it is optimal to stop the treatment if $c_t \geq c_t^*$, and to continue the treatment otherwise. Moreover, c_t^* is nonincreasing in t. That is, $c_1^* \geq c_2^* \geq \cdots \geq c_{T-1}^*$.*

Theorem EC.1 states that the optimal policy for MDP-B is a threshold policy. There exists an optimal threshold class in each epoch $t \in \mathcal{T}\backslash\{T\}$. If patient class c_t is not worse than optimal stopping class c_t^*, it is optimal to stop the treatment for the patient. Otherwise, continuing the treatment is the optimal action. The monotonicity of the optimal threshold c_t^* in time t reveals that the optimal stopping criterion can be relaxed if the patient has been treated for some periods. In the following proposition, we discuss a special case in which the optimal threshold is stationary for time t.

PROPOSITION EC.1. *Assume a subgroup of patients who never deteriorate to worse conditions under the treatment, i.e., $P(i,j) = 0$ if $j < i$ for $i, j \in \mathcal{C}$. Then the optimal threshold in Theorem EC.1 for this group of patients is independent of t. That is, $c_1^* = c_2^* = \cdots = c_{T-1}^* = c^*$. More specifically,*

$$c^* = \min\left\{ c \in \mathcal{C} : G(c) - H - \sum_{c' \in \mathcal{C}} P(c, c')G(c') \leq 0 \right\}.$$

Proposition EC.1 implies that a stationary stopping criterion can be applied to a subgroup of patients who only evolve to better conditions under treatment. This result is particularly useful because, under proper treatments, majority of patients will recover. In our data, there are 2,323 (76%) patients never experienced deterioration during ventilation. Although some patients' conditions deteriorated under ventilation, in our case study the optimal policy based on the MDP-B model is still a stationary threshold policy.

EC.2. Technical Proofs

EC.2.1. Proof of Lemma EC.1

The proof proceeds by induction on decision epoch t.

For $t = T.$, according to Equation (EC.2), $J_T^B(c_T)$ is nonincreasing in c_T since $G(c_T)$ is nonincreasing in c_T. Next, we assume that in epoch $t+1$, $J_{t+1}^B(c_{t+1})$ is nonincreasing in c_{t+1}. In epoch t, based on Equation (EC.1), we have

$$J_t^B(c_t) = \min\Big\{G(c_t),\ H(c_t) + \sum_{c_{t+1}\in\mathcal{C}} P(c_t, c_{t+1}) J_{t+1}^B(c_{t+1})\Big\}.$$

Note that Assumption EC.1 (i.e., $P_j \geq_{st} P_i$ for $i < j$) implies that

$$P_i \boldsymbol{g} \geq P_j \boldsymbol{g},\ \forall i, j \in \mathcal{C}, i < j, \tag{EC.3}$$

for any C dimensional column vector $\boldsymbol{g}$ with nonincreasing components, i.e., $g_1 \leq g_2 \leq \cdots \leq g_C$ (Müller and Stoyan 2002). According to the induction hypothesis, we have $\sum_{c_{t+1}\in\mathcal{C}} P(c_t, c_{t+1}) J_{t+1}^B(c_{t+1})$ is nonincreasing in c_t. Since we have both $G(c_t)$ and $H(c_t)$ nonincreasing in c_t, we get $J_t^B(c_t)$ is nonincreasing in c_t, which completes the proof.

EC.2.2. Proof of Theorem EC.1

To prove Theorem EC.1, we first define the following function:

$$h_t^B(c_t) := G(c_t) - H(c_t) - \sum_{c_{t+1}\in\mathcal{C}} P(c_t, c_{t+1}) J_{t+1}^B(c_{t+1}),\ c_t \in \mathcal{C}, t \in \mathcal{T}\backslash\{T\}. \tag{EC.4}$$

$h_t^B(c_t)$ represents the benefit from continuing the treatment. It is optimal to stop the treatment if $h_t^B(c_t) \leq 0$, and to continue otherwise. Then the theorem is equivalent to the following statements:

(i) $h_t^B(c_t)$ is nonincreasing in c_t for $t \in \mathcal{T}\backslash\{T\}$.

(ii) $h_t^B(c) \geq h_{t+1}^B(c)$ for all $c \in \mathcal{C}$ and $t = 1, 2, \ldots, T-2$.

Part (i) indicates that the optimal policy is a threshold policy and part (ii) implies that the optimal threshold c_t^* is nonincreasing in t. We now prove these two statements instead.

We first prove part (i) by induction on decision epoch t. In decision epoch $T-1$, we have

$$\begin{aligned} h_{T-1}^B(c_{T-1}) &= G(c_{T-1}) - H(c_{T-1}) - \sum_{c_T\in\mathcal{C}} P(c_{T-1}, c_T) J_T^B(c_T) \\ &= G(c_{T-1}) - H(c_{T-1}) - \sum_{c_T\in\mathcal{C}} P(c_{T-1}, c_T) G(c_T). \end{aligned}$$

The first equality follows from Equation (EC.4). The second equality is the boundary condition (EC.2). Assumption EC.2 immediately implies that $h_{T-1}^B(c_{T-1})$ is nonincreasing in c_{T-1}. Assume that in epoch $t+1$, $h_{t+1}^B(c_{t+1})$ is nonincreasing in c_{t+1}. Thus there exists an optimal threshold c_{t+1}^* in epoch $t+1$. For $c_{t+1} \geq c_{t+1}^*$, $J_{t+1}^B(c_{t+1}) = G(c+1)$. Otherwise, $J_{t+1}^B(c_{t+1}) = G(c_{t+1}) - h_{t+1}^B(c_{t+1})$. In epoch t, we have

$$\begin{aligned} h_t^B(c_t) &= G(c_t) - H(c_t) - \sum_{c_{t+1}\in\mathcal{C}} P(c_t, c_{t+1}) J_{t+1}^B(c_{t+1}) \\ &= G(c_t) - H(c_t) - \sum_{c_{t+1}\in\mathcal{C}} P(c_t, c_{t+1}) G(c_{t+1}) + \sum_{c_{t+1} < c_{t+1}^*} P(c_t, c_{t+1}) h_{t+1}^B(c_{t+1}). \end{aligned}$$

The first equality follows from Equation (EC.4). The second equality follows from the induction hypothesis. $\sum_{c_{t+1}<c^*_{t+1}} P(c_t, c_{t+1}) h^B_{t+1}(c_{t+1})$ is nonincreasing in c_t from Assumption EC.1 and the induction hypothesis, and $G(c_t) - H(c_t) - \sum_{c_{t+1}\in\mathcal{C}} P(c_t, c_{t+1}) G(c_{t+1})$ is nonincreasing in c_t from Assumption EC.2. Hence, $h^B_t(c_t)$ is nonincreasing in c_t, which completes the proof for part (i).

To prove part (ii), we first establish the following claim.

CLAIM EC.1. *$J^B_t(c) \le J^B_{t+1}(c)$ for $c \in \mathcal{C}$ and $t \in \mathcal{T}\backslash\{T\}$.*

Proof. We prove this claim by induction on decision epoch t. When $t = T-1$, we have

$$\begin{aligned} J^B_{T-1}(c) &= \min\Big\{G(c),\ H + \sum_{c'\in\mathcal{C}} P(c,c') J^B_T(c')\Big\} \\ &\le G(c) \\ &= J^B_T(c). \end{aligned}$$

The first equality follows the definition, and the third one comes from the boundary condition.

Assume the claim holds in epoch $t+1$. In epoch t, we have

$$\begin{aligned} J^B_t(c) &= \min\Big\{G(c),\ H + \sum_{c'\in\mathcal{C}} P(c,c') J^B_{t+1}(c')\Big\} \\ &\le \min\Big\{G(c),\ H + \sum_{c'\in\mathcal{C}} P(c,c') J^B_{t+2}(c')\Big\} \\ &= J^B_{t+1}(c). \end{aligned}$$

The inequality follows from the induction hypothesis. Thus $J^B_t(c)$ in nondecreasing in t. ■

Now we are ready to prove part (2). We have

$$\begin{aligned} h^B_t(c) &= G(c) - H - \sum_{c'\in\mathcal{C}} P(c,c') J^B_{t+1}(c') \\ &\ge G(c) - H - \sum_{c'\in\mathcal{C}} P(c,c') J^B_{t+2}(c') \\ &= h^B_{t+1}(c). \end{aligned}$$

The inequality follows from Claim EC.1. Hence, part (ii) holds, and we complete the proof for Theorem EC.1.

EC.2.3. Proof of Proposition EC.1

As stated in Proposition EC.1, let

$$c^* = \min\Big\{c \in \mathcal{C} : G(c) - H - \sum_{c'\in\mathcal{C}} P(c,c') G(c') \le 0\Big\}. \tag{EC.5}$$

To prove this proposition, we need to show that $h^B_t(c_t) \le 0$ if $c_t \ge c^*$, and $h^B_t(c_t) > 0$ otherwise for $t \in \mathcal{T}\backslash\{T\}$.

The proof proceeds by induction on decision epoch t. When $t = T-1$, we have

$$\begin{aligned} h^B_{T-1}(c_{T-1}) &= G(c_{T-1}) - H(c_{T-1}) - \sum_{c_T \in \mathcal{C}} P(c_{T-1}, c_T) J^B_T(c_T) \\ &= G(c_{T-1}) - H(c_{T-1}) - \sum_{c_T \in \mathcal{C}} P(c_{T-1}, c_T) G(c_T). \end{aligned}$$

The equalities follow the definition of $h^B_t(c_t)$ and the boundary condition. It is obvious that $h^B_{T-1}(c_{T-1}) \le 0$ if $c_{T-1} \ge c^*$, and $h^B_{T-1}(c_{T-1}) > 0$ otherwise from the definition of c^*. Assume that the condition holds for epoch $t+1$. Thus, for $c_{t+1} \ge c^*$, $J^B_{t+1}(c_{t+1}) = G(c_{t+1})$. Otherwise, $J^B_{t+1}(c_{t+1}) = G(c_{t+1}) - h^B_{t+1}(c_{t+1})$. In epoch t, we have

$$\begin{aligned} h^B_t(c_t) &= G(c_t) - H(c_t) - \sum_{c_{t+1} \in \mathcal{C}} P(c_t, c_{t+1}) J^B_{t+1}(c_{t+1}) \\ &= G(c_t) - H(c_t) - \sum_{c_{t+1} \in \mathcal{C}} P(c_t, c_{t+1}) G(c_{t+1}) + \sum_{c_{t+1} < c^*} P(c_t, c_{t+1}) h^B_{t+1}(c_{t+1}). \end{aligned}$$

The second equality follows from the induction hypothesis. Now we consider the following two cases.

(a) If $c_t \ge c^*$, $G(c_t) - H(c_t) - \sum_{c_{t+1} \in \mathcal{C}} P(c_t, c_{t+1}) G(c_{t+1}) \le 0$ according to (EC.5); $\sum_{c_{t+1} < c^*} P(c_t, c_{t+1}) h^B_{t+1}(c_{t+1}) = 0$ since $P(c_t, c_{t+1}) = 0$ when $c_{t+1} < c_t$ in this special case. Hence, $h^B_t(c_t) \le 0$ for $c_t \ge c^*$.

(b) If $c_t < c^*$, $G(c_t) - H(c_t) - \sum_{c_{t+1} \in \mathcal{C}} P(c_t, c_{t+1}) G(c_{t+1}) > 0$ according to (EC.5); and $\sum_{c_{t+1} < c^*} P(c_t, c_{t+1}) h^B_{t+1}(c_{t+1}) \ge 0$ according to the induction hypothesis. Thus $h^B_t(c_t) > 0$ for $c_t < c^*$.

Combining (a) and (b), we complete the proof.

EC.2.4. Proof of Lemma 1

To prove this lemma, we require the following result stating that the belief is larger in the sense of MLR dominance if the current class or the predicted class in next epoch is healthier.

Claim EC.2. *Given current class $c_t \in \mathcal{C}$, for $\hat{c}_{t+1}, \hat{c}'_{t+1} \in \mathcal{C}$, $\hat{c}_{t+1} \ge \hat{c}'_{t+1}$ implies $[\pi(c_{t+1} \mid c_t, \hat{c}_{t+1})]_{c_{t+1} \in \mathcal{C}} \ge_{lr} [\pi(c_{t+1} \mid c_t, \hat{c}'_{t+1})]_{c_{t+1} \in \mathcal{C}}$. Similarly, given predicted class $\hat{c}_{t+1} \in \mathcal{C}$, for $c_t, c'_t \in \mathcal{C}$, $c_t \ge c'_t$ implies $[\pi(c_{t+1} \mid c_t, \hat{c}_{t+1})]_{c_{t+1} \in \mathcal{C}} \ge_{lr} [\pi(c_{t+1} \mid c'_t, \hat{c}_{t+1})]_{c_{t+1} \in \mathcal{C}}$.*

Proof. According to the definition of MLR dominance, $[\pi(c_{t+1} \mid c_t, \hat{c}_{t+1})]_{c_{t+1} \in \mathcal{C}} \ge_{lr} [\pi(c_{t+1} \mid c_t, \hat{c}'_{t+1})]_{c_{t+1} \in \mathcal{C}}$ is equivalent to

$$\pi(c_{t+1} \mid c_t, \hat{c}_{t+1}) \pi(c'_{t+1} \mid c_t, \hat{c}'_{t+1}) \ge \pi(c'_{t+1} \mid c_t, \hat{c}_{t+1}) \pi(c_{t+1} \mid c_t, \hat{c}'_{t+1}), \quad \forall c_{t+1} > c'_{t+1}.$$

From Equation (1), the above inequality is equivalent to

$$\frac{P(c_t,c_{t+1})Q(c_{t+1},\hat{c}_{t+1})P(c_t,c'_{t+1})Q(c'_{t+1},\hat{c}'_{t+1})}{\left[\sum_{c\in\mathcal{C}}P(c_t,c)Q(c,\hat{c}_{t+1})\right]\left[\sum_{c\in\mathcal{C}}P(c_t,c)Q(c,\hat{c}'_{t+1})\right]}\geq$$
$$\frac{P(c_t,c'_{t+1})Q(c'_{t+1},\hat{c}_{t+1})P(c_t,c_{t+1})Q(c_{t+1},\hat{c}'_{t+1})}{\left[\sum_{c\in\mathcal{C}}P(c_t,c)Q(c,\hat{c}_{t+1})\right]\left[\sum_{c\in\mathcal{C}}P(c_t,c)Q(c,\hat{c}'_{t+1})\right]},\quad \forall c_{t+1}>c'_{t+1}.$$

That is,

$$P(c_t,c_{t+1})P(c_t,c'_{t+1})[Q(c_{t+1},\hat{c}_{t+1})Q(c'_{t+1},\hat{c}'_{t+1})-Q(c'_{t+1},\hat{c}_{t+1})Q(c_{t+1},\hat{c}'_{t+1})]\geq 0,\quad \forall c_{t+1}>c'_{t+1}.$$

Since $P(c_t,c_{t+1})P(c_t,c'_{t+1})\geq 0$, it is simplified to

$$Q(c_{t+1},\hat{c}_{t+1})Q(c'_{t+1},\hat{c}'_{t+1})-Q(c'_{t+1},\hat{c}_{t+1})Q(c_{t+1},\hat{c}'_{t+1})\geq 0,\quad \forall c_{t+1}>c'_{t+1}.$$

According to Assumption 2, the above inequality immediately holds if $\hat{c}_{t+1}>\hat{c}'_{t+1}$. The proof of the other part is similar and we bypass it here. ■

We then prove that $J_t^P(c_t,\hat{c}_{t+1})$ is nonincreasing in both c_t and $\hat{c}_{t+1}$ by induction on decision epoch t. First, when $t=T$,

$$J_T^P(c_T,\hat{c}_{T+1})=G(c_T).$$

The lemma immediately holds.

Assume that $J_{t+1}^P(c_{t+1},\hat{c}_{t+2})$ is nonincreasing in c_{t+1} and $\hat{c}_{t+2}$. Note that $\tilde{Q}$ is TP2 under Assumptions 1 and 2 (see Krishnamurthy 2016, Lemma 10.5.2), and MLR dominance implies first-order stochastic dominance. Meanwhile, $J_{t+1}^P(c_{t+1},\hat{c}_{t+2})$ is nonincreasing in both c_{t+1} and $\hat{c}_{t+2}$. Then, $\sum_{\hat{c}_{t+2}\in\mathcal{C}}\tilde{Q}(c_{t+1},\hat{c}_{t+2})J_{t+1}^P(c_{t+1},\hat{c}_{t+2})$ is nonincreasing in c_{t+1}. Together with Claim EC.2, $\sum_{c_{t+1}\in\mathcal{C}}\pi(c_{t+1}\mid c_t,\hat{c}_{t+1})\sum_{\hat{c}_{t+2}\in\mathcal{C}}\tilde{Q}(c_{t+1},\hat{c}_{t+2})J_{t+1}^P(c_{t+1},\hat{c}_{t+2})$ is nonincreasing in both c_t and $\hat{c}_{t+1}$. $G(c_t)$ and $H(c_t)$ are assumed as nonincreasing in c_t. Therefore, $J_t^P(c_t,\hat{c}_{t+1})$ is nonincreasing in c_t and $\hat{c}_{t+1}$.

EC.2.5. Proof of Theorem 1

We first define

$$h_t^P(c_t,\hat{c}_{t+1}):=G(c_t)-H(c_t)-\sum_{c_{t+1}\in\mathcal{C}}\left[\pi(c_{t+1}\mid c_t,\hat{c}_{t+1})\sum_{\hat{c}_{t+2}\in\mathcal{C}}\tilde{Q}(c_{t+1},\hat{c}_{t+2})J_{t+1}^P(c_{t+1},\hat{c}_{t+2})\right],$$
$$t\in\mathcal{T}\backslash\{T\},c_t\in\mathcal{C},\hat{c}_{t+1}\in\mathcal{C}.$$

Theorem 1 is then equivalent to the following statements:

(i) $h_t^P(c_t,\hat{c}_{t+1})$ is nonincreasing in c_t.

(ii) $h_t^P(c_t, \hat{c}_{t+1})$ is nondecreasing in $\hat{c}_{t+1}$.

(iii) $h_t^P(c_t, \hat{c}_{t+1})$ is nonincreasing in t.

From Section EC.2.4, $\sum_{c_{t+1}\in\mathcal{C}}\left[\pi(c_{t+1}\mid c_t,\hat{c}_{t+1})\cdot\sum_{\hat{c}_{t+2}\in\mathcal{C}}\tilde{Q}(c_{t+1},\hat{c}_{t+2})\cdot J_{t+1}^P(c_{t+1},\hat{c}_{t+2})\right]$ is nonincreasing in $\hat{c}_{t+1}$. Hence, $h_t^P(c_t,\hat{c}_{t+1})$ is nondecreasing in $\hat{c}_{t+1}$, and part (ii) holds.

We next prove part (i) by induction on decision epoch t. First, in epoch $T-1$,

$$h_{T-1}^P(c_{T-1},\hat{c}_T)=G(c_{T-1})-H(c_{T-1})-\sum_{c_T\in\mathcal{C}}\pi(c_T\mid c_{T-1},\hat{c}_T)\cdot G(c_T).$$

Part (i) holds under Assumption 3. Assume that Statement (i) holds in epoch $t+1$. Similar to the proof of Theorem EC.1 in Section EC.2.2, in epoch t, with induction hypothesis and part (ii), we have

$$\begin{aligned}h_t^P(c_t,\hat{c}_{t+1})=&G(c_t)-H(c_t)-\sum_{c_{t+1}\in\mathcal{C}}\left[\pi(c_{t+1}\mid c_t,\hat{c}_{t+1})\sum_{\hat{c}_{t+2}\in\mathcal{C}}\tilde{Q}(c_{t+1},\hat{c}_{t+2})J_{t+1}^P(c_{t+1},\hat{c}_{t+2})\right]\\ =&G(c_t)-H(c_t)-\sum_{c_{t+1}\in\mathcal{C}}\pi(c_{t+1}\mid c_t,\hat{c}_{t+1})G(c_{t+1})+\\ &\sum_{c_{t+1}<c_{t+1}^*}\left[\pi(c_{t+1}\mid c_t,\hat{c}_{t+1})\sum_{\hat{c}_{t+2}<\hat{c}_{t+2}^*}\tilde{Q}(c_{t+1},\hat{c}_{t+2})h_{t+1}^P(c_{t+1},\hat{c}_{t+2})\right],\end{aligned}$$

where c_{t+1}^* and $\hat{c}_{t+2}^*$ are the optimal thresholds for the patient class and the predicted future patient class in epoch $t+1$. Under the induction hypothesis, it is nonincreasing in c_t based on Assumptions 1–3. Therefore, part (i) holds.

We establish the following result before proving part (iii).

Claim EC.3. *$J_t^P(c,\hat{c})\le J_{t+1}^P(c,\hat{c})$, $\forall t\in\mathcal{T}\backslash\{T\}$, $c\in\mathcal{C}$ and $\hat{c}\in\mathcal{C}$.*

Proof. The proof proceeds by induction on decision epoch t. When $t=T-1$, we have

$$\begin{aligned}J_{T-1}^P(c,\hat{c})&=\min\left\{G(c),\quad H+\sum_{c'\in\mathcal{C}}\left[\pi(c'\mid c,\hat{c})\sum_{\tilde{c}\in\mathcal{C}}\tilde{Q}(c',\tilde{c})J_T^P(c',\tilde{c})\right]\right\}\\ &\le G(c)\\ &=J_T^P(c,\hat{c}).\end{aligned}$$

The first equality follows the definition, and the third one follows from the boundary condition. Assume that the property holds in epoch $t+1$. In epoch t, we have

$$\begin{aligned}J_t^P(c,\hat{c})&=\min\left\{G(c),H+\sum_{c'\in\mathcal{C}}\left[\pi(c'\mid c,\hat{c})\sum_{\tilde{c}\in\mathcal{C}}\tilde{Q}(c',\tilde{c})J_{t+1}^P(c',\tilde{c})\right]\right\}\\ &\le\min\left\{G(c),H+\sum_{c'\in\mathcal{C}}\left[\pi(c'\mid c,\hat{c})\sum_{\tilde{c}\in\mathcal{C}}\tilde{Q}(c',\tilde{c})J_{t+2}^P(c',\tilde{c})\right]\right\}\\ &=J_{t+1}^P(c,\hat{c}).\end{aligned}$$

The inequality follows from the induction hypothesis. ■

We then prove that $h_t^P(c,\hat{c})$ nonincreasing in t as follows.

$$\begin{aligned} h_t^P(c,\hat{c}) &= G(c) - H - \sum_{c'\in\mathcal{C}} \left[\pi(c' \mid c,\hat{c}) \sum_{\tilde{c}\in\mathcal{C}} \tilde{Q}(c',\tilde{c}) J_{t+1}^P(c',\tilde{c}) \right] \\ &\geq G(c) - H - \sum_{c'\in\mathcal{C}} \left[\pi(c' \mid c,\hat{c}) \sum_{\tilde{c}\in\mathcal{C}} \tilde{Q}(c',\tilde{c}) J_{t+2}^P(c',\tilde{c}) \right] \\ &= h_{t+1}^P(c,\hat{c}). \end{aligned}$$

The inequality follows from Claim EC.3. Therefore, we complete the proof.

EC.2.6. Proof of Proposition 2

We prove this proposition by induction on decision epoch t. First, for $t=T$, we have

$$\begin{aligned} J_T^B(c_T) &= G(c_T) \\ &= \sum_{\hat{c}_{T+1}\in\mathcal{C}} \tilde{Q}(c_T,\hat{c}_{T+1}) G(c_T) \\ &= \sum_{\hat{c}_{T+1}\in\mathcal{C}} \tilde{Q}(c_T,\hat{c}_{T+1}) J_T^P(c_T,\hat{c}_{T+1}) \\ &= J_T^P(c_T). \end{aligned}$$

The equalities follow from definitions and the boundary condition. Thus the proposition holds in epoch T. Next, we assume that the statement holds in epoch $t+1$. That is,

$$J_{t+1}^B(c_{t+1}) \geq J_{t+1}^P(c_{t+1}),\ \forall c_{t+1}\in\mathcal{C}.$$

Considering epoch t, we have

$$\begin{aligned} J_t^B(c_t) &= \min\left\{ G(c_t),\ H(c_t) + \sum_{c_{t+1}\in\mathcal{C}} P(c_t,c_{t+1}) J_{t+1}^B(c_{t+1}) \right\} \\ &\geq \min\left\{ G(c_t),\ H(c_t) + \sum_{c_{t+1}\in\mathcal{C}} \left[P(c_t,c_{t+1}) \sum_{\hat{c}_{t+2}\in\mathcal{C}} \tilde{Q}(c_{t+1},\hat{c}_{t+2}) J_{t+1}^P(c_{t+1},\hat{c}_{t+2}) \right] \right\} \\ &= \min\left\{ G(c_t),\ H(c_t) + \sum_{\hat{c}_{t+1}\in\mathcal{C}} \left\{ \tilde{Q}(c_t,\hat{c}_{t+1}) \sum_{c_{t+1}\in\mathcal{C}} \left[\pi(c_{t+1} \mid c_t,\hat{c}_{t+1}) \sum_{\hat{c}_{t+2}\in\mathcal{C}} \tilde{Q}(c_{t+1},\hat{c}_{t+2}) J_{t+1}^P(c_{t+1},\hat{c}_{t+2}) \right] \right\} \right\} \\ &\geq \sum_{\hat{c}_{t+1}\in\mathcal{C}} \left\{ \tilde{Q}(c_t,\hat{c}_{t+1}) \min\left\{ G(c_t),\ H(c_t) + \sum_{c_{t+1}\in\mathcal{C}} \left[\pi(c_{t+1} \mid c_t,\hat{c}_{t+1}) \sum_{\hat{c}_{t+2}\in\mathcal{C}} \tilde{Q}(c_{t+1},\hat{c}_{t+2}) J_{t+1}^P(c_{t+1},\hat{c}_{t+2}) \right] \right\} \right\} \\ &= \sum_{\hat{c}_{t+1}\in\mathcal{C}} \tilde{Q}(c_t,\hat{c}_{t+1}) J_t^P(c_t,\hat{c}_{t+1}) \\ &= J_t^P(c_t). \end{aligned}$$

The first inequality follows from the induction hypothesis, and the second inequality follows from Jensen's inequality. The second equality follows from the definitions of $\tilde{Q}$ and π. Therefore, Theorem 2 holds.

EC.2.7. Proof of Proposition 3

Note that if patients are in class 2, their treatment will be stopped because no further benefit can be gained from continued treatment. Therefore, we only need to consider the treatment cost for class 1 patients, denoted as H. Based on Theorem EC.1, we know that the optimal policy without prediction (OP-B) is a time-dependent threshold policy in which the optimal threshold is nonincreasing in epoch t. The class of monotonic threshold policies in epoch t can be fully captured by a parameter t_B in the following way:

$$a_t(c;t_B) = \begin{cases} 0 & \text{if } c=1 \text{ and } t<t_B, \\ 1 & \text{otherwise,} \end{cases}$$

where $a_t(c;t_B)$ denotes the action when the patient is in class c in epoch t. Then we can represent OP-B by t_B^*, where

$$t_B^* = \arg\min_{t_B \in \mathcal{T}} J_1^B(1;t_B),$$

and $J_1^B(1;t_B)$ denotes the expected total cost of a class 1 patient under the nonincreasing-threshold policy characterized by t_B in the MDP-B model. Similarly, from Theorem 1, the optimal policy under predictive information can be captured by two parameters, denoted as t_{P1} and t_{P2}, where $t_{P1} \le t_{P2}$ such that the action for state $(c,\hat{c})$ in epoch t is defined as follows:

$$a_t(c,\hat{c};t_{P1},t_{P2}) = \begin{cases} 0 & \text{if } c=1 \text{ and } t<t_{P1}, \\ 0 & \text{if } c=1, \hat{c}=2 \text{ and } t<t_{P2}, \\ 1 & \text{otherwise.} \end{cases}$$

The first parameter t_{P1} captures the threshold to stop the treatment regardless of the prediction, and the second parameter t_{P2} is the decision boundary when the patient is predicted to evolve into the better class in the next epoch. The optimal policy under predictive information can be represented by (t_{P1}^*, t_{P2}^*), where

$$(t_{P1}^*, t_{P2}^*) = \arg\min_{t_{P1} \le t_{P2} \in \mathcal{T}} J_1^P(1;t_{P1},t_{P2}),$$

and $J_1^P(1;t_{P1},t_{P2})$ represents the expected total cost of a class 1 patient under the policy characterized by t_{P1} and t_{P2} in the MDP-P model.

Based on these settings, we can derive the expected total cost for class 1 patients under the threshold policy parameterized by t_B in the MDP-B model as

$$\begin{aligned} J_1^B(1;t_B) =& p_1^{t_B-1}[p_1 G(1) + (1-p_1)G(2)] + H(1-p_1)\sum_{i=1}^{t_B} i p_1^{i-1} + H p_1^{t_B} \\ =& G(1)p_1^{t_B} + G(2)(1-p_1^{t_B}) + H\frac{p_1 - p_1^{t_B}}{1-p_1} \\ =& \left(\Delta G - \frac{H}{1-p_1}\right) p_1^{t_B} + G(2) + \frac{H p_1}{1-p_1}, \end{aligned} \tag{EC.6}$$

where $\Delta G = G(1) - G(2)$, representing the potential benefit of stopping the treatment in a better class. The first two terms in the first line of the above equation represent the expected terminal cost, and the last two terms represent the expected treatment cost. Note that under the threshold policy parameterized by t_B, the probability of terminating the treatment in class 1 is $p_1^{t_B}$, i.e., the patient stays in class 1 after t_B epochs. For the second equality, we need to derive the expression for $(1-p_1)\sum_{i=1}^{t_B} i p_1^{i-1}$. Denote $S := \sum_{i=1}^{t_B} i p_1^{i-1}$. Then we have $p_1 S = \sum_{i=1}^{t_B} i p_1^i$. Subtracting the two equations, we can obtain

$$\begin{aligned}(1-p_1)S =& 1 + \sum_{i=1}^{t_B-1} p_1^i - t_B p_1^{t_B} \\ =& 1 + \frac{p_1(1-p_1^{t_B-1})}{1-p_1} - t_B p_1^{t_B} \\ =& \frac{1-(t_B+1)p_1^{t_B} + t_B p_1^{t_B+1}}{1-p_1}.\end{aligned} \tag{EC.7}$$

With simple algebra, we can derive the remaining equalities in Equation (EC.6).

Similarly, the total cost under the threshold policy parameterized by (t_{P1}, t_{P2}) in the MDP-P model can be written as

$$\begin{aligned}J_1^P(1; t_{P1}, t_{P2}) =& (1-p_1)[G(2)+H] + p_1(1-p_1)[G(2)+2H] + \cdots + \\ & p_1^{t_{P1}-1}[(1-p_1)G(2) + p_1\gamma G(1) + t_{P1}H] + \\ & p_1^{t_{P1}}(1-\gamma)\{[G(2)+(t_{P1}+1)H](1-\tau) + [G(1)+(t_{P1}+1)H]\tau\gamma\} + \\ & p_1^{t_{P1}}(1-\gamma)^2\tau\{[G(2)+(t_{P1}+2)H](1-\tau) + [G(1)+(t_{P1}+2)H]\tau\gamma\} + \cdots + \\ & p_1^{t_{P1}}(1-\gamma)^{t_{P2}-t_{P1}}\tau^{t_{P2}-t_{P1}-1}[(1-\tau)G(2) + \tau G(1) + t_{P2}H] \\ =& G(2)(1-p_1)\sum_{i=1}^{t_{P1}} p_1^{i-1} + G(2)p_1^{t_{P1}}(1-\tau)(1-\gamma)\sum_{i=1}^{t_{P2}-t_{P1}}[(1-\gamma)\tau]^{i-1} + \\ & G(1)p_1^{t_{P1}}\gamma + G(1)p_1^{t_{P1}}\gamma\sum_{i=1}^{t_{P2}-t_{P1}-1}[(1-\gamma)\tau]^i + G(1)p_1^{t_{P1}}(1-\gamma)^{t_{P2}-t_{P1}}\tau^{t_{P2}-t_{P1}} + \\ & H(1-p_1)\sum_{i=1}^{t_{P1}} i\cdot p_1^{i-1} + Ht_{P1}p_1^{t_{P1}}\gamma + \\ & Hp_1^{t_{P1}}(1-\tau)(1-\gamma)\sum_{i=1}^{t_{P2}-t_{P1}}[(1-\gamma)\tau]^{i-1}(t_{P1}+i) + Ht_{P2}(1-\gamma)p_1^{t_{P1}}[(1-\gamma)\tau]^{t_{P2}-t_{P1}} \\ =& G(2)\left[(1-p_1)\sum_{i=1}^{t_{P1}} p_1^{i-1} + p_1^{t_{P1}}(1-\tau)(1-\gamma)\sum_{i=1}^{t_{P2}-t_{P1}}\lambda^{i-1}\right] + \\ & G(1)\left[p_1^{t_{P1}}\gamma + p_1^{t_{P1}}\gamma\sum_{i=1}^{t_{P2}-t_{P1}-1}\lambda^i + p_1^{t_{P1}}(1-\gamma)^{t_{P2}-t_{P1}}\tau^{t_{P2}-t_{P1}}\right] + \\ & H(1-p_1)\sum_{i=1}^{t_{P1}} i\cdot p_1^{i-1} + Ht_{P1}p_1^{t_{P1}}\gamma + Hp_1^{t_{P1}}(1-\tau)(1-\gamma)\sum_{i=1}^{t_{P2}-t_{P1}}\lambda^{i-1}(t_{P1}+i) +\end{aligned}$$

$$
\begin{aligned}
& Ht_{P2}(1-\gamma)p_1^{t_{P1}}\lambda^{t_{P2}-t_{P1}} \\
=& G(1)\left[p_1^{t_{P1}}\lambda^{t_{P2}-t_{P1}}+\frac{p_1^{t_{P1}}\gamma(1-\lambda^{t_{P2}-t_{P1}})}{1-\lambda}\right]+ \\
& G(2)\left[1-p_1^{t_{P1}}\lambda^{t_{P2}-t_{P1}}-\frac{p_1^{t_{P1}}\gamma(1-\lambda^{t_{P2}-t_{P1}})}{1-\lambda}\right]+ \\
& H\left[\frac{p_1-p_1^{t_{P1}}}{1-p_1}+\frac{p_1^{t_{P1}}(1-\gamma)(1-\lambda^{t_{P2}-t_{P1}})}{1-\lambda}\right],
\end{aligned}
$$

where

$$
\begin{aligned}
\gamma &:= \tilde{Q}(c_t=1,\hat{c}_{t+1}=1)=p_1q_1+(1-p_1)(1-q_2), \\
\tau &:= \pi(c_{t+1}=1\mid c_t=1,\hat{c}_{t+1}=2)=\frac{p_1(1-q_1)}{p_1(1-q_1)+(1-p_1)q_2}, \text{ and} \\
\lambda &:= \tilde{Q}(c_t=1,\hat{c}_{t+1}=2)\pi(c_{t+1}=1\mid c_t=1,\hat{c}_{t+1}=2)=p_1(1-q_1).
\end{aligned}
$$

For the first equality in the expected total cost, each term on the right-hand side represents the expected total cost when the treatment is stopped at epoch t multiplied by the probability that the treatment is stopped at epoch t given the initial condition. For example, the expression in the second line, $p_1^{t_{P1}-1}[(1-p_1)G(2)+p_1\gamma G(1)+t_{P1}H]$, corresponds to the proportion of the expected total cost when the treatment is stopped at epoch t_{P1}. If the treatment is stopped at epoch t_{P1}, the patient must have stayed in class 1 in the previous $t_{P1}-1$ epochs; otherwise, the treatment should have been stopped before epoch t_{P1}. The probability that the patient stays in class 1 in the previous $t_{P1}-1$ epochs is $p_1^{t_{P1}-1}$. Based on the threshold policy, if the patient transits to class 2, the terminal cost is $G(2)$, with a probability of $1-p_1$; otherwise, if the patient continues to stay in class 1 at epoch t_{P1}, the treatment will be stopped only if the predicted class is $\hat{c}_{t+1}=1$, and the terminal cost is $G(1)$. The probability for this case is $p_1\gamma$. The remaining part is the treatment cost, which is Ht_{P1} in both situations above. The rest follows from algebraic manipulation, and we also need the result in Equation (EC.7).

From Equation (EC.6), we have $t_B^*=1$ if $H\geq(1-p_1)\Delta G$, and $t_B^*=T$ otherwise. This is intuitive as when the treatment cost is higher than the expected potential benefit of the treatment, i.e., $H\geq(1-p_1)\Delta G$, it is optimal to stop the treatment. On the other hand, if the treatment cost is low, it would be better to wait for the patient to evolve into the better class until the last epoch.

To derive the optimal policy under perfect information (OP-PP), we let $q_1=q_2=1$, i.e., $Q=I$. In this case, $\gamma=p_1$, $\tau=0$, and $\lambda=0$. The cost function for the MDP-P model becomes

$$
\begin{aligned}
J_1^P(1;t_{P1},t_{P2},Q=I) &= G(1)p_1^{t_{P1}+1}+G(2)(1-p_1^{t_{P1}+1})+H\left[\frac{p_1-p_1^{t_{P1}}}{1-p_1}+p_1^{t_{P1}}(1-p_1)\right] \\
&= \left(\Delta G-H\frac{2-p_1}{1-p_1}\right)p_1^{t_{P1}+1}+G(2)+\frac{Hp_1}{1-p_1}.
\end{aligned}
$$

From the above equation, we have $t^*_{P1}=1$ if $H/\Delta G \geq (1-p_1)/(2-p_1)$, and $t^*_{P1}=T$ otherwise. For t^*_{P2}, we have $t^*_{P2}=1$ if $H \leq \Delta G$, since it is beneficial to continue the treatment for class 1 patients who are predicted to evolve to class 2 when $H \leq \Delta G$, and $t^*_{P2}=T$ otherwise.

To summarize, we can partition $H/\Delta G$ into 4 regions as depicted in Figure EC.1. We can observe that when the ratio $H/\Delta G \in [0,(1-p_1)/(2-p_1)] \cup [1,+\infty)$, OP-B is the same as OP-PP. These regions represent the extreme cases that predictive information is "useless" as the treatment cost is either too high to continue the treatment or too low to consider stopping the treatment before the patient evolves to the better class. Therefore, we only need to consider the other two regions where OP-PP is different from OP-B.

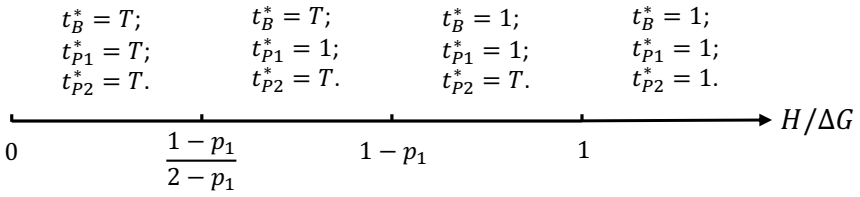


Figure EC.1 The optimal policies under different parameters.

Case 1: $H/\Delta G \in [1-p_1, 1]$.

In this case, we have $t^*_B = t^*_{P1} = 1$ and $t^*_{P2} = T$. Thus,

$$\begin{aligned} J_1^B(1;t^*_B) - J_1^P(1;t^*_{P1},t^*_{P2}) =& \Delta G \frac{p_1(1-\lambda-\gamma)(1-\lambda^{T-1})}{1-\lambda} - H\frac{p_1(1-\gamma)(1-\lambda^{T-1})}{1-\lambda} \\ =& p_1(1-p_1)q_2(\Delta G - H)\frac{1-\lambda^{T-1}}{1-\lambda} - Hp_1\frac{\lambda-\lambda^T}{1-\lambda}. \end{aligned}$$

Note that we are evaluating the performance of OP-PP in the environment where the prediction is actually imperfect. Therefore, the cost of OP-PP is computed using the cost function for the MDP-P model with a general Q. From the above equation, we can find that the gap between the two value function is linearly increasing in q_2. Moreover, it is concavely increasing in q_1 which can be validated by taking derivatives as follows:

$$\begin{aligned} d[J_1^B(1;t^*_B) - J_1^P(1;t^*_{P1},t^*_{P2})]/dq_2 =& -p_1\Big[p_1(1-p_1)q_2\Delta G\Lambda - p_1(1-p_1)q_2H\Lambda - \\ & p_1H\frac{1-\lambda^{T-1}}{1-\lambda} - p_1\lambda H\Lambda\Big] \\ \geq& -p_1^2H\left[p_1(q_1+q_2-1)\Lambda - \frac{1-\lambda^{T-1}}{1-\lambda}\right] \\ \geq& -p_1^2H\left[p_1q_1\Lambda - \frac{1-\lambda^{T-1}}{1-\lambda}\right], \end{aligned}$$

where

$$\Lambda = \frac{d\frac{1-\lambda^{T-1}}{1-\lambda}}{d\lambda} = \frac{(T-2)p_1^{T-1} - (T-1)p^{T-2}(1-q_1)^{T-2} + 1}{\left[1-p_1(1-q_1)\right]^2}.$$

We have

$$\begin{aligned}
p_1q_1\Lambda - \frac{1-\lambda^{T-1}}{1-\lambda} =& \frac{(T-2)p_1^T q_1 - (T-1)p^{T-1}(1-q_1)^{T-2}q_1 + p_1q_1}{\left[1-p_1(1-q_1)\right]^2} - \frac{1-p_1^{T-1}(1-q_1)^{T-1}}{1-p_1(1-q_1)} \\
\leq& \frac{[(T-2)p_1(1-q_1) - (T-1) - p_1(1-q_1)^2 + (1-q_1)]p_1^{T-1}(1-q_1)^{T-2}}{\left[1-p_1(1-q_1)\right]^2} \\
\leq& \frac{[(T-2)(1-q_1) - (T-1) - p_1(1-q_1)^2 + (1-q_1)]p_1^{T-1}(1-q_1)^{T-2}}{\left[1-p_1(1-q_1)\right]^2} \\
\leq& \frac{[-q_1(T-1) - p_1(1-q_1)^2]p_1^{T-1}(1-q_1)^{T-2}}{\left[1-p_1(1-q_1)\right]^2} \\
\leq& 0
\end{aligned}$$

Thus, the gap is increasing in q_1. The second order derivative is

$$p_1(1-p_1)q_2(\Delta G - H)\frac{d\left(\frac{1-\lambda^{T-1}}{1-\lambda}\right)^2}{d^2q_1} - Hp_1\frac{d\left(\frac{\lambda-\lambda^T}{1-\lambda}\right)^2}{d^2q_1} \leq -Hp_1\frac{d\left(\frac{\lambda-\lambda^T}{1-\lambda}\right)^2}{d^2q_1} \leq 0.$$

With a simple algebra, we can derive that the above cost difference is nonnegative if and only if

$$\frac{1-\lambda-\gamma}{1-\gamma} \geq \frac{H}{\Delta G}.$$

Therefore, in the environment of imperfect prediction, the cost of using OP-PP is lower than the optimal cost without using predictive information if only if the following condition holds:

$$\frac{1-\lambda-\gamma}{1-\gamma} \geq \frac{H}{\Delta G} \geq 1-p_1.$$

Here, the interval for the cost ratio is nonempty if and only if $q_1+q_2 \geq 1$, which can be derived immediately after plugging the expressions for λ and γ.

Case 2: $H/\Delta G \in [(1-p_1)/(2-p_1), 1-p_1]$.

In this case, we have $t_{P1}^* = 1$ and $t_B^* = t_{P2}^* = T$. Thus,

$$\begin{aligned}
J_1^B(1;t_B^*) - J_1^P(1;t_{P1}^*,t_{P2}^*) =& \Delta G\left[p_1^{t_B^*} - p_1^{t_{P1}^*}\lambda^{t_{P2}^*-t_{P1}^*} - \gamma\frac{p_1^{t_{P1}^*} - p_1^{t_{P1}^*}\lambda^{t_{P2}^*-t_{P1}^*}}{1-\lambda}\right] + \\
& H\left[\frac{p_1^{t_{P1}^*} - p_1^{t_B}}{1-p_1} - p_1^{t_{P1}^*}(1-\gamma)\frac{1-\lambda^{t_{P2}^*-t_{P1}^*}}{1-\lambda}\right] \\
=& \Delta G\left[p_1^T - p_1\lambda^{T-1} - \gamma\frac{p_1 - p_1\lambda^{T-1}}{1-\lambda}\right] + \\
& H\left[\frac{p_1-p_1^T}{1-p_1} - p_1(1-\gamma)\frac{1-\lambda^{T-1}}{1-\lambda}\right] \\
=& \Delta G\frac{(p_1^T - p_1\lambda^{T-1})(1-\lambda) - p_1\gamma(1-\lambda^{T-1})}{1-\lambda} + \\
& H\frac{(p_1-p_1^T)(1-\lambda) - p_1(1-p_1)(1-\gamma)(1-\lambda^{T-1})}{(1-p_1)(1-\lambda)}.
\end{aligned}$$

We can also easily verify that the gap is linear in q_2 with the same coefficient, and is concavely increasing in q_1 by taking derivatives by tedious algebra and calculus. The above cost difference is nonnegative if and only if

$$\frac{H}{\Delta G} \geq \frac{(1-p_1)\gamma(1-\lambda^{T-1}) - (1-p_1)(p_1^{T-1}-\lambda^{T-1})(1-\lambda)}{(1-p_1^{T-1})(1-\lambda) - (1-p_1)(1-\gamma)(1-\lambda^{T-1})}.$$

In this case, the cost of using OP-PP is lower than the optimal cost without using predictive information if only if the following condition holds:

$$1-p_1 \geq \frac{H}{\Delta G} \geq \frac{(1-p_1)\gamma(1-\lambda^{T-1}) - (1-p_1)(p_1^{T-1}-\lambda^{T-1})(1-\lambda)}{(1-p_1^{T-1})(1-\lambda) - (1-p_1)(1-\gamma)(1-\lambda^{T-1})}.$$

Next, we investigate when the above interval for the cost ratio is nonempty. Define the right-hand side of the above inequality as f. We can simplify f as follows:

$$\begin{aligned}
f :=& \frac{(1-p_1)\gamma(1-\lambda^{T-1}) - (1-p_1)(p_1^{T-1}-\lambda^{T-1})(1-\lambda)}{(1-p_1^{T-1})(1-\lambda) - (1-p_1)(1-\gamma)(1-\lambda^{T-1})} \\
=& (1-p_1)\frac{(1-\lambda-\gamma)\lambda^{T-1} - (1-\lambda)p_1^{T-1} + \gamma}{(1-p_1)(1-\gamma)\lambda^{T-1} - (1-\lambda)p_1^{T-1} + (1-\lambda) - (1-p_1)(1-\gamma)} \\
=& (1-p_1)\frac{(1-p_1)(1-\gamma)\lambda^{T-1} - (1-\lambda)p_1^{T-1} + (1-\lambda)p_1^{T-1} + (1-\lambda) - (1-p_1)(1-\gamma)}{(1-p_1)(1-\gamma)\lambda^{T-1} - (1-\lambda)p_1^{T-1} + (1-\lambda) - (1-p_1)(1-\gamma)} + \\
& (1-p_1)\frac{p_1(1-p_1)(1-q_1-q_2)(1-\lambda^{T-1})}{(1-p_1)(1-\gamma)\lambda^{T-1} - (1-\lambda)p_1^{T-1} + (1-\lambda) - (1-p_1)(1-\gamma)} \\
=& (1-p_1) + \frac{p_1(1-p_1)^2(1-q_1-q_2)(1-\lambda^{T-1})}{(1-p_1)(1-\gamma)\lambda^{T-1} - (1-\lambda)p_1^{T-1} + (1-\lambda) - (1-p_1)(1-\gamma)} \\
=& (1-p_1) + \frac{p_1(1-p_1)^2(1-q_1-q_2)(1-\lambda^{T-1})}{(1-p_1)(1-\gamma)(\lambda^{T-1}-1) + (1-\lambda)(1-p_1^{T-1})} \\
=& (1-p_1) + \frac{p_1(1-p_1)^2(1-q_1-q_2)}{(1-\lambda)\frac{1-p_1^{T-1}}{1-\lambda^{T-1}} - (1-p_1)(1-\gamma)}.
\end{aligned}$$

Note that the denominator of the second term in the last equation above is nonnegative, i.e.,

$$\begin{aligned}
(1-\lambda)\frac{1-p_1^{T-1}}{1-\lambda^{T-1}} - (1-p_1)(1-\gamma) &\geq (1-\lambda) - (1-p_1)(1-\gamma) \\
&= (1-q_1-q_2)(1-p_1)^2 + q_1 \\
&\geq (1-q_1-q_2)(1-p_1)^2 + q_1(1-p_1)^2 \\
&\geq (1-q_2)(1-p_1)^2 \\
&\geq 0.
\end{aligned}$$

Therefore, having a nonempty interval for the cost ratio is equivalent to requiring

$$p_1(1-p_1)^2(1-q_1-q_2) \le 0,$$

which is again equivalent to $q_1+q_2 \ge 1$.

Combining these two cases, we can conclude that in the environment of imperfect prediction, the cost of using OP-PP is lower than the optimal cost without using predictive information if only if

$$\frac{1-\lambda-\gamma}{1-\gamma} \ge \frac{H}{\Delta G} \ge \frac{(1-p_1)\gamma(1-\lambda^{T-1})-(1-p_1)(p_1^{T-1}-\lambda^{T-1})(1-\lambda)}{(1-p_1^{T-1})(1-\lambda)-(1-p_1)(1-\gamma)(1-\lambda^{T-1})},$$

where the interval is nonempty if and only if $q_1+q_2 \ge 1$. The proof is completed.

EC.3. Discounting Future Costs

In this section, we consider that the DM is more concerned with the immediate cost. Hence a discount factor $\alpha \in [0,1]$ is incorporated into our models. Then the value functions of the two settings are modified as follows:

$$J_t^B(c_t) = \min\Big\{G(c_t),\ H(c_t) + \alpha \cdot \sum_{c_{t+1}\in\mathcal{C}} P(c_t, c_{t+1}) \cdot J_{t+1}^B(c_{t+1})\Big\}, \forall c_t \in \mathcal{C}, t \in \mathcal{T}\backslash\{T\},$$

and

$$J_t^P(c_t, c_{t+1}) = \min\Big\{G(c_t),\ H(c_t) + \alpha \cdot \sum_{c_{t+2}\in\mathcal{C}} P(c_{t+1}, c_{t+2}) \cdot J_{t+1}^P(c_{t+1}, c_{t+2})\Big\},$$
$$\forall t \in \mathcal{T}\backslash\{T\}, c_t, c_{t+1} \in \mathcal{C}.$$

We show next that the theoretical results in Section 4 hold with the discount factor α. Since most of proofs are similar to those in Section EC.2, we only provide proofs that is significantly different from the previous ones.

EC.3.1. Proof of Theorem EC.1 with Discounted Future Costs

To prove Theorem EC.1, we re-define the function (EC.4) as

$$h_t^B(c_t) := G(c_t) - H(c_t) - \alpha \cdot \sum_{c_{t+1}\in\mathcal{C}} P(c_t, c_{t+1}) \cdot J_{t+1}^B(c_{t+1}), \forall c_t \in \mathcal{C}, t \in \mathcal{T}\backslash\{T\}.$$

In decision epoch $T-1$, we have

$$\begin{aligned} h_{T-1}^B(c_{T-1}) =& G(c_{T-1}) - H(c_{T-1}) - \alpha \sum_{c_T\in\mathcal{C}} P(c_{T-1}, c_T) J_T^B(c_T) \\ =& G(c_{T-1}) - H(c_{T-1}) - \sum_{c_T\in\mathcal{C}} P(c_{T-1}, c_T) G(c_T) + \\ & (1-\alpha) \sum_{c_T\in\mathcal{C}} P(c_{T-1}, c_T) J_T^B(c_T). \end{aligned}$$

The last term above is nonincreasing in c_{T-1} from Lemma EC.1 and Assumption EC.1. Together with Assumption EC.2, $h_{T-1}^B(c_{T-1})$ is nonincreasing in c_{T-1}. Assume that in epoch $t+1$, $h_{t+1}^B(c_{t+1})$ is nonincreasing in c_{t+1}. In epoch t, we have

$$\begin{aligned} h_t^B(c_t) =& G(c_t) - H(c_t) - \alpha \sum_{c_{t+1}\in\mathcal{C}} P(c_t, c_{t+1}) J_{t+1}^B(c_{t+1}) \\ =& G(c_t) - H(c_t) - \sum_{c_{t+1}\in\mathcal{C}} P(c_t, c_{t+1}) G(c_{t+1}) + \alpha \sum_{c_{t+1}<c_{t+1}^*} P(c_t, c_{t+1}) h_{t+1}^B(c_{t+1}) + \\ & (1-\alpha) \sum_{c_{t+1}\in\mathcal{C}} P(c_t, c_{t+1}) G(c_{t+1}). \end{aligned}$$

The second equality follows from the induction hypothesis. $\sum_{c_{t+1}<c_{t+1}^*} P(c_t, c_{t+1}) h_{t+1}^B(c_{t+1})$ is nonincreasing in c_t from Assumption EC.1 and the induction hypothesis, and $G(c_t) - H(c_t) -$

$\sum_{c_{t+1}\in\mathcal{C}} P(c_t,c_{t+1})G(c_{t+1})$ is nonincreasing in c_t from Assumption EC.2. $\sum_{c_{t+1}\in\mathcal{C}} P(c_t,c_{t+1})G(c_{t+1})$ is nonincreasing in c_t from Assumption EC.1. Hence, $h_t^B(c_t)$ is nonincreasing in c_t. We have

$$\begin{aligned} h_t^B(c) &= G(c) - H - \alpha \sum_{c'\in\mathcal{C}} P(c,c') J_{t+1}^B(c') \\ &\geq G(c) - H - \alpha \sum_{c'\in\mathcal{C}} P(c,c') J_{t+2}^B(c') \\ &= h_{t+1}^B(c). \end{aligned}$$

The inequality follows from Claim EC.1. Therefore, we complete the proof for Theorem EC.1 with a discount factor.

EC.3.2. Proof of Proposition EC.1 with Discounted Future Costs

After introducing the discount factor α, the stationary threshold c^* in Proposition EC.1 becomes

$$c^* = \min\Big\{c\in\mathcal{C} : G(c) - H\alpha \sum_{c'\in\mathcal{C}} P(c,c')G(c') \leq 0\Big\}.$$

To prove this proposition, we need to show that $h_t^B(c_t)\leq 0$ if $c_t\geq c^*$, and $h_t^B(c_t)>0$ otherwise for $t\in\mathcal{T}\backslash\{T\}$. The proof proceeds by induction on decision epoch t.

When $t=T-1$, we have

$$\begin{aligned} h_{T-1}^B(c_{T-1}) &= G(c_{T-1}) - H(c_{T-1}) - \alpha \sum_{c_T\in\mathcal{C}} P(c_{T-1},c_T) J_T^B(c_T) \\ &= G(c_{T-1}) - H(c_{T-1}) - \alpha \sum_{c_T\in\mathcal{C}} P(c_{T-1},c_T) G(c_T). \end{aligned}$$

The equalities follow from definitions and the boundary condition. It is obvious that $h_{T-1}^B(c_{T-1})\leq 0$ if $c_{T-1}\geq c^*$, and $h_{T-1}^B(c_{T-1})>0$ otherwise from the definition of c^*. Assume that the property holds in epoch $t+1$. In epoch t, we have

$$\begin{aligned} h_t^B(c_t) =& G(c_t) - H(c_t)\alpha \sum_{c_{t+1}\in\mathcal{C}} P(c_t,c_{t+1}) J_{t+1}^B(c_{t+1}) \\ =& G(c_t) - H(c_t) - \alpha \sum_{c_{t+1}\in\mathcal{C}} P(c_t,c_{t+1}) G(c_{t+1}) - \sum_{c_{t+1}<c^*} P(c_t,c_{t+1}) h_{t+1}^B(c_{t+1}) \\ =& G(c_t) - H(c_t) - \sum_{c_{t+1}\in\mathcal{C}} P(c_t,c_{t+1}) G(c_{t+1}) + \alpha \sum_{c_{t+1}<c^*} P(c_t,c_{t+1}) h_{t+1}^B(c_{t+1}) + \\ & (1-\alpha) \sum_{c_{t+1}\in\mathcal{C}} P(c_t,c_{t+1}) G(c_{t+1}). \end{aligned}$$

The second equality follows from the induction hypothesis. If $c_t\geq c^*$, then $G(c_t) - [H(c_t) + \sum_{c_{t+1}\in\mathcal{C}} P(c_t,c_{t+1})G(c_{t+1})] \leq 0$. $\sum_{c_{t+1}<c^*} P(c_t,c_{t+1}) h_{t+1}^B(c_{t+1}) = 0$ since $P(c_t,c_{t+1})=0$ when $c_{t+1}<c_t$. Hence, $h_t^B(c_t)\leq 0$ when $c_t\geq c^*$. Otherwise, if $c_t<c^*$, $G(c_t) - [H(c_t) + \sum_{c_{t+1}\in\mathcal{C}} P(c_t,c_{t+1})G(c_{t+1})] > 0$ and $\sum_{c_{t+1}<c^*} P(c_t,c_{t+1}) h_{t+1}^B(c_{t+1}) \geq 0$ from the induction hypothesis. Thus $h_t^B(c_t)>0$ when $c_t<c^*$. The proof is completed.

EC.3.3. Proof of Theorem 1 with Discounted Future Costs

We re-define that

$$h_t^P(c_t,\hat{c}_{t+1}) := G(c_t) - H(c_t) - \alpha \sum_{c_{t+1}\in\mathcal{C}} \left[\pi(c_{t+1} \mid c_t, \hat{c}_{t+1}) \sum_{\hat{c}_{t+2}\in\mathcal{C}} \tilde{Q}(c_{t+1},\hat{c}_{t+2}) J_{t+1}^P(c_{t+1},\pi_{t+1}) \right], \quad t\in\mathcal{T}\backslash\{T\}, c_t, \hat{c}_{t+1}\in\mathcal{C}.$$

First, $h_t^P(c_t,\hat{c}_{t+1})$ is nonincreasing in c_t following a similar argument to the proof in Section EC.3.1. From Section EC.2.4, $\sum_{c_{t+1}\in\mathcal{C}} \pi(c_{t+1} \mid c_t, \hat{c}_{t+1}) \sum_{\hat{c}_{t+2}\in\mathcal{C}} \tilde{Q}(c_{t+1},\hat{c}_{t+2}) J_{t+1}^P(c_{t+1},\hat{c}_{t+2})$ is nonincreasing in $\hat{c}_{t+1}$. Hence, $h_t^P(c_t,\hat{c}_{t+1})$ is nondecreasing in $\hat{c}_{t+1}$.

We next prove that $h_t^P(c,\hat{c})$ is nonincreasing in t as follows.

$$\begin{aligned} h_t^P(c,\hat{c}) &= G(c) - H - \alpha \sum_{c'\in\mathcal{C}} \left[\pi(c' \mid c,\hat{c}) \sum_{\hat{c}'\in\mathcal{C}} \tilde{Q}(c',\hat{c}') J_{t+1}^P(c',\hat{c}') \right] \\ &\geq G(c) - H - \alpha \sum_{c'\in\mathcal{C}} \left[\pi(c' \mid c,\hat{c}) \sum_{\hat{c}'\in\mathcal{C}} \tilde{Q}(c',\hat{c}') J_{t+2}^P(c',\hat{c}') \right] \\ &= h_{t+1}^P(c,\hat{c}). \end{aligned}$$

The inequality follows from Claim EC.3. Thus, the proof is completed.

EC.4. POMDP Formulation of the Case with Predictive Information

In this part, we develop an alternative formulation for the stopping problem with predictive information using the POMDP framework. We refer to this as the POMDP-P model.

Core states. The core state is the underlying state of the process, which is the same with MDP-P and can be described as (c_t, c_{t+1}) in epoch t. However, the core state in POMDP-P cannot be completely observed by the DM.

Observations. Since the predictive information may contain errors, the DM can only partially observe the core state (c_t, c_{t+1}) in epoch t based on the signal revealed by a predictive model. Note that the current patient class c_t can be fully observed. Meanwhile, the DM can receive a predicted class $\hat{c}_{t+1}$ on c_{t+1} from the predictive model. Thus the observation in epoch t is $(c_t, \hat{c}_{t+1})$, and the observation space is $\mathcal{C} \times \mathcal{C}$.

Information matrix. In POMDP models, the information matrix characterizes the quality of observations, which is exactly the misclassification matrix, Q, in our MDP-P model. Recall that each component of the misclassification matrix $Q(i,j) = Pr(\hat{c}_t = j \mid c_t = i)$ represents the probability that a class i patient is predicted as class j.

Belief states. The belief state is a vector that defines the probability of the patient in each class. Since the current class c_t is fully observable, the belief state $\boldsymbol{\pi}_t$ is a vector that consists of the probabilities of each class that the patient will be in epoch $t+1$, given the observation $(c_t, \hat{c}_{t+1})$. The belief space is a $C-1$ dimensional standard simplex, denoted by Π.

The optimality equation of the POMDP-P model can be written as

$$J_t^{IP}(c_t, \boldsymbol{\pi}_t) = \min \left\{ G(c_t),\ H(c_t) + \sum_{c_{t+1} \in \mathcal{C}} \left[\pi_t(c_{t+1}) \sum_{\hat{c}_{t+2} \in \mathcal{C}} \tilde{Q}(c_{t+1}, \hat{c}_{t+2}) J_{t+1}^{IP}(c_{t+1}, \boldsymbol{\pi}_{t+1}) \right] \right\},$$
$$t \in \mathcal{T} \backslash \{T\}, c_t \in \mathcal{C}, \boldsymbol{\pi}_t \in \Pi,$$

where

$$\tilde{Q}(c_{t+1}, \hat{c}_{t+2}) = \sum_{c_{t+2} \in \mathcal{C}} P(c_{t+1}, c_{t+2}) Q(c_{t+2}, \hat{c}_{t+2}),$$

and $J_t^{IP}(c_t, \boldsymbol{\pi_t})$ represents the minimum expected total cost incurred to a class c_t patient at epoch t when his or her health condition in the next epoch follows the belief $\boldsymbol{\pi}_t$. The boundary value is

$$J_T^{IP}(c_T, \boldsymbol{\pi}_T) = G(c_T),\ c_T \in \mathcal{C}.$$

The belief is updated according to the Bayes' rule as follows:

$$\pi_{t+1}(c_{t+2} \mid c_{t+1}, \hat{c}_{t+2}) = \frac{P(c_{t+1}, c_{t+2}) Q(c_{t+2}, \hat{c}_{t+2})}{\sum_{c \in \mathcal{C}} P(c_{t+1}, c) Q(c, \hat{c}_{t+2})}.$$

Based on the POMDP formulation, we have the following results on the optimal policy. The proofs are similar to those for the MDP-P model, and we omit them from here.

THEOREM EC.2. *Under Assumptions 1–3, for the POMDP-P model, there exists an optimal threshold curve* $\Gamma_t(c_t)$ *for each* c_t *in epoch* $t \in \mathcal{T}\backslash\{T\}$ *that partitions the belief space* Π *into two connected regions* $\mathcal{B}_t^0(c_t)$ *and* $\mathcal{B}_t^1(c_t)$*, such that the optimal action is*

$$a_t^*(c_t, \boldsymbol{\pi}_t) = \begin{cases} 0 & if\ \boldsymbol{\pi}_t \in \mathcal{B}_t^0(c_t), \\ 1 & if\ \boldsymbol{\pi}_t \in \mathcal{B}_t^1(c_t). \end{cases}$$

Moreover, the optimal policy has the following properties:

(1) For $c_t, c_t' \in \mathcal{C}$ *and* $c_t > c_t'$, $t \in \mathcal{C}\backslash\{T\}$ *we have* $\mathcal{B}_t^1(c_t) \supseteq \mathcal{B}_t^1(c_t')$.

(2) For $c \in \mathcal{C}$ *and* $t = 1, 2, ..., T-2$, $\mathcal{B}_t^1(c) \subseteq \mathcal{B}_{t+1}^1(c)$.

Theorem EC.2 asserts that the optimal policy for the described problem is a threshold policy. Similar to classical POMDP results, the threshold curve lies in the belief space, which is not easily interpretable. Moreover, the curve needs to be numerically computed, and the belief at every epoch also needs to be computed to compare against the curve to derive the optimal decision. To address this issue, in the following theorem we map the threshold in the optimal policy in the belief space into a threshold curve in the observation space and recover the same result for the MDP-P model.

THEOREM EC.3. *Under Assumptions 1–3, for the POMDP-P model, there exists an optimal threshold curve* Γ_t *in epoch* $t \in \mathcal{T}\backslash\{T\}$ *that partitions observation space* $\mathcal{C} \times \mathcal{C}$ *into two subsets* $\mathcal{O}_t^0$ *and* $\mathcal{O}_t^1$*, such that the optimal action is*

$$a_t^*(c_t, \hat{c}_{t+1}) = \begin{cases} 0 & if\ (c_t, \hat{c}_{t+1}) \in \mathcal{O}_t^0, \\ 1 & if\ (c_t, \hat{c}_{t+1}) \in \mathcal{O}_t^1. \end{cases}$$

The optimal threshold curve has the following properties:

(1) $a_t^*(c_t, \hat{c}_{t+1})$ *is increasing in* c_t *and decreasing in* $\hat{c}_{t+1}$.

(2) $\mathcal{O}_t^1 \subseteq \mathcal{O}_{t+1}^1$ *for* $t = 1, 2, ..., T-2$.

Theorem EC.3 is equivalent to Theorem EC.2 analytically in describing the structure of the optimal policy, but it is easier to implement since the policy in $\mathcal{C} \times \Pi$ space is projected into $\mathcal{C} \times \mathcal{C}$ space, which is much more interpretable and allows the easy design of decision protocols based on the predictions rather than computed beliefs.

Compared with the MDP-P model formulation, Theorem EC.3 is the same as Theorem 1, and all insights remain unchanged. In particular, if a patient's condition at the current period is in a good state, it is more likely that the treatment will be stopped; if the patient's condition at the next period is predicted to be worse, it is also more likely that the treatment will be stopped. Also, the likelihood of stopping treatment increases with the duration of treatment.

EC.5. Sample Selection

Table EC.1 Data selection process.

Sample	Observations	% prior	% initial
All patients receive MV	4,026	-	100%
Excluding 30-day ICU readmissions	3,378	83.9%	83.9%
Excluding the cases with over 30 days of ICU LOS	3,341	98.9%	83.0%
Excluding the cases ventilated over 7 days	3,309	99.0%	82.2%
Excluding the cases with simultaneous intubation and extubation records	3,296	99.6%	81.9%
Excluding the cases with NIV record during MV	3,282	99.6%	81.5%
Excluding the cases died within 7 days after extubation	3,147	95.9%	78.2%
Excluding the cases discharged within 6 hours after extubation	3,129	99.4%	77.7%
Excluding the cases without RSBI record	3,118	99.6%	77.4%
Excluding unplanned extubations	3,067	98.4%	76.2%

ICU: intensive care unit; LOS: length of stay; MV: mechanical ventilation; NIV: noninvasive ventilation; RSBI: rapid shallow breathing index.

EC.6. Variables Definition

Table EC.2 Definition of variables in the case study.

Variable	Explanation
GCS	A neurological score that captures the conscious state
Creatinine	Byproduct of muscle metabolism that is excreted unchanged by the kidneys
Haematocrit	Volume percentage of red blood cells in blood
Haemoglobin	Iron-containing oxygen-transport metalloprotein in red blood cells
Potassium	Amount of potassium in blood
Sodium	Amount of sodium in blood
Platelets	Amount of platelets in blood
WBC	Amount of white blood cell in blood
Systolic BP	Maximum pressure in blood vessels during heartbeats
Diastolic BP	Minimum pressure in blood vessels between two heartbeats
Arterial pO2	Partial pressure of oxygen in arterial blood
Arterial pCO2	Partial pressure of carbon oxygen in arterial blood
Arterial pH	Amount of hydrogen ions in arterial blood
Arterial SaO2	Percentage of hemoglobin binding sites occupied by oxygen
FiO2	Fraction of inspired oxygen
Tidal volume	Volume of air displaced between inhalation and exhalation
Cardiac rhythm	
Paced	Atrial paced rhythm and ventricular paced rhythm
Sinus rhythm	Normal sinus rhythm
Sinus tachycardia	Normal sinus tachycardia rhythm
Nursing assessment	
Cardiac	Pulse regular, rate 60-100 beats/min, skin warm and dry. Blood pressure less than 140/90 and no symptoms of hypotension
Respiratory	Respiratory rate 12-24 times/min. Bilateral breath sounds clear. Nail beds, mucous membranes pink and sputum clear
Vascular	Extremities are normal, pink and warm. Peripheral pulses palpable. Capillary refill within 3 seconds. No edema, numbness or tingling
Musculoskeletal	Independently able to move all extremities and perform functional activities as observed or stated
Neurological	Alert, oriented to person, place, time, and situation. Speech is coherent
Nutrition	No difficulty with chewing, swallowing or manual dexterity. Patient consuming over 50% of daily diet ordered as observed

HR: heart rate; RR: respiratory rate; GCS: Glasgow Coma Scale; WBC: white blood cell; BP: blood pressure.

EC.7. Sensitivity Analysis

EC.7.1. Adjusted Transition Matrix Estimation

As mentioned earlier, the extubation decision can also affect the estimation of the transition matrix. When the patients were extubated, we will not be able to observe their health transition if they were continued to be intubated. In other words, the estimated health transitions for patients under MV could be biased due to physicians' extubation decisions. We provide an alternative estimation procedure similar to the approach we adopted in terminal cost estimation to account for such selection as much as possible. In particular, we estimate the transition probabilities using the following equation:

$$\hat{P}(i,j) = \sum_{k=1}^{K_c} \frac{N_{ij}^k}{\sum_{l\in\mathcal{C}} N_{il}^k} \cdot \frac{N_i^k}{N_i},$$

where N_{ij}^k is the number of patients in class i and subgroup k that transit to class j. Note that $\sum_{l\in\mathcal{C}} N_{il}^k \neq N_i^k$, since N_i^k counts the extubated patients while N_{il}^k only counts those under ventilation. The idea is to use the propensity distribution to adjust the weight of transitions. The resultant transition matrix is as follows:

$$\hat{P}_{adjusted} = \begin{pmatrix} 0.35 & 0.18 & 0.20 & 0.13 & 0.14 \\ 0.08 & 0.39 & 0.31 & 0.10 & 0.12 \\ 0.01 & 0.13 & 0.49 & 0.23 & 0.14 \\ 0.01 & 0.05 & 0.23 & 0.38 & 0.33 \\ 0.00 & 0.02 & 0.06 & 0.18 & 0.74 \end{pmatrix}.$$

The performance of our models under the adjusted transition matrix is shown in Table EC.3. The results are very closed to those under the transition matrix estimated via the sample mean approach used in the main text.

Table EC.3 Performance of different extubation policies under the propensity-adjusted transition matrix.

Class	OP-B		OP-PP		OP-IP		OP-PP on MDP-P with $\hat{Q}$	
	LOS (hr)	EFR (%)	LOS (hr)	EFR (%)	LOS (hr)	EFR (%)	LOS (hr)	EFR (%)
1	105.2	21.2	104.0	18.9	104.7	18.9	104.7	18.9
2	104.2	29.6	101.5	23.4	102.6	22.8	102.6	22.8
3	97.1	21.0	96.2	20.0	97.0	20.2	97.0	20.2
4	94.6	17.7	93.4	16.4	94.4	16.9	94.4	16.9

OP-B: optimal extubation policy without predictive information; OP-PP: optimal extubation policy with perfect prediction; OP-IP: optimal extubation policy with imperfect prediction; LOS: length of stay; EFR: extubation failure rate.

EC.7.2. Alternative Patient Classification Schemes

In this section, we employ other alternative patient classification criteria to derive the results. The parameters are estimated following the same procedure described in Section 4.3.

EC.7.2.1. Six-class Stratification

In this section, we incorporate the cutoff value RSBI 80 from Chao and Scheinhorn (2007) and classify the patients in six classes $\{[105,+\infty), [80,105), [65,80), [45,65), [35,45), (0,35)\}$. Thus, $\mathcal{C}=\{1,2,3,4,5,6\}$. The terminal costs are shown in Table EC.4.

Table EC.4 Risk of extubation failure and RLOS for each patient class.

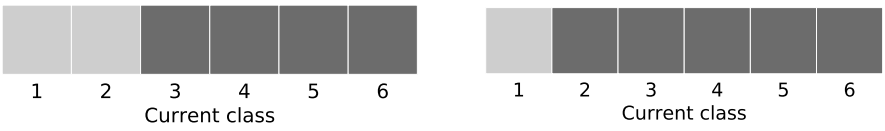
Terminal class	No. of observations	RLOS (hr)	Risk (%)
1	29	137.8	40.9
2	55	99.6	26.0
3	95	97.3	30.2
4	419	91.1	21.0
5	688	88.6	17.7
6	1,781	79.1	13.8

RLOS: remaining length of stay after extubation.

The transition probability matrix is estimated as:

$$P=\begin{pmatrix} 0.35 & 0.13 & 0.08 & 0.20 & 0.11 & 0.13 \\ 0.13 & 0.20 & 0.20 & 0.22 & 0.08 & 0.17 \\ 0.07 & 0.11 & 0.29 & 0.31 & 0.15 & 0.07 \\ 0.02 & 0.05 & 0.09 & 0.48 & 0.24 & 0.12 \\ 0.01 & 0.02 & 0.03 & 0.23 & 0.36 & 0.35 \\ 0.01 & 0.01 & 0.01 & 0.07 & 0.18 & 0.72 \end{pmatrix}.$$

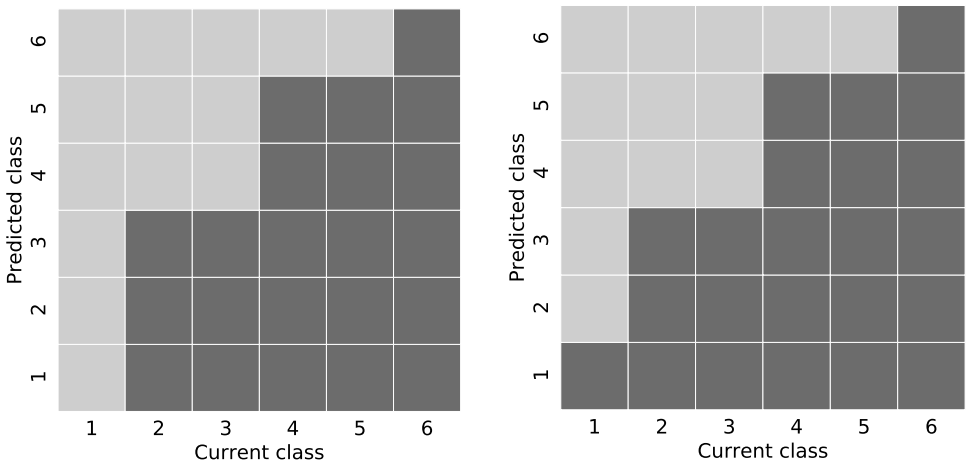
With these parameters, we derive the results under our models. The optimal policies are depicted in Figures EC.2 and EC.3. We find that classifying the patients into more classes enables more precise decisions. For example, under the basic policy, patients with RSBI in $[80,105)$ will not be extubated in the 5-class case but will be extubated in the 6-class case in the last two epochs.



(a) $t=1,2,\ldots,25$. (b) $t=26,27$.

Figure EC.2 Optimal extubation policy without predictive information. The dark gray grid corresponds to extubate and the light gray grid to continue ventilation.

The performance under different extubation policies is summarized in Table EC.5. With one more class, we can further reduce LOS and EFR using predictive information. However, the improvement is marginal. Moreover, the issues with the base model persist with finer patient classification: both LOS and EFR could get worse compared with the current practice if the model does not incorporate predictive information.



(a) $t = 1, 2, \ldots, 26$. (b) $t = 27$.

Figure EC.3 Optimal extubation policy with perfect predictive information. The dark gray grid corresponds to extubate and the light gray grid to continue ventilation.

Table EC.5 Extubation performance under different extubation policies.

Initial Class	Data		OP-B		OP-PP	
	LOS (hr)	EFR (%)	LOS (hr)	EFR (%)	LOS (hr)	EFR (%)
1	115.3	21.4	105.8	20.2	104.6	19.0
2	104.4	27.8	104.7	21.3	101.9	21.9
3	103.7	29.9	103.3	30.2	101.5	23.9
4	99.5	20.2	97.1	20.9	96.4	20.1
5	93.9	16.6	94.6	17.7	93.3	16.3
6	86.7	13.9	85.0	13.8	85.0	13.8

OP-B: optimal extubation policy without predictive information; OP-PP: optimal extubation policy with perfect prediction; LOS: length of stay; EFR: extubation failure rate.

EC.7.2.2. Seven-class Stratification

In this section, we further refine the classification and grouped the patients into seven RSBI classes $\{[105, +\infty), [80, 105), [65, 80), [55, 65), [45, 55), [35, 45), (0, 35)\}$. Thus, $\mathcal{C} = \{1, 2, 3, 4, 5, 6, 7\}$. The terminal costs are shown in Table EC.6.

The transition probability matrix is estimated as:

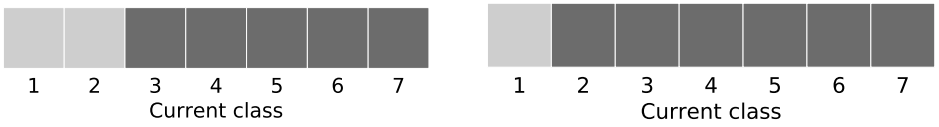
$$P = \begin{pmatrix} 0.35 & 0.13 & 0.08 & 0.07 & 0.13 & 0.11 & 0.13 \\ 0.13 & 0.20 & 0.20 & 0.13 & 0.09 & 0.08 & 0.17 \\ 0.07 & 0.11 & 0.29 & 0.18 & 0.13 & 0.15 & 0.07 \\ 0.03 & 0.08 & 0.14 & 0.26 & 0.29 & 0.10 & 0.10 \\ 0.01 & 0.04 & 0.08 & 0.14 & 0.32 & 0.29 & 0.12 \\ 0.01 & 0.02 & 0.03 & 0.06 & 0.17 & 0.36 & 0.35 \\ 0.01 & 0.01 & 0.01 & 0.02 & 0.05 & 0.18 & 0.72 \end{pmatrix}.$$

The optimal policies are shown in Figures EC.4 and EC.5, and their performance is summarized in Table EC.7. All the observations and insights are consistent with the results reported in the main text with five patient classes.

Table EC.6 Risk of extubation failure and RLOS for each patient class.

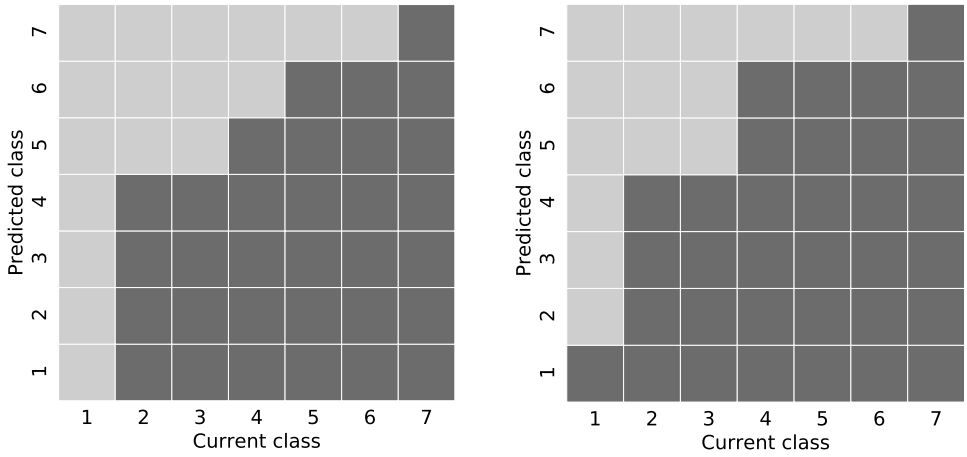
Terminal class	No. of observations	RLOS, hr	Risk, %
1	29	137.8	40.9
2	55	99.6	26.0
3	95	97.3	30.2
4	121	94.7	21.4
5	298	89.4	20.7
6	688	88.6	17.7
7	1,781	79.1	13.8

RLOS: remaining length of stay after extubation.



(a) $t = 1, 2, \ldots, 24$. (b) $t = 25, 26, 27$.

Figure EC.4 Optimal extubation policy without predictive information. The dark gray grid corresponds to extubate and the light gray grid to continue ventilation.



(a) $t = 1, 2, \ldots, 26$. (b) $t = 27$.

Figure EC.5 Optimal extubation policy with perfect predictive information. The dark gray grid corresponds to extubate and the light gray grid to continue ventilation.

EC.7.3. Different Weights on Terminal Cost

Physicians consider two important factors when making extubation decisions: ICU LOS and extubation failure. In the main analysis, we construct the objective function using only ICU LOS as the sole consideration. We show that the optimal policies that failed to consider predictive information can result in more extubation failures than the current practice. However, if the predictive information can be properly captured in the model, the optimal policies can lead to much lower

Table EC.7 Extubation performance under different extubation policies.

Initial Class	Data		OP-B		OP-PP ($Q = I$)	
	LOS (hr)	EFR (%)	LOS (hr)	EFR (%)	LOS (hr)	EFR (%)
1	115.3	21.4	105.9	20.2	104.7	19.2
2	104.4	27.8	105.2	21.3	102.2	21.8
3	103.7	29.9	103.3	30.2	101.5	25.6
4	104.6	26.3	100.7	21.4	99.6	20.1
5	98.3	18.7	95.4	20.7	94.9	19.9
6	93.9	16.6	94.6	17.7	93.3	16.3
7	86.7	13.9	85.0	13.8	85.0	13.8

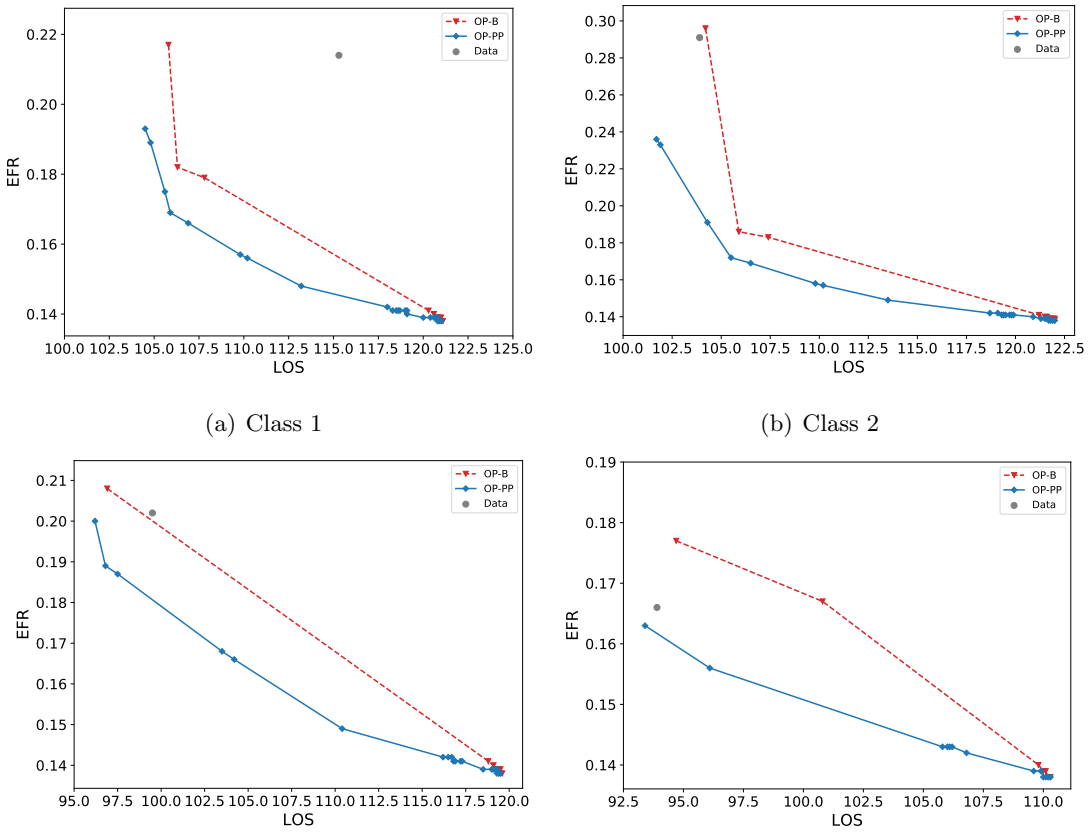
OP-B: optimal extubation policy without predictive information; OP-PP: optimal extubation policy with perfect prediction (i.e., $Q = I$); LOS: length of stay; EFR: extubation failure rate.

extubation failures while minimizing LOS only. In this part, we conduct sensitivity analysis by modifying the value function to consider extubation failures in the models directly. To this end, we introduce a tuning parameter λ to model the trade-off between ventilation duration and terminal cost. Specifically, the value function of the MDP-P model becomes

$$J_t^P(c_t, \hat{c}_{t+1}) = \min\Big\{\lambda G(c_t),\ H(c_t) + \sum_{c_{t+1}\in\mathcal{C}} \sum_{\hat{c}_{t+2}\in\mathcal{C}} P(c_{t+1}, \hat{c}_{t+2} \mid c_t, \hat{c}_{t+1}) J_{t+1}^P(c_{t+1}, \hat{c}_{t+2})\Big\}.$$

Since the terminal cost for each class is increasing with the risk of extubation failure, λ can be viewed as the weight of extubation failure over LOS; $\lambda = 1$ corresponds to the case reported in the main text. With a larger λ, the DM tends to ventilate patients longer and extubate patients when they are in better conditions. We vary λ from 1 to 10 with an interval 0.01, then run our model and find that the model with $\lambda = 1$ can reasonably characterize the reality. The results are summarized in Figure EC.6. Since class 5 represents the best condition, the weight has no impact on the optimal action for this class of patient; therefore, we only show the results for the other four classes.

In Figure EC.6, x-axis represents the expected total LOS from intubation, and y-axis corresponds to the EFR. Each point corresponds to a specific λ. In these charts, the points at the lower left corners represent more effective policies with lower LOS and lower EFR. Since larger λ results in longer LOS, λ increases as the point moves from left to right, with increasing LOS and decreasing EFR. The leftmost points on the curves correspond to cases with $\lambda = 1$, which are the results reported in the main text. With these charts, one can visualize the benefit of using predictive information and identify the inefficiency in the base model without predictive information compared with the current practice. The DM can use these charts to balance the trade-off between the two objectives, i.e., LOS and EFR, and choose a larger λ if he or she deems EFR should have a higher weight in the policies.



(a) Class 1 (b) Class 2

(c) Class 3 (d) Class 4

Figure EC.6 Trade-off in extubation outcomes under different weights on terminal costs.

OP-B: optimal extubation policy without predictive information; OP-PP: optimal extubation policy with perfect prediction (i.e., $Q = I$); LOS: length of stay; EFR: extubation failure rate.

EC.8. Regression Models in Terminal Costs Estimation

In all of the regression models, we not only incorporate the original clinical variables, but also transform the numerical variables into categorical ones according to the normal ranges, with the goal of producing more accurate regression models. For example, for heart rate (HR), we encode $\text{HR} < 60$, $\text{HR} \in [60, 100]$ and $\text{HR} > 100$ into three categories. The models that estimate RLOS, EFR and propensity to extubate are shown in Table EC.8.

Table EC.8: Regression models in parameters estimation.

Variables	RLOS	Failure	Extubate
Age	0.331***	0.000	-0.001
Gender: Male	-3.938	-0.005	0.325***
Race			
Chinese	1	1	1
Malay	-0.575	-0.021	0.051
India	-3.349	-0.026	0.082
Others	-9.732***	-0.021	0.021

Weight	-0.157	-0.001	0.003
Class			
1	1	1	1
2	5.932	-0.001	0.205
3	-0.799	-0.017	0.134
4	1.834	0.011	0.166
5	-1.500	-0.001	0.273
GCS	-0.037	-0.005*	0.170***
FiO2	0.326**	0.003***	-0.012***
Tidal volume	0.013	0.000	0.000
SOFA	2.303***	-0.001	-0.124***
RSBI	0.228*	0.000	0.000
HR			
60-100	1	1	1
<60	14.823	0.072	0.024
>100	-5.075	0.004	0.263
RR			
12-20	1	1	1
<12	-3.288	-0.060**	0.351***
>20	2.775	0.083***	-0.134
Creatinine			
45-110	1	1	1
<45	-5.949	-0.038	-0.006
>110	24.174***	0.115***	-0.456***
Arterial pH			
7.35-7.45	1	1	1
<7.35	5.816*	0.088***	-0.362***
>7.45**	8.972	0.015	-0.782***
Potassium			
3.5-5	1	1	1
<3.5	20.343	0.110***	-0.373***
>5	-0.205***	-0.023	0.103
Sodium			
135-145	1	1	1
<135	1.041	0.014	-0.175**
>145	31.120***	0.172***	-0.416***
Platelets			
150-450	1	1	1
<150	8.326***	0.045***	-0.077
>450	-1.313	0.073	0.568
WBC			
4.5-11	1	1	1
<4.5	-14.271	-0.056	0.603**
>11	6.565***	0.008	-0.018
MBP			
70-100	1	1	1
<70	4.557	0.046**	-0.070
>100	14.594*	0.079	0.067
Systolic BP			
90-120	1	1	1
<90	17.022**	0.001	0.002
>120	0.405	0.015	0.126**
Diastolic BP			
60-80	1	1	1
<60	4.011	0.022	-0.292***

>80	5.630	0.005	0.298
Temperature			
36.1-37.2	1	1	1
<36.1	0.082	0.022	0.010
>37.2	-4.368	-0.011	0.277***
Arterial PaO2			
80-100	1	1	1
<80	2.555	0.028	0.258***
>100	0.331	0.024	-0.117
Arterial SaO2			
93-97	1	1	1
<93	-3.026	-0.068***	0.055
>97	-6.222	-0.071***	0.438***
Haematocrit			
37-52	1	1	1
<37	-15.121**	-0.063	0.177
>52			-9.396
Haemoglobin			
12-17.5	1	1	1
<12	10.415*	0.025	-0.185
PaO2/FiO2	-0.740	-0.032***	-0.028
Cardiac rhythm: Paced	3.543	-0.018	0.119*
Cardiac rhythm: Sinus rhythm	12.053	-0.035	0.387
Cardiac rhythm: Sinus tachycardia	42.415	0.050	-0.341
Nursing assessment: Cardiac	2.739	0.039	-0.476***
Nursing assessment: Respiratory	9.159***	0.053***	-0.188**
Nursing assessment: Vascular	5.386	0.026	-0.338***
Nursing assessment: Musculoskeletal	-11.603***	-0.051***	-0.077
Nursing assessment: Neurological	-13.633	-0.095**	0.042
Nursing assessment: Nutrition	1.993	-0.074	-0.445*
No. of observations	3,067	3,067	7,281
(Pseudo) R-squared	0.15	0.16	0.27
Prob > Chi-squared			< 0.001

HR: heart rate; RR: respiratory rate; GCS: Glasgow Coma Scale; WBC: white blood cell; BP: blood pressure.
$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

EC.9. Constrained Estimation of Parameters

Threshold-type policies are widely used in practice due to better interpretability and ease of implementation. Our analytical result in Section 4.1—in particular, Theorem 1—guarantees the optimality of threshold-type polices as long as the transition probability matrix and misclassification matrix are TP2 (i.e., Assumptions 1 and 2). As discussed before, both assumptions are reasonable and are also made in other healthcare studies (e.g., Boloori et al. 2020). However, even the true transition probability matrix and misclassification matrix are TP2; because estimation errors, the naive estimators for these matrices may not satisfy the TP2 condition. As a result, the optimal policies derived using the biased estimators may not be threshold-type policies, which could be difficult to interpret and implement in practice, as shown in the above case.

We propose an estimation procedure here that embeds the TP2 requirement in estimation of the transition probability matrix and misclassification matrix by solving constrained MLE problems. Using this estimated transition probability matrix and misclassification matrix as the input, we are guaranteed to obtain threshold-type policies by solving the MDP-P model. We then evaluate the performance difference between the policies obtained via this approach and those obtained earlier using matrix parameters estimated via the naive MLE approach.

We formulate the constrained estimation problem for the transition probability matrix as

$$\begin{aligned}
\max_{P} \quad & \sum_{i,j\in\mathcal{C}} N(i,j)\log P(i,j) \\
\text{s.t.} \quad & P(i,j)P(i+1,j+1) \geq P(i,j+1)P(i+1,j),\ \forall i,j \in \mathcal{C}/\{C\} \\
& \sum_{j\in\mathcal{C}} P(i,j) = 1,\ \forall i \in \mathcal{C} \\
& 0 \leq P(i,j) \leq 1,\ \forall i,j \in \mathcal{C}
\end{aligned}$$

where N is a matrix that consists of the counts of patient transitions. The above problem maximizes the log-likelihood of observing transition data over the space of the transition probability matrix, which is constrained to be TP2. The constrained estimation problem for the misclassification matrix can be formulated in a similar way. The resulting estimation problem is a non-convex quadratically constrained optimization problem. We attempt to solve this problem via the sequential quadratic programming (SQP) algorithm, but not all of the constraints can be satisfied with the SQP algorithm; otherwise, it is infeasible. The same problem occurs with the TP2 misclassification matrix estimation. Note that the main reason we want to estimate a TP2 transition probability matrix is to obtain a threshold policy, so it is more important to guarantee feasibility than optimality when we solve the estimation problem.

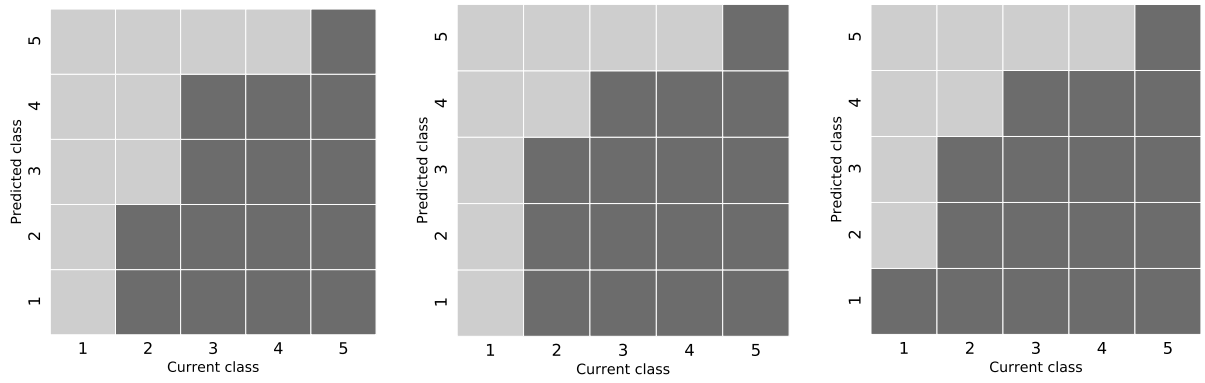
To address the above issue, we propose the following heuristic that solves a series of convex optimization problems based on the equivalence between the TP2 condition and row-wise MLR dominance, as stated in Section 4.1. In particular, we estimate the transition probability matrix row by row by requiring the MLR dominance between consecutive rows. For the first row, the transition probabilities are estimated via the typical MLE with only the simplex constraint. For the next row, we solve a constrained MLE with an additional condition whereby the second row MLR dominates the first row. This leads to a convex optimization problem that can be efficiently solved. We continue the process to estimate each of the remaining rows of the transition probability matrix. Although this procedure is a greedy heuristic, we find that it produces larger likelihoods

for both the transition matrix and misclassification matrix compared with those estimated by the SQP algorithm. The estimated matrices are given below:

$$\hat{P}_{TP2} = \begin{pmatrix} 0.35 & 0.27 & 0.22 & 0.11 & 0.05 \\ 0.12 & 0.37 & 0.29 & 0.15 & 0.07 \\ 0.04 & 0.12 & 0.48 & 0.24 & 0.12 \\ 0.02 & 0.06 & 0.22 & 0.35 & 0.35 \\ 0.01 & 0.02 & 0.07 & 0.18 & 0.72 \end{pmatrix},$$

$$\hat{Q}_{TP2} = \begin{pmatrix} 0.60 & 0.27 & 0.09 & 0.04 & 0.00 \\ 0.11 & 0.59 & 0.2 & 0.08 & 0.02 \\ 0.02 & 0.08 & 0.61 & 0.23 & 0.06 \\ 0.00 & 0.03 & 0.3 & 0.53 & 0.14 \\ 0.00 & 0.01 & 0.11 & 0.28 & 0.60 \end{pmatrix}.$$

Next, we resolve the MDP-B and MDP-P models using the above estimation. The OP-B and OP-PP policies are the same as those depicted in Figures 2 and 3, respectively. The new OP-IP policy is presented in Figure EC.7, which is a threshold policy as opposed to Figure 4. In particular, in epoch 27, class 3 patients who are predicted to improve to class 5 in epoch 28 will now be recommended to continue ventilation under the new policy in Figure EC.7.



(a) $t = 1, 2, \ldots, 25$. (b) $t = 26$. (c) $t = 27$.

Figure EC.7 Optimal extubation policy with imperfect predictive information using the adjusted parameters. The dark gray grid corresponds to extubate and the light gray grid to continue ventilation.

The performance of different policies under TP2 matrices is summarized in Table EC.9. All of the insights observed before continue to hold under the new parameters. We also derive the performance of the optimal extubation policy with imperfect prediction under the TP2 matrices while the system dynamics follow the non-TP2 matrices estimated in Sections 4.3 and 5.2. As shown in Table EC.10, even assuming the true system dynamics follow the parameters estimated using the naive MLE, the performance of the threshold policy obtained under the TP2 matrices is very close to the performance of the optimal policy reported in Figure 4 and Table 5. Therefore, we can use the proposed method to estimate TP2 matrices and then obtain a threshold policy for

practical implementation. Although it might not be optimal, the threshold policy is much easier to interpret and implement, without much loss of performance.

Table EC.9 Performance of different extubation policies under the adjusted parameters.

Class	OP-B		OP-PP		OP-IP	
	LOS (hr)	EFR (%)	LOS (hr)	EFR (%)	LOS (hr)	EFR (%)
1	107.9	23.5	106.6	20.5	107.4	20.5
2	104.2	29.6	102.0	23.8	103.1	23.4
3	97.1	21.0	96.4	20.1	97.0	20.3
4	94.6	17.7	93.3	16.3	94.2	16.9
5	85.0	13.8	85.0	13.8	85.0	13.8

OP-B: optimal extubation policy without predictive information; OP-PP: optimal extubation policy with perfect prediction; OP-IP: optimal extubation policy with imperfect prediction; LOS: length of stay; EFR: extubation failure rate.

Table EC.10 Performance of the optimal extubation policy with imperfect prediction derived using $\hat{P}_{TP2}$ and $\hat{Q}_{TP2}$ while the environment follows $\hat{P}$ and $\hat{Q}$.

Class	1	2	3	4	5
LOS (hr)	105.3	102.9	97.1	94.3	85.0
EFR (%)	19.3	23.0	20.3	16.9	13.8

LOS: length of stay; EFR: extubation failure rate.