10-2024

Extubation decisions with predictive information for mechanically ventilated patients in ICU

Guang CHENG

Jingui XIE

Zhichao ZHENG
Singapore Management University, DANIELZHENG@smu.edu.sg

Haidong LUO

Oon Cheong OOI

Extubation Decisions 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 crucial decision in intensive care units (ICUs), significantly affecting patient outcomes and the throughput of ICUs. This study aims to improve the current extubation protocols by incorporating predictive information on patient health conditions. We develop a discrete-time, finite-horizon Markov decision process with predictions of future state to support extubation decisions. We characterize the structure of the optimal policy and provide important insights into how predictive information can lead to different decision protocols. We demonstrate that adding predictive information is always beneficial, even if physicians place excessive trust in the predictions, as long as the predictive model is moderately accurate. Using a comprehensive dataset from an ICU in a tertiary hospital in Singapore, we evaluate the effectiveness of various policies and demonstrate that incorporating predictive information can reduce ICU length of stay by up to 3.4% and, simultaneously, decrease the extubation failure rate by up to 20.3%, compared to the optimal policy that does not utilize prediction. These benefits are more significant for patients with poor initial conditions upon ICU admission. Both our analytical and numerical findings suggest that predictive information is particularly valuable in identifying patients who could benefit from continued intubation, thereby allowing for personalized and delayed extubation for these patients.

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

1. Introduction

In recent years, healthcare-related costs have escalated significantly, with considerable wastage noted in areas such as failures in care delivery, lack of care coordination, overtreatment, and low-value care practices (Shrank et al. 2019). Concurrently, the rising healthcare demand drives service providers to adopt fast-track management, which requires swift and precise decision-making at every stage of patient care, including the services provided by intensive care units (ICUs) (Wilmore and Kehlet 2001, Zhu et al. 2012). On the other hand, with growing data availability, advances

in machine learning techniques, and enhanced computational power, highly accurate predictive models are becoming more attainable. Although significant efforts have been made to develop predictive models in the healthcare context, how to leverage predictive information to improve the quality of medical decisions remains underexplored. In this paper, we aim to integrate predictive information with prescriptive analytics to assess how predictive information can enhance decision-making quality in healthcare service delivery. In particular, we investigate *how and to what extent predictive information can improve the outcomes of the extubation decision—i.e., weaning patients from mechanical ventilators—in ICUs.*

In ICUs, patients who cannot breathe spontaneously may require invasive mechanical ventilation (MV) to assist with breathing. Patients are not allowed to be discharged from ICUs while intubated (i.e., receiving MV support). Since critical care is expensive for both patients and hospitals (Barrett et al. 2014, Halpern and Pastores 2015), the decision to extubate is vital for patients, and the time to extubation is usually considered the primary service outcome for surgical care in hospitals (Wong et al. 2016). On one hand, early extubation poses the risk of extubation failure (defined as the need for re-intubation or other breathing support measures) if the patient is unable to breathe spontaneously. Extubation failure is significantly associated with a longer ICU length of stay (LOS) and a 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 congestion in the units. Moreover, due to endotracheal tube placement, invasive ventilation is painful for patients and can lead to ventilator-associated complications, including airway injury, alveolar damage, and ventilator-associated pneumonia, which exacerbate patients' respiratory conditions (Hess 2011).

Typically, healthcare decisions follow standard protocols that consider current physiological data or historical clinical events based on clinical trials and experience. However, this approach can be shortsighted in medical decision-making, especially in critical care. For instance, a patient could deteriorate rapidly without intervention, and it can be too late to act once the patient's condition worsens. Early detection of imminent deterioration and preventive interventions can help save lives in critical care settings (Blum 2018). In practice, physicians often make predictions of patients' future conditions based on their own experience. However, the accuracy of subjective prediction is limited and can vary significantly among individuals. Massive amounts of data are underutilized, and important signals hidden in the data can be easily overlooked. By using appropriately validated and accurate predictive models, we aim to shed light on the creation of medical decision protocols that incorporate predictive information. As we will demonstrate in the subsequent case study, these protocols can potentially outperform prevailing decision-making practices, particularly in reducing both ICU LOS and extubation failures.

Motivated by this, our goal is to investigate the optimal time to extubate patients under MV. In recent years, advancements in machine learning and econometric analysis have made predictions of patients' future conditions and treatment effects more accessible and accurate. In this paper, we develop models that leverage 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 then 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 viewed as a medical intervention, and discharge can be seen as an optimal stopping problem. It is noteworthy that potential ethical issues may arise from these guidelines. While ethical considerations are an essential part of the discussion in developing practical policies, they fall outside the scope of this paper.

The main contributions and findings of our work are summarized as follows.

• We demonstrate that the classic MDP model can be adapted to incorporate predictions of future states into extubation decisions, and we characterize the structure of the optimal policy. Through this proposed modeling framework, we theoretically explore the value derived from integrating predictive information. We establish that predictive information is always valuable if the misclassification matrix, which will be defined later in the model development, is known.

• We investigate the conditions under which predictive information is strictly beneficial in a special case involving two patient classes. We derive the necessary and sufficient condition for this when the misclassification matrix is fully known to the physicians. We show that the benefit of predictive information increases concavely with sensitivity and linearly with specificity. We also characterize the condition under which the rate of benefit increase from sensitivity is greater than that from specificity. We provide an upper bound and a lower bound on the benefit of predictive information, serving as a rule of thumb for deciding whether to invest in a prediction model. When the misclassification matrix is unknown, we determine the necessary and sufficient condition under which predictive information can still be beneficial even when decisions are made under the assumption that the noisy prediction is perfectly accurate—i.e., the optimal policy with perfect prediction is applied in an environment with noisy predictions. This intuitively suggests that when prediction accuracy is high, benefits can be derived from using predictive information, even if prediction errors are completely disregarded.

• We apply our models to the extubation problem in ICUs, a crucial component of fast-track management, to assess the effectiveness of incorporating predictive information into 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 the ICU LOS and extubation failure rate—a measure closely related to in-hospital mortality. We further analyze the impact of prediction accuracy and show that even moderately accurate forecasts can substantially improve decision quality. These results provide support for more precise and personalized fast-track management in ICUs.

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 decisions 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 decisions, and Section 6 concludes.

2. Literature Review

Our work is related to four streams of literature: (1) ICU extubation decision-making; (2) integration of predictive and prescriptive analytics; (3) MDP and its applications in healthcare; and (4) optimal stopping and its applications.

2.1. ICU Extubation Problem

Being one of the most critical decisions in ICUs, the extubation problem has received substantial 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. A growing body of studies has begun to apply machine learning models to the extubation decision problem in recent years. The majority of these studies aim to predict extubation outcomes using machine learning methods. For instance, Wang et al. (2010) employed the adaptive neuro-fuzzy inference system to predict blood gas exchange for the management of ventilated patients in ICUs, and Kuo et al. (2015) applied an artificial neural network to predict extubation outcomes for mechanically ventilated patients. Tsai et al. (2019) integrated machine learning techniques with Bayesian decision analysis. They first predicted extubation outcomes, and then used Bayes' rule to update beliefs about patients' health conditions. However, as Kwong et al. (2019) noted, only a limited number of studies have started to assess the efficacy and effectiveness of machine learning techniques for MV in ICUs. Pursuing this direction, Prasad et al. (2017) and Jagannatha et al. (2018) applied reinforcement learning to determine the sedation prescription and extubation decision for ventilated patients. They assumed a Markov transition for the patient's health condition and formulated the reinforcement learning model as an MDP. Using the fitted Q-iteration algorithm (Riedmiller 2005),

they learned the reward function from the data and then selected the policy with the maximum reward. However, the policy thus learned only indicates 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 policy learning. More recently, Grand-Clément et al. (2021) examined a sequential decision-making problem of ventilator triage and reassessment for coronavirus (COVID-19) patients. They aimed to establish guidelines that are both data-driven and interpretable, and they developed optimal policies for MDP models based on decision trees. Their results emphasized the importance of tailoring the guidelines specifically to COVID-19 for better outcomes.

Unlike this body of literature, our work aims to leverage the predictive information to make more precise and personalized extubation decisions, which is not considered in existing models. We also seek to derive a policy that is interpretable and implementable in practice. Therefore, we adopt the classical MDP model, which is widely used in medical research and known for its interpretability among physicians. Additionally, we attempt to handle the selection bias for parameter estimation using advanced econometric 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 that apply operations research techniques to personalized medical decision-making (Mišić et al. 2010, Akan et al. 2012, Chan et al. 2013, Bertsimas et al. 2017, Hu et al. 2018, Ayer et al. 2019). Unlike these studies, we focus on how to incorporate predictive information into sequential decision-making to provide personalized recommendations for treatment continuation decisions.

2.2. Integration of Predictive and Prescriptive Analytics

The recent surge in the integration of predictive information into prescriptive analytics has drawn significant attention in the operations research and management science community. Bertsimas and Kallus (2020) 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 for single-stage decision problems. They argued that the parameters conventionally computed using the sample mean were insufficient. These parameters could be more accurately predicted using machine learning methods, and thus they used the results from a predictive model as input parameters in optimization models. Spencer et al. (2014) conducted a pilot study on an admission control problem with predicted future arrivals. Formulating this problem as a queueing model, they showed that future information could significantly reduce the system delay. However, the improvement came at the expense of a lengthy look-ahead

window, which required more sophisticated and accurate predictive models. To assess whether less future information could still yield significant improvement, Xu (2015) discussed the necessity of a nontrivial forecast window for arrivals in queueing models. The results showed that the performance would be no better than an online policy if the forecast windows were not sufficiently long. Building on these findings, Xu and Chan (2016) developed a proactive admission control policy for emergency departments based on the assumption that arrivals for several future time slots can be predicted. Simulation results showed that the proposed proactive policy outperformed the online policy, even when the predictions contained noise. 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. Dai and Shi (2019) developed a framework that supported inpatient overflow decisions in hospitals. They formulated the multi-class parallel-queue system as an MDP, where the system state included the count of waiting patients and patients who would be discharged on 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. (2021) formulated an MDP that prescribed ICU discharge decisions. In their model, the state variable was the number of patients within each class, corresponding to a particular readmission risk. The evolution of readmission risk was assumed to depend solely 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. More recently, Chen et al. (2023) explored how to integrate future demand arrival rates in the formulation of routing strategies for queueing systems experiencing demand surges. Using transient fluid control analysis, they proposed a two-stage index-based look-ahead routing policy, which excels in balancing operational costs and delivers superior performance, even in the presence of prediction errors. An increasing number of studies have incorporated predictive information into prescriptive analytics with applications across various domains (Chen and Farias 2013, Huang et al. 2019). Our research, while sharing similarities with the aforementioned literature regarding leveraging future predictive information in decision-making, differs with respect to problem formulation and methodology. In particular, to our knowledge, this paper is among the first to incorporate the predictive information into the state space of an MDP, while others use prediction information differently.

2.3. MDP and Its Applications in Healthcare

Our model builds on the theory of MDP (Bellman 1957). Markov chain models have been used extensively in the medical community for modeling disease progression, forming the backbone of many healthcare decision-making problems. MDP models enable the decision maker (DM) to select

actions that balance immediate rewards with uncertain future gains, taking into consideration the current situation, to yield the most optimal solution. MDP has seen widespread use in the healthcare domain when dealing with dynamic and stochastic decision-making problems (Schaefer et al. 2005). For instance, Alagoz et al. (2004) employed an MDP model to determine the optimal timing for living-donor liver transplantation. Likewise, MDP was leveraged to obtain optimal policies for choosing between living-donor and cadaveric livers (Alagoz et al. 2007a), and deciding whether to accept or reject a cadaveric liver for a patient in need of liver transplantation (Alagoz et al. 2007b). MDP has also been applied 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).

In recent years, robust MDP models have been developed and employed to handle parameter uncertainty in healthcare problems (Zhang et al. 2017, Goh et al. 2018, Grand-Clément et al. 2020). The robust MDP primarily focuses on the uncertainty in the transition probability matrix. Strong technical assumptions on the structure of the uncertainty set are often imposed for tractability, e.g., the rectangularity condition that mandates parameters of different states to be unrelated (Givan et al. 2000, Iyengar 2005, Nilim and El Ghaoui 2005, Wiesemann et al. 2013, Zhang et al. 2017), or a row-wise structure that demands each row of a possible transition matrix is a convex combination of a given set of probability vectors (Goh et al. 2018, Grand-Clément et al. 2020, Goyal and Grand-Clément 2022). Verifying the applicability of these assumptions in our problem context is not trivial. Moreover, our problem is further complicated by the uncertainty originating from the misclassification matrix, which is not addressed in traditional robust MDP models. Therefore, following the approach used in the literature (Chan et al. 2012, Xu and Chan 2016, Hu et al. 2018), we perform more sensitivity analyses to assess the robustness of our policies. We will leave the development of robust extubation policies with predictive information for future research.

It is worth mentioning that our model can also be formulated as a partially observable Markov decision process (POMDP), which is a generalization of MDP that accommodates states that are not completely observable. POMDP has garnered considerable attention in the healthcare community recently and has been successfully applied 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 distinguishes itself from classical MDP and POMDP models by incorporating predictions of future states into the state space. In practice, these predictions are typically generated from machine learning models constructed and validated using historical data. From this perspective, our model shares similarities with higher-order Markov chains, wherein system transition is contingent upon historical states (Raftery 1985, Ching et al. 2008, 2013). However, higher-order

Markov chains require that exponential numbers of parameters be characterized; otherwise, restrictive assumptions must be made on the dependence structure to reduce the parameter count. As such, higher-order Markov decision processes (HMDP) are not straightforward to optimize and have limited application in practice (Ching et al. 2004). Our approach permits greater flexibility, both in capturing historical information—not confined to state information as in HMDP—and in using advanced machine learning methods to project historical information onto a single prediction state. Thus, our model can efficiently reduce both the number of parameters and the computational complexity, which HMDP encounters.

2.4. Optimal Stopping and Its Applications

Our model also connects to the literature on the optimal stopping problem (Chow et al. 1971), which arises in numerous 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). Contemporary work on the optimal stopping problem involves approximating value functions and pinpointing greedy policies by means of approximate dynamic programming, simulation, and duality (Goldberg and Chen 2018). For instance, Desai et al. (2012) used the pathwise optimization method to calculate upper and lower bounds on the value function of an optimal stopping problem. Li and Linetsky (2013) solved a class of optimal stopping problems via eigenfunction expansion. More recently, Ciocan and Mišić (2022) proposed a binary tree-based greedy policy for the optimal stopping problem and characterized the depth required for a tree policy to approximate any optimal policy to a specified precision. They framed the problem of learning such policies from data as a sample average approximation problem. The advantage of their approach is its data-driven and interpretable nature, making it easy to implement in practice. The proposed policy outperformed other methods when applied to an option pricing problem. In a different vein, Iancu et al. (2020) considered a robust optimal stopping problem in continuous time, in which the system state cannot be observed except for a limited number of monitoring times, which the DM must select along with a suitable stopping policy. Rather than using conventional models that depict system evolution through specified stochastic processes, they applied a robust optimization framework and described the system states using an uncertainty set that captures possible system values at future monitoring times. The system evolution was modeled by updating the bounding functions—which place lower and upper bounds on future system values in the uncertainty set—after each observation. The DM's objective was to maximize the worst-case total reward. Our work adopts the conventional approach in which system evolution is characterized by a Markov process but diverges from the typical optimal stopping problem by incorporating predictions regarding future information into the state space.

3. Clinical Settings and Data

Our study is based on data collected from a cardiothoracic ICU of a tertiary hospital in Singapore. The dataset comprises 5,566 ICU admissions from March 2010 to October 2016. For each ICU admission, we compiled 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 dataset. Clinical variables were typically updated every 15 minutes to 4 hours, except for laboratory tests, which were often updated daily. The dataset provides detailed information about MV. The nursing staff documented the time of intubation and extubation, the type of endotracheal tube, and the ventilator mode at MV initiation. Patient conditions were recorded every 15 minutes to 4 hours on a ventilator. Based on this rich dataset, we can identify patients who received MV, 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 reinforcement learning models (Epstein et al. 1997, Fontela et al. 2005, Prasad et al. 2017, Chen et al. 2019). First, we eliminate readmitted cases within 30 days ($n = 648$). Note that we keep the first admission in this step. Due to the unavailability of historical data for patients admitted to the ICU in the initial month of our study period, i.e., March 2010, these patients cannot be classified as new admissions or readmissions. Hence, they must be excluded from our analysis. Nevertheless, our dataset does not include any patients requiring MV during March 2010. Patients with excessively long LOS typically have rare conditions and their treatment plans could be far from standard protocols. Hence, patients with over a 30-day ICU LOS ($n = 37$) are excluded from our study. Similarly, patients with more than one week of ventilation ($n = 32$) are 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 ($n = 13$), which are likely incorrect and are thus excluded. Cases with an NIV record during MV ($n = 14$) are excluded since their exact extubation times cannot be determined. Patients who died within 7 days after extubation ($n = 135$) are excluded since identifying their extubation outcomes is difficult. These patients do not count toward extubation failures in our numerical experiments. In our study ICU, patients are not allowed to be discharged within 6 hours after extubation. There are a few violations

in the data ($n = 18$), and we exclude these cases because the cause of the violations cannot be determined. We also remove cases without records of respiratory rate and tidal volume ($n = 11$) since these are necessary variables used in assessing patients' health conditions under MV. We will elaborate later in detail on how we assess and classify patient health conditions. Cases of unplanned extubation (i.e., accidental or patient-induced endotracheal tube removal, $n = 51$) are excluded from our analysis because they are not the result of regular extubation decisions. Finally, 3,067 separate entries are included in our study sample. Details of our data selection process are provided in Table EC.1 in Appendix EC.5.

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
Systolic BP (mmHg)	126.1	22.6	Nutrition	1,812	59.1
Diastolic BP (mmHg)	64.0	14.9	Initial class		
Temperature (°C)	34.9	3.1	1	22	0.7
Arterial pO2 (mmHg)	169.9	113.0	2	87	2.8
Arterial pCO2 (mmHg)	41.6	7.4	3	289	9.4
Arterial pH	7.4	0.1	4	705	23.0
Arterial SaO2 (%)	92.5	10.8	5	1,964	64.0
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.

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 an average of 11.7 hours. 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 Analysis

In this section, we describe the ICU extubation problem and formulate the problem as an MDP with perfect or imperfect predictive information, which we term the MDP-P model. We also present an MDP formulation that disregards predictions of future states, referred to as the MDP-B model, in Appendix EC.1 for benchmarking purposes. We derive structural results for the optimal policy and compare optimal value functions across different scenarios. Lastly, we outline the methodology for estimating parameters based on our dataset. We will use the terms ventilation and treatment interchangeably for the remainder of the paper.

The extubation decision-making process is graphically represented in Figure 1. Suppose a patient is initiated on MV. The DM, usually a physician, routinely monitors the patient's condition and determines at every checkpoint (or decision epoch) whether to extubate the patient. The decision problem concludes once the patient is taken off the ventilator.

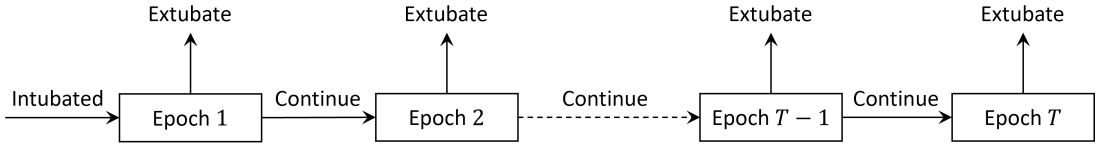


Figure 1 An illustration of the extubation problem.

4.1. Model Formulation and Theoretical Results

In this section, we formulate a mathematical model to characterize the extubation problem with predictive information (i.e., the MDP-P model) and derive the optimal policy. We assume that the DM has access to predictive information, which could potentially be erroneous. We first elaborate on the basic components of the MDP-P model.

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 epoch. The decision process for patients ventilated for more than a week is highly complex, significantly deviating from routine procedures. In our ICU dataset, only 1.6% of patients ($n = 67$)

were on ventilators for more than a week. Hence, we cap the maximum ventilation duration at one week (168 hours) for this study. In our study ICU, physicians assess a patient's condition every six hours to decide on extubation. Therefore, the decision interval is set to six hours, and consequently, there are a total of 28 epochs—i.e., $T = 28$, with the final epoch being terminal, meaning all patients are extubated.

States. To capture the patient's breathing condition, we categorize patients into one of C classes based on observed clinical variables, with each class $c \in \mathcal{C} = \{1, 2, \ldots, C\}$ corresponding to a distinct respiratory condition. The worst condition is represented by class 1, while class C indicates the best. We denote the patient's class at epoch t as c_t. Given access to predictive information, the DM can observe not only the current class c_t but also a prediction for the patient's class at the subsequent epoch, denoted as $\hat{c}_{t+1} \in \mathcal{C}$. Consequently, the state at 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), a well-established predictor for extubation failure (Fitch et al. 2014, Mahle et al. 2016, Yang and Tobin 1991). 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 is the tidal volume (L)—a lower RSBI value implies better respiratory function. The real-time monitoring of patients' respiratory conditions is facilitated by ventilators, which regularly record the respiratory rate and tidal volume of intubated patients. Various studies have suggested different RSBI cut-off points for predicting extubation outcomes. For instance, Yang and Tobin (1991) proposed an RSBI of 105, while Meade et al. (2001) suggested an RSBI of less than 65 for extubation. Another study by Chao and Scheinhorn (2007) concluded that an RSBI cut-off of less than 80 was overly conservative and that a higher threshold could enable more patients to benefit from successful early extubation. Recently, McConville and Kress (2012) reaffirmed the RSBI threshold of 105 proposed by Yang and Tobin (1991) while proposing a set of early extubation strategies.

In this study, we adhere to the medical literature by using RSBI as the classification criterion to stratify patients into different risk groups for extubation. Ventilators record respiratory rate and tidal volume, which are updated in our dataset every 15 minutes to 1 hour. We divide RSBI into five classes $\{[105, +\infty), [65, 105), [45, 65), [35, 45), (0, 35)\}$ based on threshold values derived from the medical literature reviewed above, and physicians' suggestions from our study ICU. Thus, $C = 5$ in our case. A sensitivity analysis conducted with alternative, more granular patient classification schemes in Appendix EC.7.4 provides findings consistent with those reported here in the base case.

Actions. At each decision epoch, the DM must decide whether to extubate the patient. The action at epoch $t \in \mathcal{T}$ is $a_t \in \mathcal{A} = \{0, 1\}$, with 0 indicating continuation and 1 indicating extubation.

As previously stated, we assume the treatment is discontinued, regardless of the patient's health condition at the final epoch, i.e., $a_T = 1$.

Transitions. If the DM chooses to continue at epoch t, i.e., $a_t = 0$, the patient in class i will transition to class j at epoch $t+1$ with a probability of $P(i,j)$, where $i, j \in \mathcal{C}$. On the other hand, if the treatment is terminated, i.e., $a_t = 1$, the decision problem concludes. In the machine learning community, a confusion matrix is typically used to evaluate the performance of predictive models in binary classification problems. To capture prediction errors, we introduce a misclassification matrix, Q, an analog to the confusion matrix for multi-class classification problems. Each component of the misclassification matrix $Q(i,j) := \mathbb{P}(\hat{c} = j \mid c = i)$ represents the probability that a patient transitioning to class i at the next epoch is predicted to transition to class j when the prediction is made at the current epoch. Note that the misclassification matrix is stationary, meaning prediction errors do not depend on time. The diagonal elements of Q represent the probabilities that the corresponding class of patients in the next epoch can be correctly predicted. If the prediction is perfectly accurate, Q is the identity matrix, denoted as I. 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 at epoch $t+1$ can be computed as $\tilde{Q}(c_t, \hat{c}_{t+1}) = \sum_{c \in \mathcal{C}} P(c_t, c) Q(c, \hat{c}_{t+1})$. Using Bayes' rule, given the current state $(c_t, \hat{c}_{t+1})$, the probability that the true class at 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 $\boldsymbol{\pi}$ can be interpreted as a belief vector of the DM on the patient's future condition given the current class c_t and the 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

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

Costs. At each epoch, the associated cost depends on the patient's current class and the DM's action, while the predicted class does not affect the immediate cost. We set our objective to minimize the expected ICU LOS, a primary service outcome for surgical care provided by hospitals (Wong et al. 2016). This objective aligns with the goals of fast-track management, which encourages physicians to extubate patients aggressively to reduce their stays in the ICU (Sato et al. 2009, Fitch et al. 2014). We demonstrate in our numerical analysis that, while primarily targeting the reduction of LOS, optimal policies that incorporate predictive information can also effectively reduce the extubation failure rate (EFR). We further demonstrate in Appendix EC.7.5 that EFR can be directly calibrated into the cost parameters to capture the trade-off between service and clinical outcomes. In our dataset, there is no significant correlation between the remaining LOS

after extubation and ventilation duration (correlation coefficient $= 0.01$, p-value $= 0.92$). Thus, we assume the remaining LOS is independent of the length of the patient's stay. This is consistent with the assumption in the literature that ICU LOS is geometrically distributed (Chan et al. 2012, Dai and Shi 2019, 2021). The extubation decision does not impact ICU stays before intubation. Therefore, we only consider the ICU stay from intubation as the objective in this study. In the remainder of this paper, we use LOS to denote ICU LOS after MV initiation. Under this premise, the immediate cost H incurred by continuing ventilation equals the length of the decision interval, which is 6 hours. The terminal cost $G(c)$ is the expected remaining LOS after extubation. We assume that $G(c)$ decreases[1] with c, indicating that the remaining LOS shortens when a patient is extubated in a better condition. In Appendix EC.3, we show that our structural results remain valid under the same set of assumptions (as stated below) when the cost to continue is multiplied by a discount factor.

The optimality equation of the MDP-P model can be expressed as

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

where the transition probability $\mathbb{P}(c_{t+1}, \hat{c}_{t+2} \mid c_t, \hat{c}_{t+1})$ is computed as in Equation (2). Here, $J_t^P(c_t, \hat{c}_{t+1})$ represents the minimum expected total cost incurred by a patient in class c_t at epoch t, given a prediction of the breathing condition at epoch $t+1$ as $\hat{c}_{t+1}$. The boundary condition is $J_T^P(c_T, \hat{c}_{T+1}) = G(c_T)$, for $c_T \in \mathcal{C}$. Note that at epoch T, the predicted class $\hat{c}_{T+1}$ beyond time horizon T does not affect the value function since the process ends at epoch T, and the terminal cost depends solely on the patient class at epoch T.

To obtain the structural results for MDP-P, we need certain assumptions. Two related concepts are introduced below. The first is monotone likelihood ratio (MLR) dominance. A probability vector $\boldsymbol{x}$ is said to dominate $\boldsymbol{y}$ in MLR ordering if $\boldsymbol{x}(j)\boldsymbol{y}(i) \geq \boldsymbol{x}(i)\boldsymbol{y}(j)$, for $i < j$, represented as $\boldsymbol{x} \geq_{lr} \boldsymbol{y}$. MLR dominance is frequently used in POMDPs since it is preserved under conditional expectations (e.g., Bayesian updates). Next, we introduce the definition of total positivity of order 2 (TP2). A matrix P is considered TP2 if all its second-order minors are nonnegative. Equivalently, P is TP2 if and only if $P_j \geq_{lr} P_i$ for $i < j$, where P_i is the i^{th} row of P (Krishnamurthy 2016). With these concepts, we are ready to introduce the required assumptions for the MDP-P model, which are deemed reasonable in our problem context and have also been made in other healthcare studies (e.g., Boloori et al. 2020).

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

ASSUMPTION 1. *Transition matrix P is TP2.*

Assumption 1 ensures that a healthier patient is expected to evolve towards a better condition. 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 $j > i+1$, meaning that patients will not transition more than one class in a single epoch, then Assumption 1 holds if $P(i,i)P(i+1,i+1) \geq P(i,i+1)P(i+1,i)$.

ASSUMPTION 2. *The misclassification matrix Q is TP2.*

Assumption 2 suggests that a patient with a better health state is more likely to be predicted to be in a more favorable state at the next epoch compared to a patient 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 means that a healthier patient gains less benefit from one additional epoch of ventilation, regardless of the prediction for the future health state. Given these three assumptions, we can demonstrate the following properties of the value function for the MDP-P model, followed by the structure of the optimal policy. Proofs for all technical results are relegated to Appendix EC.2.

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

Lemma 1 suggests that optimal ventilation cost decreases with patient class. Furthermore, if a patient is predicted to be in a better health state at epoch $t+1$, the expected optimal cost will 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 at 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 in state $(c_t, \hat{c}_{t+1})$ is*

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

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

(1) For all $t \in \mathcal{T} \backslash \{T\}$, $a_t^*(c_t, \hat{c}_{t+1})$ is increasing in c_t but decreasing in $\hat{c}_{t+1}$.

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

Theorem 1 establishes that the optimal policy for MDP-P is a switching curve policy. We note that the inclination to extubate the patient increases with the patient's current class and decreases in the predicted class. As outlined in Appendix EC.1, the optimal policy in MDP-B, i.e., the base MDP model without predictive information, is a time-dependent threshold policy that solely

depends on the current class. With the inclusion of predictive information, patients who would have been extubated according to the optimal policy in MDP-B may continue receiving ventilation if their condition is expected to improve at the next epoch. Similarly, patients who would continue receiving ventilation based on the optimal policy in MDP-B may be extubated if their condition is predicted to worsen at the next epoch. This suggests that predictive information allows for more accurate decision-making via personalized treatment effect prediction, a factor confirmed to be significant in other medical decision-making problems (Chan et al. 2012, Kim et al. 2014). Furthermore, the extubation region $\mathcal{R}_t^1$ expands with time t (equivalently, the continuing region $\mathcal{R}_t^0$ contracts with time t), indicating that the stopping criterion should be more relaxed as the treatment duration t increases.

Remark 1. We can also formulate the optimal stopping problem with predictive information using the POMDP model. Under the POMDP framework, predictive information is modeled as an observation at each epoch, and the patient's condition at the next epoch is depicted as a partially observable state. The DM forms a belief about the patient's actual condition at the next epoch using the misclassification matrix (referred to as the information matrix in POMDP terminology) and makes the stopping decision based on this 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 at the next epoch—is revealed by the predictive model, and the process continues. More details on the POMDP formulation are provided in Appendix EC.4, where we show that all structural results regarding the optimal policies can be proven 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, as, in our model, the partially observed state—i.e., the patient's condition at the next epoch—is disclosed at 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 epoch.

4.2. The Benefit of Predictive Information

In this section, we theoretically explore the value of 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 satisfactory, the DM might use the predictive information directly as if it were precise. Hence, we further discuss a case where the policy is derived using identity matrix I as the misclassification matrix, while the actual misclassification matrix Q remains unknown to the DM. In a special case with two patient classes, we derive the necessary and sufficient condition for the predictive information to be beneficial.

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

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

The following result can be derived.

THEOREM 2. *The optimal policy with predictive information consistently outperforms policies without predictive information, i.e., $J_t^P(c_t) \leq J_t^B(c_t)$, for all $t \in \mathcal{T}$ and $c_t \in \mathcal{C}$, where $J_t^B(c_t)$ represents the minimum cost incurred by a class c_t patient at epoch t under the base model without prediction.*

Theorem 2 suggests that any predictive information (even if inaccurate) is valuable. That is, for any class c patient, the expected cost under the optimal policy with predictive information is not higher than that under any policy without predictive information. This result is based on the fact that the true misclassification matrix Q is known to the DM.

Even though predictive information is always beneficial when the underlying misclassification matrix is known, the relation between prediction errors and the associated improvement remains unclear. To address this question, a special case with two patient classes in an infinite time horizon is examined. We confine ourselves to this simplified setting because the value function for the general case is challenging to evaluate. Moreover, the prediction accuracy of the multi-class setting cannot be easily summarized into a few parameters from the misclassification matrix without imposing strong assumptions. Hence, we aim to glean some insights from analyzing this special case, where the prediction accuracy can be represented by the well-known sensitivity and specificity measures, and the findings can be numerically validated in the multi-class setting. Note that we consider the infinite-horizon setting for notational simplicity, and all the following results hold under the finite-horizon setting.

Suppose there are two classes of patients: class 1, representing patients in a severe condition, and class 2, denoting those in a better state. The transition matrix can be fully captured by its diagonal elements: $p_1 := P(1,1)$ and $p_2 := P(2,2)$, with the assumption that $0 < p_1 < 1$ to avoid any trivial circumstances. Similarly, the misclassification matrix can be represented by its diagonal elements, i.e., $q_1 := Q(1,1)$ and $q_2 := Q(2,2)$, which stand for sensitivity and specificity, respectively, in binary classification problems. We assume that Assumption 2, which in this case simplifies to the condition $q_1 + q_2 > 1$, holds in the following analysis. In this setup, the optimal decision for class 2 patients is to discontinue ventilation because it does not offer any further benefit. Hence, we only need to consider the cost related to class 1 patients. In our analysis of this special case, we define the benefit of predictive information as the difference in expected total costs between

the MDP-B model and the MDP-P model for class 1 patients, i.e., $J^B(1) - J^P(1)$. We will also assume that $\Delta G := G(1) - G(2) > H$. Otherwise, ventilation provides no benefit, and all patients should be taken off ventilators immediately. The following theorem outlines the necessary and sufficient condition for predictive information to be strictly beneficial, demonstrating how the value of predictive information varies with the model parameters.

THEOREM 3. *Given two patient classes and under Assumption 2, we have the following results:*
(a) Predictive information is strictly beneficial if and only if

$$\kappa_1 < \frac{H}{\Delta G} < \kappa_2, \tag{5}$$

where

$$\begin{aligned}
\kappa_1 &= \mathbb{P}(c_{t+1} = 2 \mid \hat{c}_{t+1} = 1) = \frac{(1-p_1)\gamma}{p_1 q_1 + (1-p_1)\gamma}, \\
\kappa_2 &= \mathbb{P}(c_{t+1} = 2 \mid \hat{c}_{t+1} = 2) = \frac{1-\lambda-\gamma}{1-\gamma}, \\
\gamma &= \mathbb{P}(\hat{c}_{t+1} = 1 \mid c_t = 1) = p_1 q_1 + (1-p_1)(1-q_2), \text{ and} \\
\lambda &= \mathbb{P}(\hat{c}_{t+1} = 2, c_{t+1} = 1 \mid c_t = 1) = p_1(1-q_1).
\end{aligned}$$

(b) When condition (5) holds, the benefit of predictive information is concavely increasing in sensitivity q_1 and linearly increasing in specificity q_2. Specifically, the rate of benefit increase from q_1 is greater than that from q_2 if

$$\left(1 - \frac{H}{\Delta G}\right)[1 - p_1(1-p_1)(1-q_1-q_2)] - p_1 \le 0. \tag{6}$$

Otherwise, the rate of benefit increase from q_2 is greater than that from q_1.

(c) When condition (5) holds, the benefit of predictive information is bounded above and below by $\min\{\beta_1,\ \beta_2\}$ and $q_1\beta_1 + q_2\beta_2 - \max\{\beta_1, \beta_2\}$, respectively, where $\beta_1 = p_1 H/(1-p_1) - p_1(\Delta G - H)$, and $\beta_2 = (1-p_1)(\Delta G - H)$.

Part (a) of Theorem 3 provides a necessary and sufficient condition under which the predictive information is strictly beneficial when dealing with two patient classes. Note that $0 < \kappa_1 \le \kappa_2 < 1$. This condition conveys that the holding cost must not be excessively high or low in comparison to the difference between the terminal costs. Intuitively, a prohibitively high holding cost negates any benefit derived from awaiting a patient's condition improvement. Patients in class 1 would consequently be extubated immediately unless the negative predictive value of correctly detecting condition improvement (i.e., κ_2) is sufficiently high. More specifically, the second inequality of condition (5) can be expressed as $(1-\gamma)H < (1-\lambda-\gamma)\Delta G$. It is noteworthy that $1-\gamma$ represents the probability that a class 1 patient is predicted to transition to class 2, while $1-\lambda-\gamma$ is the likelihood that a class 1 patient predicted to transition to class 2 indeed makes that transition. If a

class 1 patient is predicted to transition to class 2, and if the policy leverages this prediction to keep the patient ventilated for one more epoch, then $(1-\gamma)H$ represents the expected holding cost of this policy. Conversely, $(1-\lambda-\gamma)\Delta G$ captures its expected benefit. Thus, this condition corresponds to a situation where the predictive information is valuable, advising the DM to postpone extubation if the patient is likely to improve at the subsequent epoch.

In contrast, if the holding cost is negligibly low, the continuation of ventilation and waiting for a patient to improve to class 2 is more beneficial, even in the absence of predictions. The first inequality of condition (5) sets the threshold for such a scenario. In particular, if the prediction indicates that a patient's condition will not improve at the next epoch, extubating the patient might be optimal, even when the patient is in class 1 under the given conditions. In other words, awaiting a patient's transition to class 2 may not be worthwhile. By substituting the expression of κ_1 and rearranging the terms, the inequality can be rewritten as follows:

$$\left(\frac{1}{1-p_1}-\frac{1-\gamma}{1-\lambda}\right)H > \frac{\gamma}{1-\lambda}\Delta G.$$

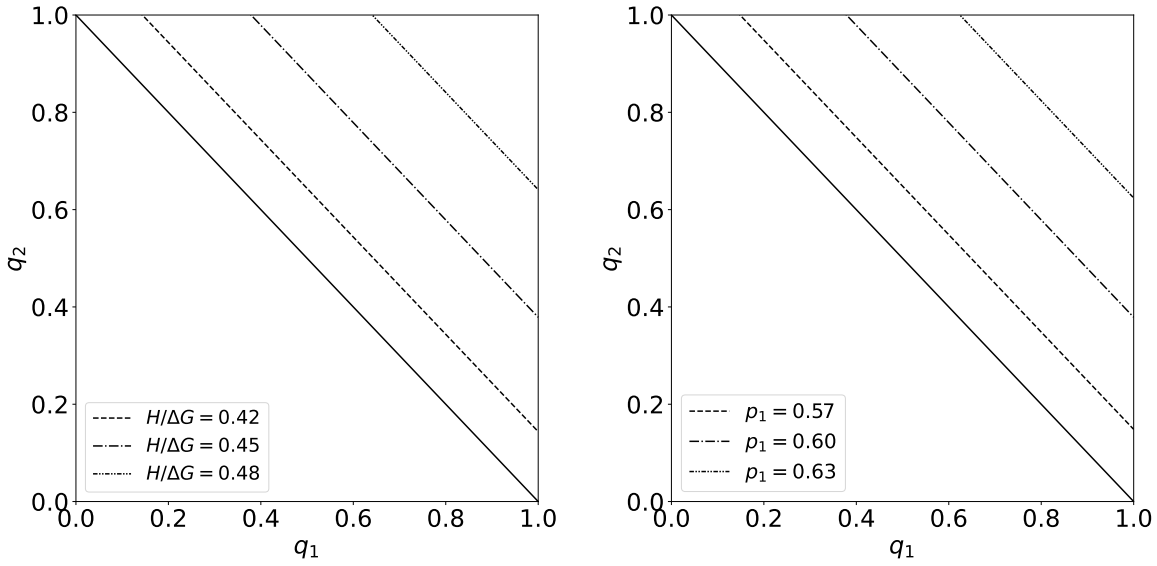
Here, $1/(1-p_1)=\sum_{i=0}^{\infty} p_1^i$, which represents the expected number of epochs needed for a class 1 patient to transition to class 2. Similarly, $(1-\gamma)/(1-\lambda)=(1-\gamma)\sum_{i=0}^{\infty}\lambda^i$ accounts for the expected number of epochs for a class 1 patient who remains in class 1 but is erroneously predicted to transition to class 2 at the succeeding epoch. The left-hand side of the above inequality represents the expected holding cost saved by early extubation under the policy with predictive information—i.e., extubating class 1 patients predicted to remain in class 1 at the next epoch—compared to the policy of extubating only class 2 patients, or, in other words, keeping all class 1 patients ventilated until their condition improves. The right-hand side of the inequality represents the expected increase in terminal costs due to not waiting for a patient's condition to improve. In essence, this condition states that the expected holding cost reduction resulting from early extubation should exceed the expected increase in terminal costs induced by early extubation. Therefore, this condition illustrates a situation in which predictive information is crucial in instructing the DM to proceed with early extubation if the patient is unlikely to improve at the next epoch. Lastly, we would like to highlight that condition (5) can be rewritten in terms of q_1 and q_2. The detailed derivation of alternative expressions can be found within Theorem 3's proof in Appendix EC.2.

Part (b) of Theorem 3 sheds light on how the benefit of predictive information varies with the parameters related to prediction accuracy, namely, sensitivity q_1 and specificity q_2. As anticipated, the benefit is enhanced with increased sensitivity and specificity. However, the second-order conditions suggest that the benefit accrues at a diminishing rate with an increase in sensitivity, making it marginally less beneficial to correctly identify a Class 1 patient who will remain in Class 1 at

the next epoch. We further characterize the condition under which the rate of benefit increase from sensitivity is greater than that from specificity—in other words, when sensitivity is more valuable to improve than specificity. From expression (6), it can be deduced that as the cost ratio $H/\Delta G$ increases, the region where sensitivity holds more importance expands. Furthermore, if $p_1^2 + (1-2p_1)(1-H/\Delta G) \geq 0$, it can be demonstrated that the region where sensitivity is more important grows as p_1 increases. Figure 2 depicts these trends. In Figure 2(a), where $p_1 = 0.60$, we illustrate three separating lines corresponding to $H/\Delta G = 0.42$, 0.45, and 0.48. The solid line represents the condition $q_1 + q_2 = 1$. Within the region between the 145-degree solid line and each separating line, sensitivity is more important than specificity. As $H/\Delta G$ increases from 0.42 to 0.48, the region where sensitivity holds more importance expands. In Figure 2(b), with $H/\Delta G$ fixed at 0.45, we draw the separating lines by varying p_1 from 0.57 to 0.63. It is clear that the separating lines move upward as p_1 rises. Moreover, sensitivity is always more valuable to improve when $H/\Delta G \geq (1-p_1^2)/[1+p_1(1-p_1)]$. Conversely, specificity becomes more valuable when $H/\Delta G \leq (1-p_1)^2/[1-p_1(1-p_1)]$. A notable observation from condition (6) is that it depends on the sum of q_1 and q_2, rather than on their individual values. Even if q_1 is significantly larger than q_2, the advantage of increasing q_1 remains greater than that of increasing q_2, as long as condition (6) is met. It is important to note that this pertains to the benefit of predictive information, which does not account for the costs associated with increasing sensitivity or specificity. In practice, the DM can assess whether condition (6) is satisfied and decide which parameter to prioritize for improvement when additional efforts are feasible, considering the cost and benefits of such improvements.

Finally, part (c) of Theorem 3 provides an upper bound and a lower bound on the benefit of predictive information. The upper and lower bounds collectively offer guidance on the viability of investing in a predictive model. The upper bound indicates the maximum benefit one could potentially achieve using predictive information. If the upper bound suggests minimal potential benefits, it may not be worth pursuing. But if it is significant, the lower bound offers further insight into the feasibility and required accuracy for a predictive model.

For a clearer understanding of the bounds, note that β_1 represents the benefit derived from perfect prediction when only class 2 patients are extubated by the base model without prediction. Here, the benefit of predictive information stems from extubating class 1 patients who do not benefit from extended ventilation. To elaborate, the expected cost under the base model is $H/(1-p_1) + G(2)$, where $H/(1-p_1)$ denotes the expected holding cost, and $G(2)$ is the terminal cost. With perfect prediction, a class 1 patient, who will not show improvement at the following epoch, will be extubated at the initial decision epoch. Otherwise, the patient will transition to class 2 at the subsequent epoch and then be extubated. Thus, the expected cost is $p_1 G(1) + (1-p_1)[G(2)+H]$. The difference between these costs equates to β_1, denoting the benefit of predictive information in this



(a) $p_1 = 0.60$. (b) $H/\Delta G = 0.46$.

Figure 2 Relative importance of sensitivity and specificity. The region below each separating line represents the area where the rate of benefit increase from sensitivity q_1 is greater than that from specificity q_2.

scenario. Analogously, β_2 depicts the benefit of perfect prediction when all patients are extubated by the base model. In this situation, the potential cost-saving emerges from not extubating those patients transitioning to class 2 at the next epoch, which equals ΔG minus the holding cost H. The transition probability for a class 1 patient to class 2 is $1 - p_1$. Thus, $\beta_2 = (1 - p_1)(\Delta G - H)$ represents the expected benefit of predictive information in this context. When condition (5) holds, the benefit of predictive information is bounded above by $\min\{\beta_1, \beta_2\}$. Interestingly, the bound is not always increasing in $\Delta G - H$, which represents the benefit of waiting for one more epoch to extubate in a better condition. Intuitively, a larger $\Delta G - H$ seems to offer a stronger incentive to know the patient's condition at the next epoch. However, if $\Delta G - H$ is large enough, then one will wait for the patient's condition to improve even in the absence of a prediction, aligning with the optimal policy with predictive data.

Concerning the lower bound, recall that q_1 is the likelihood of identifying class 1 patients who will stay in class 1 at the next epoch. This probability scales the benefit of employing predictive information to extubate class 1 patients not showing any improvement with extended ventilation. This justifies the first term in the lower bound, namely, $q_1\beta_1$. In addition, q_2 is the likelihood of detecting class 1 patients transitioning to class 2 at the subsequent epoch. This probability scales the benefit of using predictive information to defer the extubation decision for class 1 patients likely to benefit from continued ventilation. This logic underpins the second term in the lower bound, i.e., $q_2\beta_2$. It is also noteworthy that the upper and lower bounds coincide when $q_1 = q_2 = 1$.

In practice, even though both sensitivity and specificity might not be immediately available for a specific case, one could approximate these values through a small-scale study or a literature review. This would facilitate the estimation of a lower bound on the benefit of predictive information using our findings.

The results presented above depend on the underlying assumption that the true misclassification matrix Q is known to the DM. In practice, however, obtaining the true misclassification matrix may not be straightforward. Typically, the misclassification matrix is derived from out-of-sample tests using data. The accuracy of this estimation largely depends on the data quality, which could be compromised. In some scenarios, the DM might base decisions entirely on the prediction, provided the predictive model's accuracy is deemed satisfactory. To capture this feature, we analyze the cost of the MDP-P model under the optimal policy that assumes I to be the misclassification matrix, as detailed in Theorem 4.

THEOREM 4. *Suppose that the true misclassification matrix of the prediction model is unknown to the DM, and Assumption 2 holds. If the DM applies an imperfect predictive model as if it were perfect, i.e., the optimal extubation policy derived from assuming perfect prediction, we have*

(a) If $\kappa_1 < H/\Delta G < \kappa_2$, the optimal policy derived from assuming perfect prediction is better than the optimal policy without prediction.

(b) If $\kappa_2 \le H/\Delta G < 1$ or $(1-p_1)/(2-p_1) < H/\Delta G \le \kappa_1$, the optimal policy derived from assuming perfect prediction is worse than the optimal policy without prediction.

(c) If $H/\Delta G \le (1-p_1)/(2-p_1)$, the optimal policy derived from assuming perfect prediction is equivalent to the optimal policy without prediction

Part (a) of Theorem 4 suggests that for a reasonable predictive model, an interval exists such that if the cost ratio $H/\Delta G$ resides within this span, employing predictive information is beneficial, even when its true accuracy is overlooked and perceived as perfect. Part (b) of Theorem 4 conveys that under certain circumstances, for instance, when either q_1 or q_2 is excessively small, assuming perfect prediction could lead to erroneous decisions and inferior outcomes. When the holding cost is small enough, both policies adopt the identical decision rule—always keeping class 1 patients intubated, irrespective of the prediction, as deduced in part (c) of Theorem 4.

REMARK 2. It is noteworthy that the condition specified in part (a) of Theorem 4 coincides with condition (5) in the case where the misclassification matrix is known, as described in Theorem 3. Combining these results, within the scope of a two-patient-class scenario, we can infer that if the predictive information is validated as beneficial (as per condition (5) presented in Theorem 3), then implementing the optimal policy that regards the imperfect prediction as if it were perfect is not worse than the optimal policy without predictive information. An immediate practical implication

is that once the data analytics team affirms the value of a predictive model by assessing its accuracy and cost parameters, it is feasible to disregard the misclassification matrix during the actual rollout. In such cases, it would suffice to communicate only the predictions to the physicians for their deliberation. Adopting this strategy might mitigate the complexity physicians face in interpreting the misclassification matrix, thereby facilitating smoother adoption of the predictive model in real-world settings. However, it is crucial to acknowledge that this conclusion is predominantly based on the framework of a two-patient-class setup. In situations involving multiple classes, one should exercise caution regarding this implication.

4.3. Parameter Estimation

In this section, we estimate the unknown parameters of the proposed model. Specifically, we first estimate terminal costs using a two-step approach. Transition probabilities are then estimated using the sample mean. Subsequently, we train a practical prediction model and estimate the misclassification matrix via out-of-sample testing.

Terminal costs. The terminal cost $G(c)$ incurred by extubation is the expected remaining LOS after extubation (RLOS) conditioned on patient class, i.e., $G(c) := \mathbf{E}[\text{RLOS} \mid c]$. We expect $G(c)$ to be positively correlated with the risk of extubation failure $r(c)$ since patients with failed extubation require additional respiratory support. A failed attempt might even exacerbate a patient's condition, prolonging their stay. In our dataset, patients who experienced extubation failure had a significantly longer RLOS (144.8 hours vs. 66.5 hours, p-value $= 0.000$). Our estimation results confirm this conjecture. By accounting for both the treatment and terminal costs in the objective, we effectively incorporate the EFR when minimizing the overall cost. It is possible to assign a higher weight to the terminal cost, making EFR a more dominant factor in the optimization. We present a sensitivity analysis of different weights on the terminal costs in the objective in Appendix EC.7.5. As demonstrated in subsequent numerical results, the policies derived from our models with predictive information not only minimize LOS but also significantly reduce EFR, even in the baseline setup without placing higher weights on terminal costs.

Estimating the terminal cost presents significant challenges, primarily due to selection bias. Within each patient class, those who were extubated were likely healthier than those who were not, as physicians made extubation decisions. Consequently, a sample average approach might produce a biased estimator for the terminal cost. To mitigate this bias, we employ a two-step procedure, leveraging our rich dataset. The specifics of the estimation process are discussed below.

Using additional information from our dataset, the expression for $G(c)$ can be reformulated as

$$G(c) = \mathbf{E}\left[\mathbf{E}\left[\text{RLOS} \mid c, \boldsymbol{x}\right] \mid c\right], \tag{7}$$

where $\boldsymbol{x}$ represents a vector of patient characteristics and clinical data just before extubation, as listed in Table 1, excluding patient class. We also incorporate the sequential organ failure assessment (SOFA) score, a widely-recognized severity metric used in ICUs (Vincent et al. 1996), to account for the underlying health status. The first step is to estimate $\mathbf{E}[\text{RLOS} \mid c, \boldsymbol{x}]$ and subsequently determine the distribution of patient attributes $\boldsymbol{x}$ for each patient class using a propensity score approach to deduce $\mathbf{E}[\text{RLOS} \mid c]$. The first step naturally lends itself to linear regression. The fitted value derived from the regression model,

$$\text{RLOS} = b_0 + \boldsymbol{b}_1'\boldsymbol{x} + \boldsymbol{b}_2'\boldsymbol{c} + u, \tag{8}$$

serves as an estimator for $\mathbf{E}[\text{RLOS} \mid c, \boldsymbol{x}]$, where (with a slight abuse of notation) $\boldsymbol{c}$ represents the set of dummy variables for patient class; b_0, $\boldsymbol{b}_1$, and $\boldsymbol{b}_2$ are the corresponding coefficients; and u is the random error term. Our methodology aligns with numerous empirical studies that use untransformed LOS as the dependent variable (e.g., Moran et al. 2008, Niskanen et al. 2009, Kc 2020). Verburg et al. (2014) demonstrated that a linear regression with an untransformed LOS truncated at 30 days yields the highest R-square in comparison with other forms of LOS and regression models. In the literature, it is also common to apply a logarithmic transformation to LOS as the dependent variable (KC and Terwiesch 2012, Hu et al. 2018), in which, however, the goal is usually to identify the risk factors for extended LOS. In our scenario, a logarithmic transformation is infeasible since we intend to estimate the expected RLOS, and this expectation does not hold under the logarithmic transformation. Nevertheless, we also estimate the model with a logarithmically transformed RLOS and revert the fitted log(RLOS) to the RLOS scale using the exponential function. This method results in a higher root mean squared error than Equation (8) (61.9 vs. 60.0).

In the second step, we need to estimate $G(c)$ using Equation (7). A significant hurdle is estimating the distribution of covariates for each patient class. Due to selection bias, conditions at the terminal epoch cannot represent the distribution of these variables within each class. To counteract this bias, we adopt a stratification-based propensity weighting technique (Rosenbaum and Rubin 1983). The core concept is leveraging the propensity-to-treat distribution (in our case, the propensity to be extubated) as an estimator for the covariate population distribution. This approach is feasible because our dataset provides real-time patient tracking data from the moment of ICU admission. Following this strategy, we first compute the propensity score for each patient at every epoch using logistic regression, resulting in an AUC of 0.762. Then, within each class c, samples are stratified into K_c equally-sized subgroups based on their propensity scores. Given the limited data points for the most adversely affected class (i.e., class 1), we opt for five strata for class 1 and ten for the

remaining classes. Cochran (1968) established that five groups effectively eliminate 90% of the bias introduced by imbalanced covariates. Subsequently, we calculate the average $\widehat{RLOS}$ from Equation (8) for extubated patients in each class c subgroup k, denoted as $\overline{\widehat{RLOS}}_{c,k}$. Our final step is to use the weighted average of means from all subgroups within a patient class as an estimator for the terminal cost:

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

where N_c^k is the sample count 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 by following a similar two-step procedure but with a linear probability model (LPM) in the first step. We opt for LPM over logistic regression for the same reasons as using linear regression to estimate $G(c)$. In our data, the LPM produces an almost identical AUC when compared to that of logistic regression (0.775 vs. 0.776). It is worth noting that the estimated EFR is not used in the model for optimizing the extubation protocol, but rather in the numerical analysis for evaluating the performance of different extubation policies.

Estimation results are shown in Table 2. As anticipated, both the risk of extubation failure $r(c)$ and RLOS(c) decrease with patient class c. Further details on the regression models used in this section can be found in Appendix EC.9. Note that there may be some unobserved factors correlated with both LOS and patient class. Our estimation method does not completely eliminate the potential selection bias. Consequently, we provide a more comprehensive sensitivity analysis on terminal costs in Appendix EC.7.1. The results suggest that the optimal policy remains robust in the presence of estimation bias.

Table 2 Estimated RLOS and EFR.

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.

Class transition matrix. To estimate transition probabilities, we first divide the ventilation period into epochs at six-hour intervals for each patient. Subsequently, we identify patient classes at each epoch using the given RSBI values and the aforementioned classification criteria. If multiple RSBI records exist within a single epoch for a patient, we use the average value to represent the patient's condition during that epoch. Thus, 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 sample points in class i and n_{ij} represents the number of sample points in class i that transition to j at the subsequent epoch. One-step transition probabilities for each patient class are presented as follows:

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

The element-wise 95% confidence interval $\mathcal{P}$, estimated using the bootstrap method, is provided as follows:

$$\mathcal{P} = \begin{pmatrix} [0.23,0.48] & [0.10,0.32] & [0.10,0.32] & [0.04,0.20] & [0.05,0.23] \\ [0.05,0.13] & [0.34,0.47] & [0.22,0.34] & [0.08,0.17] & [0.06,0.15] \\ [0.01,0.03] & [0.11,0.17] & [0.45,0.53] & [0.20,0.27] & [0.09,0.14] \\ [0.01,0.02] & [0.04,0.06] & [0.21,0.26] & [0.33,0.38] & [0.32,0.38] \\ [0.00,0.01] & [0.01,0.02] & [0.07,0.09] & [0.17,0.20] & [0.70,0.73] \end{pmatrix}.$$

We also recognize that selection bias can influence the estimation of the transition matrix. To adjust for this bias in our sensitivity analysis, we employ the propensity score weighting method, with details available in Appendix EC.7.3. Our primary findings align with those obtained using the propensity score weighting method.

Misclassification matrix. In this section, we develop a practical predictive model using the random forest algorithm on our dataset and explore the benefit of integrating the predictions generated by this model. Time-series data from each epoch of every patient serve as sample points, while the patient class at the following epoch is set as the target. Our dataset comprises 4,376 samples, 75% of which are allocated for training, with the remaining 25% designated for testing and estimation of the misclassification matrix Q. Note that our primary aim is not to create a model with maximal predictive capacity. Rather, our goal is to showcase how a reasonably accurate predictive model can influence decision quality when its predictions are incorporated into the extubation protocol. We use all variables in Table 1 in our predictive model to forecast patients' RSBI class at the upcoming epoch. Additionally, we classify continuous variables based on their typical ranges and incorporate them as extra features. For instance, the heart rate (HR) is segmented into $HR < 60$, $60 \leq HR \leq 100$, and $HR > 100$ categories. The out-of-sample misclassification matrix $\hat{Q}$ of the predictive model is presented as follows:

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

The element-wise 95% confidence interval $\mathcal{Q}$, derived via the bootstrap method, is as follows:

$$\mathcal{Q} = \begin{pmatrix} [0.53, 0.66] & [0.05, 0.12] & [0.12, 0.22] & [0.10, 0.20] & [0.00, 0.02] \\ [0.06, 0.14] & [0.53, 0.67] & [0.11, 0.21] & [0.05, 0.13] & [0.02, 0.08] \\ [0.00, 0.03] & [0.05, 0.13] & [0.54, 0.68] & [0.16, 0.27] & [0.04, 0.12] \\ [0.00, 0.02] & [0.00, 0.04] & [0.24, 0.37] & [0.46, 0.60] & [0.09, 0.19] \\ [0.00, 0.01] & [0.00, 0.03] & [0.07, 0.15] & [0.21, 0.34] & [0.54, 0.67] \end{pmatrix}.$$

Note that the estimated matrices $\hat{P}$ and $\hat{Q}$ are not TP2. However, we will later show that the optimal policy closely resembles a switching-curve policy. Additionally, we demonstrate how to harness Theorem 1 to derive near-optimal policies that are both more interpretable and easier to implement. This involves enforcing Assumptions 1 and 2 during parameter estimation via constrained maximum likelihood estimation. A more detailed elaboration of the procedures and results can be found in Appendix EC.10.

Analogous to terminal costs, transition and misclassification matrices might be subject to estimation errors. Therefore, we undertake a sensitivity analysis to assess the influence of such errors on the optimal policy in Appendix EC.7.2. Our investigations reveal that the optimal policy remains largely resilient to estimation errors in both transition and misclassification matrices.

5. Numerical Results

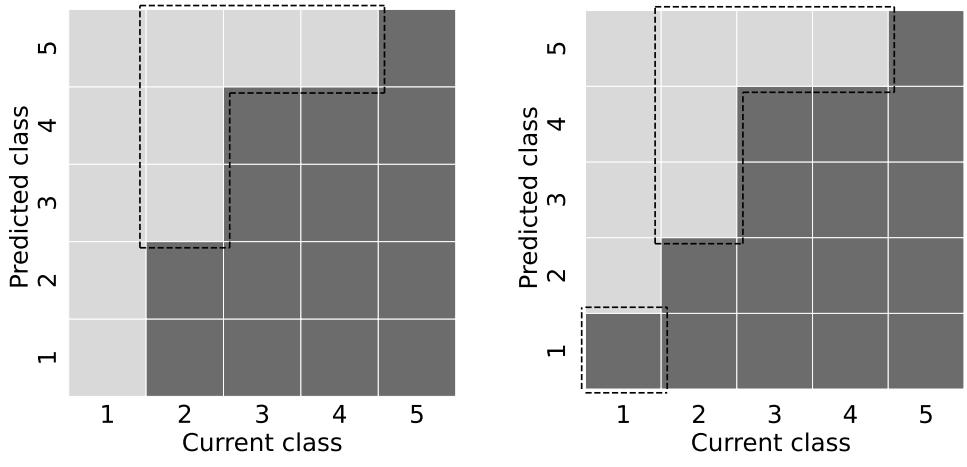
In this section, we use the parameters estimated above to determine optimal policies for extubation in various settings. We first derive the optimal policies for extubation without prediction and with perfect prediction, then compare the objectives (i.e., LOS) and the EFR for each patient class under different policies. Our results validate the value of predictive information. We also show that if the states for the next two or three periods are predictable, there is a more significant improvement in performance. Next, we assess the performance using the prediction model trained in the previous section. While the prediction accuracy is moderate, a significant performance enhancement is observed compared to the scenario without any prediction. We also propose several straightforward heuristics and guidelines to assist in applying predictive information practically. Lastly, to delve deeper into the influence of prediction accuracy on system performance, we conduct a series of numerical comparisons involving different prediction errors.

5.1. Perfect Prediction and the Value of Future Information

In this section, we first investigate the benefit of perfect prediction (i.e., $Q = I$), which provides an upper bound for the value of predictive information. With the estimated parameters at hand, we can computationally determine the optimal extubation policy without prediction (labelled as OP-B; Figure 3), as well as the optimal extubation policy with perfect predictions (labelled as OP-PP; Figure 4) by setting $Q = I$. The outcomes reveal that in the base model without prediction, the

optimal strategy is to extubate the patient whenever they progress to class 2 or a better class—analogous to a cut-off value of 105 in RSBI. This aligns with the suggestions of Yang and Tobin (1991) and McConville and Kress (2012). Conversely, if we are aware that a class 2 patient will advance to a better class by continuing ventilation for an additional epoch, it becomes optimal to continue with ventilation for this patient instead of extubation, as depicted in Figure 4, and the same holds for a class 3 patient. Figure 4 shows that the OP-PP policy at every decision epoch is contingent upon both present class data and future class information. The black dashed contours in Figure 4 highlight the patient subgroup whose extubation recommendations under the OP-PP policy diverge from those under the OP-B policy. Note that the extubation decision at the last epoch is more aggressive compared to earlier epochs due to the terminal effect. This policy can be readily adapted, allowing physicians to exercise discretion when a patient remains intubated nearing the conclusion of the decision timeline.

Figure 3 **Optimal extubation policy without prediction. The dark gray grid indicates extubation, while the light gray grid represents continued ventilation. The policy is stationary across all decision epochs.**



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

Figure 4 **Optimal extubation policy with perfect prediction. The dark gray grid indicates extubation, while the light gray grid represents continued ventilation.**

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

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

where $a_t^{B*}(\cdot)$ follows the OP-B policy and $r(\cdot)$ corresponds to 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 strategy with predictive information (for any general misclassification matrix) can be obtained from Equations (3) and (4). The corresponding EFR for a patient in class c_t at epoch t, $R_t^P(c_t)$, 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 present the current LOS and EFR for each initial class in our dataset, which are computed using sample averages. These results are summarized in Table 3.

Table 3 Performance of 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	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 prediction; OP-PP: optimal extubation policy with perfect prediction; LOS: length of stay; EFR: extubation failure rate.

From Table 3, we observe that the OP-B policy results in a longer LOS compared to current practices, confirming our earlier concern that physicians might be incorporating their own predictions into extubation decisions. However, it is worth noting that comparing current practice with our model might not be entirely fair, as real-world decisions are multifaceted, and numerous factors come into play. While we have tried to reduce estimation bias in model parameters and conducted thorough sensitivity analyses, we cannot claim that the comparison is perfectly fair. Hence, the comparison between our data and the base model performance primarily serves to illustrate the

potential pitfalls of the base model in certain situations relative to current practices. Going forward, our focus will be on the comparison between the base model and the model that incorporates predictive information.

When predictive information is considered, not only does LOS decrease relative to the OP-B policy, but EFR is also reduced. The reduction in EFR largely stems from identifying a patient subgroup that can be ventilated longer, allowing them to achieve improved conditions, which significantly lowers their EFR. Simultaneously, accurate predictions can pinpoint patients at risk of complications from prolonged ventilation, facilitating their earlier extubation. This highlights the significance of precision medicine and personalized medical decision-making. Class 2 patients, who are at the borderline between extubation and continued ventilation, experience the most benefit from predictive information, with a 3.4% reduction in LOS and, more importantly, a 20.3% reduction in EFR compared to the optimal policy that does not utilize prediction. These results partially underscore that, even without explicitly incorporating EFR into the objective of our model, predictive information enables the optimized extubation policy to significantly reduce EFR, thereby improving patients' clinical outcomes. It is logical that patients with an initial class of 5 see no benefit from predictive information since they are typically extubated shortly after ICU admission. This aligns with observations from our partner hospital, where patients in better initial respiratory health are generally put on ventilation due to needs related to surgery and anesthesia and are safely extubated soon after ICU admission. As such, we will omit results for class 5 patients in subsequent analyses.

To contextualize the reductions in ICU LOS, imagine a hospital with 3,000 ICU admissions requiring MV support annually (our study ICU is merely one of five specialty ICUs in the hospital). If we can decrease the average ICU stay by 2 hours per patient, this would be equivalent to saving 250 days annually, which is analogous to adding an additional ICU bed, assuming a 70% utilization rate. If the hospital retains its current bed count, this LOS reduction could alleviate congestion. Given that high congestion levels are correlated with poorer outcomes (KC and Terwiesch 2009, Kuntz et al. 2014), a reduced LOS can enhance patient care quality. Operationally, a shorter LOS can reduce wait times for subsequent ICU admissions. As delays can propagate in a queueing system (Luo et al. 2017), a 2-hour cut might result in substantial throughput improvements for both the ICU and the broader hospital (Kim et al. 2020).

One might expect that if a predictive model could forecast patient conditions beyond a six-hour window, further LOS reductions could be achieved. We next investigate the effect of extending the prediction window from one to two or even three epochs—i.e., from 6 hours to 12 or 18 hours. Table 4 shows that both LOS and EFR can be further reduced with extended predictive windows. Echoing the effect of diminishing returns, we notice that while there are benefits to longer prediction

windows, the primary benefits are captured within a single epoch. Another contributing factor might be health transitions: if a patient's condition improves within a few epochs, then acquiring extended future predictions might not add significant value. This implies that, in practice, the DM might prioritize enhancing predictions for just the subsequent epoch.

Table 4 Performance of the optimal extubation policy with perfect prediction under longer prediction windows.

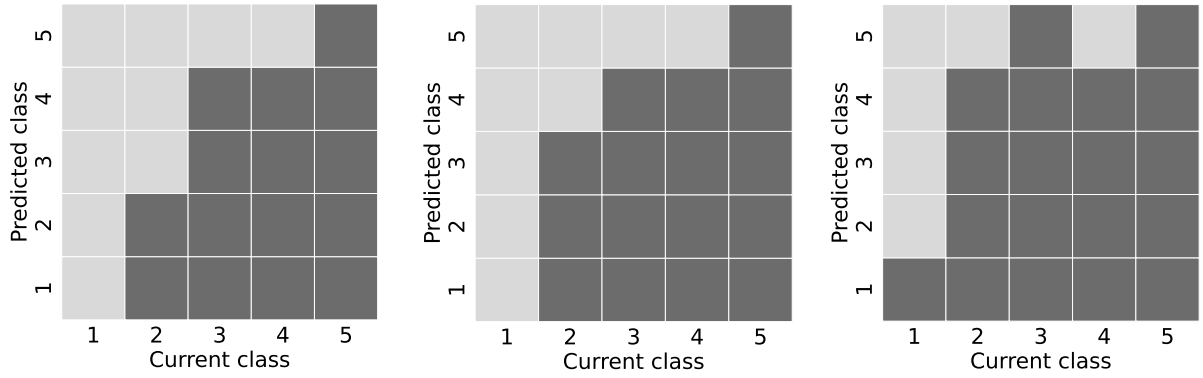
Initial class	2-epoch prediction		3-epoch prediction	
	LOS (hr)	EFR (%)	LOS (hr)	EFR (%)
1	104.4	19.1	104.4	19.1
2	101.4	23.1	101.3	22.6
3	96.4	19.7	96.4	17.7
4	93.3	16.3	93.3	16.3

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

5.2. A Practical Predictive Model and Imperfect Prediction

In the aforementioned analysis, our comparison of a model with perfect prediction to MDP-B provides an upper limit on the value of predictive information. Yet, in the real world, predictions are seldom entirely accurate. Employing the estimated misclassification matrix, $\hat{Q}$, from Section 4.3, we derive the optimal policy for a model with imperfect prediction as illustrated in Figure 5. Interestingly, at decision epoch 27, the optimal policy deviates from a threshold structure. Specifically, for a class 3 patient predicted to transition to class 5 (i.e., the best condition) at the next epoch, extubation is recommended despite anticipated health improvements. This deviation arises since the accuracy of a class 3 patient predicted to be class 5 at the next epoch is much lower than that of a class 4 patient predicted to be class 5 at the next epoch (0.48 vs. 0.75). Thus, the policy appears skeptical about the likelihood of a class 3 patient transitioning to class 5 at the next epoch and recommends extubation. Note that this does not contravene our analytical findings, as neither the estimated misclassification matrix nor the transition probability matrix satisfies the TP2 condition in our dataset (i.e., Assumptions 1 and 2). As such, a threshold policy is not assured. We will delve deeper into this in Section EC.10, proposing a new strategy to derive a more pragmatically implementable policy by leveraging the analytical results from Section 4.1. Nonetheless, this behavior can also be attributed to terminal effects, implying its implications in real-world practices might be minimal.

Table 5 presents the performance of the optimal extubation policy with imperfect predictions, labeled as OP-IP. We observe that the benefit of predictive information is reduced when predictions are imperfect. Nevertheless, LOS and EFR still show improvement compared to the model without predictive information. Next, as discussed in Section 4.2, we consider the scenario where the DM



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

Figure 5 Optimal extubation policy with imperfect prediction. The dark gray grid indicates extubation, while the light gray grid represents continued ventilation.

is unaware of the misclassification matrix and naïvely applies the optimal policy derived from assuming perfect predictions (i.e., the OP-PP policy). Although the OP-PP policy is not optimal in this context, its performance is remarkably close to that of the optimal OP-IP policy, with differences being marginal within our reporting accuracy. This outcome suggests that our prediction model is considerably accurate, rendering the loss from overlooking its true misclassification matrix negligible. This result confirms our findings in Theorem 4.

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

Initial class	OP-IP		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 derived from assuming perfect prediction (i.e., $Q = I$); LOS: length of stay; EFR: extubation failure rate.

5.3. Rule of Thumb

A defining characteristic of optimal extubation policies with predictive information is that if a patient is predicted to benefit from continued intubation, extubation is not pursued. Conversely, without predictive information, the patient would be extubated. Stemming from this observation, we aim to determine the threshold of prediction accuracy at which the decision shifts from extubating to continuing the ventilation. To investigate this, we modify a specific entry in the misclassification matrix related to predicting transitions to a better class. By gradually increasing this entry

while randomizing the others, we aim to determine the point at which the optimal policy begins to shift. Specifically, we first focus on the entry $Q(3,3)$, representing the probability of correctly predicting a transition to class 3 at the next epoch. We generate values for $Q(3,3)$ by drawing from a uniform distribution over $[0,1]$. The remaining entries in the third row of Q are regenerated via a uniform distribution over $[0,1]$ and are normalized such that their sum is $1-Q(3,3)$. The rest of the misclassification matrix is then generated from an uncertainty set constructed by element-wise confidence intervals, as detailed in Appendix EC.7.2. This procedure is repeated 50,000 times, with the matrices arranged in ascending order based on the values of $Q(3,3)$, facilitating the analysis of policy shifts correlated with its variations. Similar analyses are conducted for $Q(4,4)$ and $Q(5,5)$.

Given that the optimal policies are time-dependent, providing a summary of policy changes for all decision epochs can become convoluted. Thus, we focus on decisions made within the first 10 epochs (i.e., 60 hours), since only 90 patients (2.9% of our dataset) were subjected to ventilation beyond 10 decision epochs. Based on our observations, we propose the following rules of thumb to leverage predictive information:

- For patients predicted to improve by one class, continuing the ventilation is recommended if the prediction accuracy exceeds 0.65, except for class 3 patients.
- For patients predicted to improve by two classes, continuing the ventilation is recommended if the prediction accuracy exceeds 0.5.

From our simulations, we observe that 100% of the optimal policies with prediction accuracy beyond these thresholds switch from recommending extubation to recommending continued intubation when patients are predicted to improve under ventilation. This guidance also extends to physicians' own prognostications, with accuracy aligning with their confidence levels.

Note that the proposed heuristic emerges from the model parameters we estimated and might not translate seamlessly to other contexts due to potential estimation discrepancies and varying environments. Hence, we encourage practitioners to recalibrate the parameters and modify the guidelines to suit their specific context.

5.4. Impact of Prediction Accuracy and Validation of Analytical Results

The accuracy of predictions hinges on both data availability and the efficacy of a predictive model. We carry out a sensitivity analysis to investigate the influence of prediction accuracy on outcomes. We set the diagonal elements of the misclassification matrix $Q(i,i)$ to values ranging from 0.1 to 0.9, incrementing by 0.1, and then randomize the off-diagonal entries. Recall that the i^{th} diagonal element of Q represents the probability of correctly predicting a patient's future state as class i when they indeed transition to class i. Hence, the diagonal elements $Q(i,i)$ capture the model's predictive accuracy. For each accuracy level, we generate 1,000 instances of the misclassification

Table 6 Performance of the optimal extubation policy with imperfect prediction under different prediction accuracy.

Accuracy	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.

matrix (with random off-diagonal elements) and execute our model to compute the average LOS and EFR. The findings are presented in Table 6.

From Table 6, we can observe that the model with imperfect predictive information outperforms the model without prediction, even if the prediction accuracy is moderate. The misclassification matrix in our model enables 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. The Bayesian update allows the value of even noisy prediction to be extracted to improve decision quality.

In some situations, the DM might not be aware of the misclassification matrix or may use it solely as an indicator of whether to believe the predictions. Consequently, decisions might be made based purely on the predictions as if they were perfectly accurate. To probe into the implications of this, we adopt the aforementioned procedure to evaluate whether the policy derived from assuming perfect prediction (i.e., the OP-PP policy) remains beneficial even if the prediction is imperfect. The results are presented in Table 7. The insights from the table suggest that, in cases of low prediction accuracy, the model without predictive information might yield better outcomes mainly in terms of LOS compared to when overtrusting the predictions. However, with moderate prediction accuracy, the optimal policy derived from assuming perfect prediction can surpass the OP-B policy in an environment of imperfect prediction. This complements the conclusions drawn in Theorem 4 for two patient classes.

In Remark 2, we note that for two patient classes, if predictive information is proven beneficial, then the optimal policy derived from assuming perfect prediction remains beneficial relative to the optimal policy without prediction. We would like to explore to what extent this insight remains valid

Table 7 **Performance of the optimal extubation policy derived from assuming perfect prediction under different actual prediction accuracy.**

Accuracy	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.

in multi-class scenarios. To achieve this, we randomly generate 50,000 misclassification matrices from the uncertainty set of Q and compute the corresponding outcomes from different policies. Details of constructing the uncertainty set can be found in Appendix EC.7.2. The results show that when the average benefit (OP-IP vs. OP-B) in terms of LOS for patients in classes 1, 2, 3, and 4 exceeds 0.5 hours (45,334 out of 50,000 sampled Q matrices), or when the benefit for class 1 patients exceeds 0.6 hours (45,812 out of 50,000 sampled Q matrices), the OP-PP policy can also outperform the OP-B policy in more than 95% of these instances. Hence, we suggest that if the predictive information can reduce the LOS for a patient by at least 0.5 hours, viewing the prediction model as perfectly accurate is not detrimental to decision quality in our case.

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 predictions of future states, 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.

To validate the proposed framework, we apply our models to the extubation problem in ICUs and calibrate the models using a comprehensive dataset 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 patients whose initial condition is poor. In contrast, optimal policies from the benchmark model without predictive information might 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 predict a patient's condition at the next decision epoch and measure the quality of the prediction using an out-of-sample misclassification matrix. Although the accuracy of the 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 accuracy 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 overlooked by applying the policy derived from assuming perfect prediction. In both scenarios, we observe that predictive information can still enhance decision quality, even with a moderately accurate predictive model.

Our research can be extended in several directions. First, our proposed framework focuses on integrating predictive information and does not consider many operational aspects, such as capacity rationing issues under stochastic demands. Future research could incorporate more operational aspects into the analysis. Second, this paper studies the optimal stopping problem for medical treatment. In practice, deciding when to initiate treatment, and how to control medication dosage and intensity, as well as other modalities during treatment, is also critical. 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, we demonstrate numerically that the value of predictive information grows as the prediction window increases. Nevertheless, conducting a theoretical analysis becomes considerably more challenging with an extended prediction window. Fourth, as mentioned in the literature review, developing robust MDP models to capture parameter uncertainties in both transition and misclassification matrices would be beneficial in obtaining robust extubation policies with predictive information. Fifth, it is important to note that all extubated patients in our study remained in the ICU post-extubation. In contrast, other ICUs may transfer patients to a transitional care unit. Hence, it would be interesting to extend our model to include a broader range of post-extubation pathways, considering that prediction accuracy might vary depending on the options taken. Extending our model to capture more actions, where the prediction accuracy under different actions may vary, would be insightful.

Sixth, our model assumes that extubation decisions for different patients are independent, which may not always be the case in reality. Indeed, system-level factors can influence individual-level decisions. However, incorporating system-level factors explicitly into the state could result in a high-dimensional MDP. We provide a partial solution to implicitly incorporate the impact of utilization in Appendix EC.8. Exploring ways to enhance the model by integrating system-level factors more explicitly into the decision-making process would be valuable. Finally, randomized controlled trials are warranted to validate the findings in this paper and to establish protocols that incorporate predictive information. We leave these and other challenges for 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 (2020) From predictive to prescriptive analytics. *Management Science* 66(3):1025–1044.

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.

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* 22(5):1066–1087.

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

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 J, Dong J, Shi P (2023) Optimal routing under demand surges: The value of future arrival rates. *Operations Research* .

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 (2022) Interpretable optimal stopping. *Management Science* 68(3):1616–1638.

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

Dai JG, Shi P (2019) Inpatient overflow: An approximate dynamic programming approach. *Manufacturing & Service Operations Management* 21(4):894–911.

Dai JG, Shi P (2021) Recent modeling and analytical advances in hospital inpatient flow management. *Production and Operations Management* 30(6):1838–1862.

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.

Givan R, Leach S, Dean T (2000) Bounded-parameter markov decision processes. *Artificial Intelligence* 122(1-2):71–109.

Goh J, Bayati M, Zenios SA, Singh S, Moore D (2018) Data uncertainty in Markov chains: application to cost-effectiveness analyses of medical innovations. *Operations Research* 66(3):697–715.

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

Goyal V, Grand-Clément J (2022) Robust Markov decision processes: beyond rectangularity. *Mathematics of Operations Research* .

Grand-Clément J, Chan C, Goyal V, Chuang E (2021) Interpretable machine learning for resource allocation with application to ventilator triage. *arXiv preprint arXiv:2110.10994* .

Grand-Clément J, Chan CW, Goyal V, Escobar G (2020) Robust policies for proactive ICU transfers. *arXiv preprint arXiv:2002.06247* .

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 (2019) Predictive and prescriptive analytics for location selection of add-on retail products. *Production and Operations Management* 28(7):1858–1877.

Iancu DA, Trichakis N, Yoon DY (2020) Monitoring with limited information. *Management Science* 67(7):4233–4251.

Iyengar GN (2005) Robust dynamic programming. *Mathematics of Operations Research* 30(2):257–280.

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

Kc DS (2020) Heuristic thinking in patient care. *Management Science* 66(6):2545–2563.

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, Chan CW, Olivares M, Escobar GJ (2016) Association among ICU congestion, ICU admission decision, and patient outcomes. *Critical care medicine* 44(10):1814–1821.

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 (2017) 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).

Nilim A, El Ghaoui L (2005) Robust control of Markov decision processes with uncertain transition matrices. *Operations Research* 53(5):780–798.

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.

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.

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.

Shen Y, Chan CW, Zheng F, Argenziano M, Kurlansky P (2021) The impact of surgeon daily workload and its implications for operating room scheduling .

Shi P, Helm JE, Deglise-Hawkinson J, Pan J (2021) Timing it right: Balancing inpatient congestion vs. readmission risk at discharge. *Operations Research* 69(6):1842–1865.

Shrank WH, Rogstad TL, Parekh N (2019) Waste in the us health care system: estimated costs and potential for savings. *Jama* 322(15):1501–1509.

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.

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.

van Dijk N, Kortbeek N (2009) Erlang loss bounds for OT–ICU systems. *Queueing Systems* 63:253–280.

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.

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

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):1004–1011.

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.

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.

Zhang Y, Steimle LN, Denton BT (2017) Robust Markov decision processes for medical treatment decisions. *Optimization Online* .

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. Benchmark Model without Predictive Information

In the base scenario, we address the extubation problem without predictive information, approaching it in a conventional manner. This decision problem is formulated as a finite-state, finite-horizon, discrete-time MDP model, which we denote as the MDP-B model. Although it shares many components with the MDP-P model, they differ in state and transition probabilities. In this setting, the system state is solely defined by the patients' current class, so the transition probabilities are entirely characterized by matrix P.

The optimality equation for MDP-B is given by

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

where $J_t^B(c_t)$ represents the minimum expected total cost for a class c_t patient at epoch t. The boundary condition is

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

To solve the optimality equation (EC.1) and derive the optimal policy for the MDP-B model, we begin with certain assumptions, which are deemed reasonable in a healthcare context. The first assumption concerns 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, this suggests that $\boldsymbol{x}$ is more favorable than $\boldsymbol{y}$.

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

Assumption EC.1 defines the structure of the transition probability matrix P in the sense of first-order stochastic dominance. It indicates that a healthier patient in class j is more likely to transition to a better state compared to a less healthy patient in class i, given $i, j \in \mathcal{C}$ and $i < j$.

The next assumption links the transition probability matrix to the cost functions. For a class c patient, $\sum_{c'\in\mathcal{C}} P(c, c')G(c')$ denotes the expected terminal cost after one more epoch of treatment. Thus, the expression $G(c) - \sum_{c'\in\mathcal{C}} P(c, c')G(c') - H$ quantifies the net benefit of an additional treatment epoch. The following assumption asserts that the net benefit from one more epoch of treatment declines as the patient's health improves.

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

In other words, a healthier patient gets less benefit from one more epoch 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. Technical proofs of all the results in this appendix are presented in Appendix EC.2.

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

Lemma EC.1 implies that a healthier patient can expect a reduced residual treatment cost, after undergoing t epochs of treatment. 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^* for each decision epoch $t \in \mathcal{T}\backslash\{T\}$ in 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. Furthermore, c_t^* decreases as t increases, i.e., $c_1^* \geq c_2^* \geq \cdots \geq c_{T-1}^*$.*

Theorem EC.1 states that the MDP-B model's optimal policy is a threshold policy. For each epoch $t \in \mathcal{T}\backslash\{T\}$, there exists an optimal class threshold. If a patient's class c_t is equal to or better than this optimal stopping threshold c_t^*, then it is optimal to terminate the treatment; otherwise, it is optimal to continue with it. The monotonicity of the optimal threshold c_t^* in time t indicates a more lenient stopping criterion for patients who have undergone multiple treatment epochs. In the following proposition, we discuss a special case in which the optimal threshold remains stationary over time.

PROPOSITION EC.1. *For a subgroup of patients who consistently improve under treatment (i.e., $P(i,j) = 0$ if $j < i$ for $i, j \in \mathcal{C}$), the optimal threshold from Theorem EC.1 remains invariant with time. That is, $c_1^* = c_2^* = \cdots = c_{T-1}^* = c^*$. More explicitly, the optimal threshold is defined as*

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

Proposition EC.1 suggests that a consistent stopping criterion can be applied to a patient subgroup that consistently transitions to healthier states under treatment. This is invaluable, especially since, under appropriate treatments, the majority of patients recover. In our dataset, 2,323 patients (76%) never experienced any deterioration during ventilation. Even though some patients deteriorated while on ventilation, the optimal policy derived from the MDP-B model in our case study remained a stationary threshold policy.

EC.2. Proofs of the Mathematical Results

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 decreasing in c_T since $G(c_T)$ is decreasing in c_T. Next, let's assume that for epoch $t+1$, $J_{t+1}^B(c_{t+1})$ is decreasing in c_{t+1}. At epoch t, based on Equation (EC.1), we have

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

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

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

for any C-dimensional column vector $\boldsymbol{g}$ with decreasing components, i.e., $g_1\leq g_2\leq\cdots\leq g_C$ (Müller and Stoyan 2002). Given the induction hypothesis, we infer that $\sum_{c_{t+1}\in\mathcal{C}}P(c_t,c_{t+1})J_{t+1}^B(c_{t+1})$ is decreasing in c_t. Since both $G(c_t)$ and H decrease in c_t, it follows that $J_t^B(c_t)$ is decreasing in c_t, thus completing the proof.

EC.2.2. Proof of Theorem EC.1

To prove Theorem EC.1, we first define the function to represent the benefit of continuing the treatment under the base model as follows:

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

It is optimal to stop the treatment if $h_t^B(c_t)\leq 0$, and to continue otherwise. This theorem can then be recast into the following statement:

(i) $h_t^B(c_t)$ decreases with 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) suggests that the optimal threshold c_t^* decreases with t. We now turn to proving these statements.

We begin by proving part (i) using induction on the decision epoch t. For the decision epoch $T-1$, we have

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

The first equality follows from Equation (EC.4), and the second one is derived from the boundary condition given by Equation (EC.2). Based on Assumption EC.2, it can be inferred that $h_{T-1}^B(c_{T-1})$ decreases as c_{T-1} increases.

Assume that for epoch $t+1$, $h^B_{t+1}(c_{t+1})$ decreases as c_{t+1} increases. Consequently, there exists an optimal threshold c^*_{t+1} for epoch $t+1$. When $c_{t+1} \geq c^*_{t+1}$, we have $J^B_{t+1}(c_{t+1}) = G(c_{t+1})$. Otherwise, $J^B_{t+1}(c_{t+1}) = G(c_{t+1}) - h^B_{t+1}(c_{t+1})$. For epoch t, we have

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

The first equality follows from Equation (EC.4), and the second one is due to the induction hypothesis. From Assumption EC.1 combined with the induction hypothesis, we note that $\sum_{c_{t+1}<c^*_{t+1}} P(c_t, c_{t+1}) h^B_{t+1}(c_{t+1})$ decreases as c_t increases. Moreover, due to Assumption EC.2, $G(c_t) - H - \sum_{c_{t+1}\in\mathcal{C}} P(c_t, c_{t+1}) G(c_{t+1})$ decreases as c_t increases. Consequently, $h^B_t(c_t)$ decreases with an increase in c_t, which concludes the proof for part (i).

Moving on to prove part (ii), we begin with the following claim.

Claim EC.1. *For all $c \in \mathcal{C}$ and $t \in \mathcal{T}\backslash\{T\}$, $J^B_t(c) \leq J^B_{t+1}(c)$.*

Proof. To prove this claim, we use induction on the decision epoch t. For $t = T-1$, we have

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

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

Assume that the claim is valid for epoch $t+1$. For epoch t, we have

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

This inequality follows from the induction hypothesis. Consequently, $J^B_t(c)$ increases with t. ■

With the above established, we can now proceed to prove part (ii). We have

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

This inequality follows from Claim EC.1. Thus, part (ii) is established, completing the proof for Theorem EC.1.

EC.2.3. Proof of Proposition EC.1

As presented in Proposition EC.1, we define

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

To prove this proposition, we need to show that $h_t^B(c_t) \le 0$ if $c_t \ge c^*$, 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

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

The first equality follows from the definition of $h_t^B(c_t)$, whereas the second is a result of the boundary condition of the MDP-B model. Clearly, based on the definition of c^*, we find that $h_{T-1}^B(c_{T-1}) \le 0$ when $c_{T-1} \ge c^*$ and $h_{T-1}^B(c_{T-1}) > 0$ otherwise. Assume the condition holds true for epoch $t+1$. Then, for $c_{t+1} \ge c^*$, we have $J_{t+1}^B(c_{t+1}) = G(c_{t+1})$. Otherwise, $J_{t+1}^B(c_{t+1}) = G(c_{t+1}) - h_{t+1}^B(c_{t+1})$. At epoch t, we have

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

The second equality follows from the induction hypothesis. We can consider the following two cases.
(a) If $c_t \ge c^*$, according to Equation (EC.5), we have $G(c_t) - H - \sum_{c_{t+1} \in \mathcal{C}} P(c_t, c_{t+1}) G(c_{t+1}) \le 0$. Also, since $P(c_t, c_{t+1}) = 0$ when $c_{t+1} < c_t$ in this particular case, $\sum_{c_{t+1} < c^*} P(c_t, c_{t+1}) h_{t+1}^B(c_{t+1}) = 0$ As a result, $h_t^B(c_t) \le 0$ for $c_t \ge c^*$.
(b) If $c_t < c^*$, according to Equation (EC.5), we have $G(c_t) - H - \sum_{c_{t+1} \in \mathcal{C}} P(c_t, c_{t+1}) G(c_{t+1}) > 0$. Furthermore, according to our induction hypothesis, $\sum_{c_{t+1} < c^*} P(c_t, c_{t+1}) h_{t+1}^B(c_{t+1}) \ge 0$. Thus, $h_t^B(c_t) > 0$ when $c_t < c^*$.

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

EC.2.4. Proof of Lemma 1

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

Claim EC.2. *Given the 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 $[\pi(c_{t+1} \mid c_t, \hat{c}_{t+1})]_{c_{t+1} \in \mathcal{C}} \ge_{lr} [\pi(c_{t+1} \mid c_t, \hat{c}'_{t+1})]_{c_{t+1} \in \mathcal{C}}$, where $[\pi(c_{t+1} \mid c_t, \cdot)]_{c_{t+1} \in \mathcal{C}}$ represents the belief vector of dimension C. Similarly, given the 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, $[\pi(c_{t+1} \mid c_t, \hat{c}_{t+1})]_{c_{t+1}\in\mathcal{C}} \geq_{lr} [\pi(c_{t+1} \mid c_t, \hat{c}'_{t+1})]_{c_{t+1}\in\mathcal{C}}$ is equivalent to

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

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

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

That is,

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

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

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

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, so we omit it. ■

We then prove that $J_t^P(c_t, \hat{c}_{t+1})$ is decreasing in both c_t and $\hat{c}_{t+1}$ by induction on the decision epoch t. First, when $t = T$, we have $J_T^P(c_T, \hat{c}_{T+1}) = G(c_T)$. The lemma immediately holds.

Assume that $J_{t+1}^P(c_{t+1}, \hat{c}_{t+2})$ is decreasing 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 decreasing 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 decreasing 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 decreasing in both c_t and $\hat{c}_{t+1}$. $G(c_t)$ and H are assumed as decreasing in c_t. Therefore, $J_t^P(c_t, \hat{c}_{t+1})$ is decreasing in both c_t and $\hat{c}_{t+1}$.

EC.2.5. Proof of Theorem 1

We first define the function to represent the benefit of continuing the treatment under the MDP-P model as follows:

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

Theorem 1 is equivalent to the following statements:

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

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

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

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

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

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

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

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

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

Before proving part (iii), we establish the following result.

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 the decision epoch t. When $t = T-1$, we have

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

The first equality follows the definition, and the third follows from the boundary condition. Assume that the property holds for epoch $t+1$. At 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 demonstrate that $h_t^P(c,\hat{c})$ is decreasing with respect to 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. Thus, the proof is completed.

EC.2.6. Proof of Theorem 2

We prove this theorem by induction on the 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 equations follow from the definitions and the boundary condition. Thus, the proposition holds for epoch T. Next, assuming that the statement holds for epoch $t+1$, i.e., $J_{t+1}^B(c_{t+1}) \geq J_{t+1}^P(c_{t+1})$, $\forall c_{t+1}\in\mathcal{C}$. For epoch t, we have

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

The first line follows from the optimality equation. The second line is a result of the induction hypothesis. The third line derives from the definitions of $\tilde{Q}$ and π. The fourth line follows from Jensen's inequality. The fifth line uses the optimality equation, while the last line follows from the definition of $J_t^P(c_t)$. Thus, Theorem 2 holds. The proof is completed.

EC.2.7. Proof of Theorem 3

Note that if patients are in class 2, they will be extubated, as no further benefit can be gained from continued ventilation. Consequently, we need only consider the cost for class 1 patients. The optimal decision for class 1 patients, when there is no predictive information, depends solely on their current condition. This can be simplified to a comparison of the costs of extubation and continued ventilation for class 1, i.e., $G(1)$ vs. $G(2)+H/(1-p_1)$. Here, $G(2)+H/(1-p_1)$ represents the expected cost of not extubating a class 1 patient, or in other words, extubating only when patients reach class 2. The first term represents the terminal cost of extubation in class 2, and the second term is the expected holding cost of transitioning from class 1 to class 2, represented by the mean of a geometric distribution with success probability $1-p_1$. Comparing the two policies, we deduce that if $H/\Delta G > 1-p_1$, that is, the holding cost is significantly larger relative to the cost difference between extubating in class 1 and class 2, then the optimal action is to extubate in class 1. Otherwise, continuing ventilation is optimal. Thus, the optimal cost can be formulated as

$$J^B(1) = \begin{cases} G(2)+H/(1-p_1) & \text{if } H/\Delta G \le 1-p_1, \\ G(1) & \text{if } H/\Delta G > 1-p_1. \end{cases}$$

There are four potential policies when considering predictive information: $a(1,1)=a(1,2)=1$; $a(1,1)=1$, $a(1,2)=0$; $a(1,1)=0$, $a(1,2)=1$; and $a(1,1)=a(1,2)=0$.

Under the first policy, where $a(1,1)=a(1,2)=1$, meaning all class 1 patients are extubated regardless of predictive information, the expected cost is $J^P(1)=G(1)$.

For the second policy, $a(1,1)=1$, $a(1,2)=0$, where class 1 patients who are predicted to remain in class 1 are extubated and those who are predicted to transition to class 2 continue with ventilation, the expected cost is

$$\begin{aligned} J^P(1) &= \left[\gamma\sum_{i=0}^{\infty}\lambda^i\right]G(1) + \left[1-\gamma\sum_{i=0}^{\infty}\lambda^i\right]G(2) + (1-\gamma)\left[\sum_{i=0}^{\infty}\lambda^i\right]H \\ &= \frac{\gamma}{1-\lambda}G(1) + \frac{1-\lambda-\gamma}{1-\lambda}G(2) + \frac{1-\gamma}{1-\lambda}H, \end{aligned}$$

where

$$\gamma = \tilde{Q}(c_t=1,\hat{c}_{t+1}=1) = p_1q_1 + (1-p_1)(1-q_2) \text{ and}$$

$$\lambda = \tilde{Q}(c_t=1,\hat{c}_{t+1}=2)\pi(c_{t+1}=1 \mid c_t=1,\hat{c}_{t+1}=2) = p_1(1-q_1).$$

For the third policy, $a(1,1)=0$, $a(1,2)=1$, where class 1 patients who are predicted to transition to class 2 are extubated, while those who are predicted to stay in class 1 continue with ventilation, the expected cost is

$$J^P(1) = \left[(1-\gamma)\sum_{i=0}^{\infty}(p_1q_1)^i\right]G(1) + \left[1-(1-\gamma)\sum_{i=0}^{\infty}(p_1q_1)^i\right]G(2) + \gamma\left[\sum_{i=0}^{\infty}(p_1q_1)^i\right]H$$

$$= \frac{1-\gamma}{1-p_1q_1}G(1) + \frac{\gamma - p_1q_1}{1-p_1q_1}G(2) + \frac{\gamma}{1-p_1q_1}H.$$

Under the fourth policy, $a(1,1) = a(1,2) = 0$, where all class 1 patients continue with ventilation irrespective of predictive information, the expected cost is

$$\begin{aligned} J^P(1) &= G(2) + \sum_{i=0}^{\infty} p_1^i H \\ &= G(2) + \frac{1}{1-p_1}H. \end{aligned}$$

Note that the first and fourth policies are the same as the policies without predictive information. In what follows, we focus on the second and third policies that utilize predictive information. First, we identify the condition under which the second policy, $a(1,1) = 1, a(1,2) = 0$, is strictly better than the other three policies. This is equivalent to the following inequalities:

$$\frac{\gamma}{1-\lambda}G(1) + \frac{1-\lambda-\gamma}{1-\lambda}G(2) + \frac{1-\gamma}{1-\lambda}H < G(1), \tag{EC.6}$$

$$\frac{\gamma}{1-\lambda}G(1) + \frac{1-\lambda-\gamma}{1-\lambda}G(2) + \frac{1-\gamma}{1-\lambda}H < G(2) + \frac{1}{1-p_1}H, \tag{EC.7}$$

$$\frac{\gamma}{1-\lambda}G(1) + \frac{1-\lambda-\gamma}{1-\lambda}G(2) + \frac{1-\gamma}{1-\lambda}H < \frac{1-\gamma}{1-p_1q_1}G(1) + \frac{\gamma - p_1q_1}{1-p_1q_1}G(2) + \frac{\gamma}{1-p_1q_1}H. \tag{EC.8}$$

The first two inequalities are equivalent to

$$\begin{aligned} \frac{H}{\Delta G} &< \frac{1-\lambda-\gamma}{1-\gamma} = \kappa_2, \\ \frac{H}{\Delta G} &> \frac{(1-p_1)\gamma}{p_1q_1 + (1-p_1)\gamma} = \kappa_1. \end{aligned}$$

We show that the third inequality is implied by the first two. To demonstrate this, it suffices to show

$$\frac{1-\gamma}{1-p_1q_1}G(1) + \frac{\gamma - p_1q_1}{1-p_1q_1}G(2) + \frac{\gamma}{1-p_1q_1}H > G(1),$$

which is equivalent to

$$\frac{H}{\Delta G} > \frac{\gamma - p_1q_1}{\gamma}.$$

Note that

$$\kappa_1 - \frac{\gamma - p_1q_1}{\gamma} = \frac{(1-p_1)p_1^2q_1(q_1+q_2-1)}{\gamma\left[p_1q_1 + (1-p_1)\gamma\right]} \geq 0.$$

The inequality holds by Assumption 2, that is, $q_1 + q_2 \geq 1$. Thus, when inequalities (EC.6) and (EC.7) hold, inequality (EC.8) also holds. Therefore, the second policy is the unique optimal policy if and only if

$$\kappa_1 < \frac{H}{\Delta G} < \kappa_2. \tag{EC.9}$$

Also note that

$$\kappa_2 - \kappa_1 = p_1(1-p_1)(1-\lambda)(q_1+q_2-1) \geq 0.$$

This inequality holds by Assumption (2). Since we assume $0 < p_1 < 1$, condition (EC.9) is an empty set if and only if $q_1 + q_2 = 1$.

Now we consider the conditions under which the third policy, $a(1,1) = 0, a(1,2) = 1$, is strictly better than the other three policies. This is equivalent to the following inequalities:

$$\begin{aligned}
&\frac{1-\gamma}{1-p_1q_1}G(1) + \frac{\gamma - p_1q_1}{1-p_1q_1}G(2) + \frac{\gamma}{1-p_1q_1}H < G(1),\\
&\frac{1-\gamma}{1-p_1q_1}G(1) + \frac{\gamma - p_1q_1}{1-p_1q_1}G(2) + \frac{\gamma}{1-p_1q_1}H < G(2) + \frac{1}{1-p_1}H,\\
&\frac{1-\gamma}{1-p_1q_1}G(1) + \frac{\gamma - p_1q_1}{1-p_1q_1}G(2) + \frac{\gamma}{1-p_1q_1}H < \frac{\gamma}{1-\lambda}G(1) + \frac{1-\lambda-\gamma}{1-\lambda}G(2) + \frac{1-\gamma}{1-\lambda}H.
\end{aligned}$$

The first two inequalities are equivalent to

$$\begin{aligned}
&\frac{H}{\Delta G} < \frac{\gamma - p_1q_1}{\gamma},\\
&\frac{H}{\Delta G} > \frac{(1-p_1)(1-\gamma)}{1-p_1q_1-(1-p_1)\gamma}.
\end{aligned}$$

However, we have

$$\frac{\gamma - p_1q_1}{\gamma} - \frac{(1-p_1)(1-\gamma)}{1-p_1q_1-(1-p_1)\gamma} = \frac{p_1(1-p_1)(q_1+q_2-1)(p_1q_1-1)}{\gamma\left[1-p_1q_1-(1-p_1)\gamma\right]} \leq 0.$$

This inequality holds because $p_1 < 1$, $q_1 \leq 1$, and $q_1 + q_2 \geq 1$ (i.e., Assumption 2). Thus, under Assumption 2, these inequalities cannot hold simultaneously, which indicates that the third policy can never be strictly better than the other policies.

By combining the above analyses of the two policies that utilize predictive information, we conclude that predictive information is strictly beneficial if and only if condition (EC.9) holds. Furthermore, condition (EC.9) can be expressed in alternative forms by incorporating the definitions of γ and λ as follows:

$$\frac{H}{\Delta G} > \frac{1-p_1}{2-p_1} \quad \text{and} \quad q_1 > \max\left\{\frac{(1-p_1)^2(1-q_2)\left(1-\frac{H}{\Delta G}\right)}{p_1\frac{H}{\Delta G} - p_1(1-p_1)\left(1-\frac{H}{\Delta G}\right)},\ 1 - \frac{(1-p_1)\left(1-\frac{H}{\Delta G}\right)q_2}{p_1\frac{H}{\Delta G}}\right\},$$

or

$$\frac{H}{\Delta G} > \frac{1-p_1}{2-p_1} \quad \text{and} \quad q_2 > \max\left\{\frac{\left(1-\frac{H}{\Delta G}\right)\left[(1-p_1)^2 + p_1(2-p_1)q_1\right] - p_1q_1}{(1-p_1)^2\left(1-\frac{H}{\Delta G}\right)},\ \frac{p_1(1-q_1)\frac{H}{\Delta G}}{(1-p_1)\left(1-\frac{H}{\Delta G}\right)}\right\}.$$

Part (a) of the theorem is thus proven. Moreover, the second policy is the preferable way to utilize the predictive information (compared to the third policy) when the predictive information proves beneficial, given Assumption 2.

Next, we examine how the benefit of predictive information varies with q_1 and q_2. Since policies without predictive information do not account for q_1 or q_2, we only need to determine how the cost function of the policies with predictive information is affected by these two parameters. Observe that condition (EC.9) can be expressed as continuous intervals with respect to q_1 or q_2. Thus, we can derive the cost function to see how it changes with respect to q_1 and q_2 when the conditions are met.

Under condition (EC.9), we have

$$J^P(1) = \frac{\gamma}{1-\lambda}G(1) + \frac{1-\lambda-\gamma}{1-\lambda}G(2) + \frac{1-\gamma}{1-\lambda}H,$$

which is simplified to

$$J^P(1) = \frac{\gamma}{1-\lambda}\Delta G + G(2) + \frac{1-\gamma}{1-\lambda}H = \frac{\Delta G - H}{1-\lambda}\gamma + \frac{H}{1-\lambda} + G(2).$$

Recall that $\gamma = p_1q_1 + (1-p_1)(1-q_2)$, and $\lambda = p_1(1-q_1)$. It is immediately apparent that $J^P(1)$ increases linearly with γ, and therefore decreases linearly with q_2. Hence, the benefit of predictive information increases linearly with q_2. To understand the relationship with q_1, we differentiate with respect to q_1:

$$\begin{aligned}
\frac{\partial J^P(1)}{\partial q_1} =& \left[\frac{p_1}{1-\lambda} - \frac{p_1\gamma}{(1-\lambda)^2}\right](\Delta G - H) - \frac{p_1 H}{(1-\lambda)^2} \\
=& p_1 \frac{(1-\lambda-\gamma)(\Delta G - H) - H}{(1-\lambda)^2} \\
\leq& p_1 \frac{(1-\lambda-\gamma)(\frac{H}{\kappa_1} - H) - H}{(1-\lambda)^2} \\
=& p_1 \frac{(1-\lambda-\gamma)(\frac{1}{\kappa_1} - 1) - 1}{(1-\lambda)^2} H \\
=& p_1 \frac{(1-\lambda-\gamma)\left(\frac{1-\lambda-(1-p_1)(1-\gamma)}{(1-p_1)\gamma} - 1\right) - 1}{(1-\lambda)^2} H \\
=& p_1 \frac{(1-\lambda-\gamma)\frac{1-\lambda-(1-p_1)}{(1-p_1)\gamma} - 1}{(1-\lambda)^2} H \\
=& p_1 \frac{(1-\lambda-\gamma)\frac{1-p_1(1-q_1)-(1-p_1)}{(1-p_1)\gamma} - 1}{(1-\lambda)^2} H \\
=& p_1 \frac{(1-\lambda-\gamma)\frac{p_1q_1}{(1-p_1)\gamma} - 1}{(1-\lambda)^2} H \\
=& p_1 \frac{\frac{p_1q_1q_2}{\gamma} - 1}{(1-\lambda)^2} H \\
=& p_1 \frac{p_1q_1q_2 - \gamma}{\gamma(1-\lambda)^2} H \\
=& p_1 \frac{-(1-\lambda)(1-q_2)}{\gamma(1-\lambda)^2} H
\end{aligned}$$

$$\leq 0,$$

where the first inequality follows from condition (EC.9), while the last one is based on the observations that $\lambda \leq 1$ and $q_2 \leq 1$. Note that $1-\lambda-\gamma = (1-p_1)q_2$ is independent of q_1. For the second-order derivative, we find

$$\frac{\partial^2 J^P(1)}{\partial q_1^2} = -\frac{p_1^2\left[(1-\lambda-\gamma)(\Delta G - H) - H\right]}{(1-\lambda)^3} \geq 0,$$

where the inequality is drawn from the conclusion of the first-order derivative, specifically, $(1-\lambda-\gamma)(\Delta G - H) - H \leq 0$. Thus, $J^P(1)$ is convex decreasing in q_1, implying that the benefit of predictive information is concave increasing in q_1.

To compare the rates of benefit increase from sensitivity q_1 and specificity q_2, we examine the first-order derivatives of $J^P(1)$ with respect to q_1 and q_2. Recall that $J^P(1)$ denotes the expected cost. Consequently, the rate of benefit increase with respect to q_1 is greater than that with respect to q_2 if

$$\begin{aligned}
&-\frac{\partial J^P(1)}{\partial q_1} \geq -\frac{\partial J^P(1)}{\partial q_2} \\
\iff &-p_1\frac{(1-\lambda-\gamma)(\Delta G - H) - H}{(1-\lambda)^2} \geq \frac{1-p_1}{1-\lambda}(\Delta G - H) \\
\iff &-p_1\frac{(1-\lambda-\gamma)(\Delta G - H) - H}{1-\lambda} \geq (1-p_1)(\Delta G - H) \\
\iff &p_1\frac{H + (1-\lambda-\gamma)H}{1-\lambda} + (1-p_1)H \geq (1-p_1)\Delta G + \frac{p_1(1-\lambda-\gamma)}{1-\lambda}\Delta G \\
\iff &\left[p_1\frac{2-\lambda-\gamma}{1-\lambda} + (1-p_1)\right]H \geq \left[(1-p_1) + \frac{p_1(1-\lambda-\gamma)}{1-\lambda}\right]\Delta G \\
\iff &\frac{H}{\Delta G} \geq \frac{(1-p_1)(1-\lambda) + p_1(1-\lambda-\gamma)}{(1-p_1)(1-\lambda) + p_1(2-\lambda-\gamma)} \\
\iff &\frac{H}{\Delta G} \geq \frac{1-\lambda-p_1\gamma}{p_1 + 1 - \lambda - p_1\gamma}.
\end{aligned}$$

Hence, the rate of benefit increase from q_1 is greater than that from q_2 if

$$\left(1 - \frac{H}{\Delta G}\right)\left[1 - p_1(1-p_1)(1-q_1-q_2)\right] - p_1 \leq 0.$$

Otherwise, the rate of benefit increase from q_2 is greater than that from q_1.

To establish an upper bound for the benefit of predictive information, we set $q_1 = q_2 = 1$ in $J^P(1)$, where the cost is minimized. This results in

$$J^P(1) \geq p_1 G(1) + (1-p_1)G(2) + (1-p_1)H.$$

Note that

$$J^B(1) = \min\left\{G(1),\ G(2) + \frac{H}{1-p_1}\right\}.$$

Thus, the upper bound of the benefit of predictive information is given by

$$\begin{aligned} V := J^B(1) - J^P(1) &\le \min\left\{(1-p_1)(\Delta G - H),\ \frac{p_1 H}{1-p_1} - p_1(\Delta G - H)\right\} \\ &= \min\{\beta_1,\ \beta_2\}. \end{aligned}$$

To derive a lower bound for the benefit of predictive information, we can express V as

$$V = \min\left\{\Delta G,\ \frac{H}{1-p_1}\right\} - \frac{\gamma \Delta G + (1-\gamma)H}{1-\lambda}.$$

For $q_1 = 0$, $q_2 = 1$, we have $\lambda = p_1$, $\gamma = 0$, and $V = H/(1-p_1)$. For $q_1 = 1$, $q_2 = 0$, we have $\lambda = 0$, $\gamma = 1$, and $V = \Delta G$. For $q_1 = q_2 = 1$, we have $\lambda = 0$, $\gamma = p_1$, and $V = p_1 \Delta G + (1-p_1)H$. The affine function connecting these three points can be defined as follows:

$$\begin{aligned} \mathcal{L}(q_1, q_2) = &\min\left\{\Delta G,\ \frac{H}{1-p_1}\right\} + \left[\frac{p_1 H}{1-p_1} - p_1(\Delta G - H)\right] q_1 + \\ &(1-p_1)(\Delta G - H) q_2 - (1-p_1)(\Delta G - H) - \frac{H}{1-p_1} \\ = &q_1 \beta_1 + q_2 \beta_2 - \max\{\beta_1, \beta_2\}. \end{aligned}$$

As previously demonstrated, the benefit function is concave in q_1 and linear in q_2. Therefore, it is jointly concave in (q_1, q_2). Since the affine function $\mathcal{L}(q_1, q_2)$ traverses the three extreme points of the benefit function (within the feasible space defined by $0 \le q_1, q_2 \le 1$ and $q_1 + q_2 \ge 1$), it serves as a lower bound for the benefit.

The proof is now complete.

EC.2.8. Proof of Theorem 4

Under our assumption, the DM employs the optimal policy with perfect prediction by setting $q_1 = q_2 = 1$. From Theorem 3, given $q_1 = q_2 = 1$, we deduce that the optimal policy with perfect prediction is strictly better than the optimal policy without prediction if and only if

$$1 > \frac{H}{\Delta G} > \frac{1-p_1}{2-p_1},$$

where the optimal policy with perfect prediction uses the decision rule $a(1,1) = 1, a(1,2) = 0$.

Let d represent the decision rule $a(1,1) = 1, a(1,2) = 0$. Thus, the expected cost for this policy under imperfect prediction is

$$J^d(1) = \frac{\gamma}{1-\lambda} G(1) + \frac{1-\lambda-\gamma}{1-\lambda} G(2) + \frac{1-\gamma}{1-\lambda} H.$$

As established in Appendix EC.2.7, this cost function is strictly lower than the costs associated with the other two policies if and only if condition (EC.9) is satisfied. Note that $\kappa_2 \le 1$. Furthermore, we obtain

$$\begin{aligned} \kappa_1 - \frac{1-p_1}{2-p_1} &= \frac{(1-p_1)\gamma}{p_1 q_1 + (1-p_1)\gamma} - \frac{1-p_1}{2-p_1} \\ &= \frac{(1-p_1)^2(1-q_2)}{[p_1 q_1 + (1-p_1)\gamma](2-p_1)} \\ &\ge 0. \end{aligned}$$

Thus, when $\kappa_1 < H/\Delta G < \kappa_2$, the optimal policy with perfect prediction is $a(1,1) = 1, a(1,2) = 0$, and its associated cost is lower than that of the optimal policy without prediction.

If $\kappa_2 \leq H/\Delta G < 1$ or $(1-p_1)/(2-p_1) < H/\Delta G \leq \kappa_1$, the DM still applies the optimal policy with perfect prediction, which is $a(1,1) = 1, a(1,2) = 0$. However, according to the insights from Appendix EC.2.7, the cost is higher than that of the optimal policy without prediction. Consequently, in this case, using the optimal policy with perfect prediction in an imperfect environment proves counterproductive.

When $H/\Delta G \leq (1-p_1)/(2-p_1)$, both the optimal policy with perfect prediction and the optimal policy without prediction propose not to extubate class 1 patients, irrespective of the prediction. As such, the costs remain consistent. The proof concludes here.

EC.3. Discounting Future Costs

In this section, we consider the scenario where the DM is more concerned about the immediate cost. Therefore, a discount factor $\alpha \in [0,1]$ is incorporated into our models. The value functions of the two settings are then modified as follows:

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

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

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

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

To prove Theorem EC.1, we redefine the function (EC.4) as

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

For epoch $T-1$, we have

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

The last term above is decreasing in c_{T-1} due to Lemma EC.1 and Assumption EC.1. Combined with Assumption EC.2, this implies that $h_{T-1}^B(c_{T-1})$ is decreasing in c_{T-1}. Assume that at epoch $t+1$, $h_{t+1}^B(c_{t+1})$ is decreasing in c_{t+1}. For epoch t, we have

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

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

EC.2, and $\sum_{c_{t+1}\in\mathcal{C}} P(c_t, c_{t+1})G(c_{t+1})$ is decreasing in c_t due to Assumption EC.1. Hence, $h_t^B(c_t)$ is decreasing 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, the proof for Theorem EC.1 with a discount factor is complete.

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 is given by

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

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

For $t = T-1$, we have

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

The above equalities arise from definitions and the boundary condition. Clearly, $h_{T-1}^B(c_{T-1})\leq 0$ if $c_{T-1}\geq c^*$, and $h_{T-1}^B(c_{T-1}) > 0$ otherwise, based on the definition of c^*. Assume that the property holds at epoch $t+1$. For epoch t, we have

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

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

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

We redefine that

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

First, note that $h_t^P(c_t,\hat{c}_{t+1})$ decreases with respect to c_t, which can be proved from a similar argument presented in Appendix EC.3.1. Furthermore, according to Appendix EC.2.4, the expression $\sum_{c_{t+1}\in\mathcal{C}} \pi(c_{t+1} \mid c_t,\hat{c}_{t+1}) \sum_{\hat{c}_{t+2}\in\mathcal{C}} \tilde{Q}(c_{t+1},\hat{c}_{t+2}) J_{t+1}^P(c_{t+1},\hat{c}_{t+2})$ decreases as $\hat{c}_{t+1}$ increases. Consequently, $h_t^P(c_t,\hat{c}_{t+1})$ is increasing in $\hat{c}_{t+1}$.

Next, we demonstrate that $h_t^P(c,\hat{c})$ decreases with respect to 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. With this, our proof is complete.

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

In this appendix, we develop an alternative formulation for the optimal stopping problem with predictive information using the POMDP framework. This is referred to as the POMDP-P model.

Core states. The core state at epoch t is the health state of the patient at the current epoch t and the next epoch $t+1$, denoted as (c_t, c_{t+1}). In POMDP-P, the core state cannot be completely observed by the DM.

Observations. Given that the predictive information may contain errors, the DM can only partially observe the core state (c_t, c_{t+1}) at epoch t based on the signal provided by a predictive model. Note that the current patient class c_t is fully observable. Meanwhile, the DM receives a predicted class $\hat{c}_{t+1}$ for c_{t+1} from the predictive model. Thus, the observation at epoch t is represented as $(c_t, \hat{c}_{t+1})$, and the observation space is $\mathcal{C} \times \mathcal{C}$.

Information matrix. In POMDP models, the information matrix characterizes the quality of observations. This matrix is exactly the misclassification matrix, Q, from our MDP-P model. Recall that each component of the misclassification matrix $Q(i,j) := \mathbb{P}(\hat{c} = j \mid c = i)$ represents the probability that a patient transitioning to class i at the next epoch is predicted to transition to class j when the prediction is made at the current epoch.

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

The optimality equation of the POMDP-P model is expressed as

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

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 associated with a class c_t patient at epoch t when their health condition at the subsequent epoch follows the belief $\boldsymbol{\pi}_t$. The boundary condition is

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

The belief is updated following 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. Their proofs are analogous to those for the MDP-P model and are therefore omitted here.

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

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

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

(1) For all $c_t, c_t' \in \mathcal{C}$ 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 all $c \in \mathcal{C}$ and $t = 1, 2, ..., T-2$, we have $\mathcal{B}_t^1(c) \subseteq \mathcal{B}_{t+1}^1(c)$.

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 switching curve at epoch $t \in \mathcal{T}\backslash\{T\}$ that partitions the observation space into two subsets $\mathcal{O}_t^0$ and $\mathcal{O}_t^1$, such that the optimal action given observation $(c_t, \hat{c}_{t+1})$ is*

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

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

(1) For all $t \in \mathcal{C}\backslash\{T\}$, $a_t^*(c_t, \hat{c}_{t+1})$ is increasing in c_t and decreasing in $\hat{c}_{t+1}$.

(2) For all $t = 1, 2, ..., T-2$, we have $\mathcal{O}_t^1 \subseteq \mathcal{O}_{t+1}^1$.

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 the belief space is projected into the observation 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 epoch is in a good state, it is more likely that the treatment will be stopped; if the patient's condition at the next epoch 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 left	% of previous	% of initial
All patients receiving 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 records during MV	3,282	99.6%	81.5%
Excluding the cases that 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%

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

EC.6. Definitions of Demographic and Clinical Variables

Table EC.2 Definition of demographic and clinical variables in the case study.

Variable	Definition
Age	Patient's age at ICU admission
Gender	Patient's gender
Race	Patient's race
Weight	Patient's weight at ICU admission
HR	Frequency of heartbeats
RR	Frequency of breaths
GCS	Neurological score capturing the conscious state
Creatinine	Byproduct of muscle metabolism, excreted unchanged by the kidneys
Haematocrit	Volume percentage of red blood cells in blood
Haemoglobin	Amount of iron-containing oxygen-transport metalloprotein in red blood cells
Potassium	Amount of potassium in blood
Sodium	Amount of sodium in blood
WBC	Amount of white blood cells in blood
Platelets	Amount of platelets in blood
Potassium	Amount of potassium in blood
Sodium	Amount of sodium in blood
Platelets	Amount of platelets in blood
Systolic BP	Maximum pressure in blood vessels during heartbeats
Diastolic BP	Minimum pressure in blood vessels between heartbeats
Temperature	Body temperature
Arterial pO2	Partial pressure of oxygen in arterial blood
Arterial pCO2	Partial pressure of carbon dioxide 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	Paced atrial and ventricular rhythms
Sinus rhythm	Normal sinus rhythm
Sinus tachycardia	Normal sinus tachycardia rhythm
Nursing assessment	
Cardiac	The pulse is regular; the heart rate is between 60 and 100 beats/min; the skin is warm and dry; the blood pressure is less than 140/90 mmHg; there are no symptoms of hypotension.
Respiratory	The respiratory rate is between 12 and 24 times/min; bilateral breath sounds are clear; nail beds and mucous membranes are pink; the sputum is clear.
Vascular	The extremities are normal, pink, and warm; peripheral pulses are palpable; capillary refill is 3 seconds or less; there is no edema, numbness, or tingling.
Musculoskeletal	The patient can move all extremities independently and perform functional activities.
Neurological	The patient is alert, oriented to person, place, time, and situation; the speech is coherent.
Nutrition	There is no difficulty in chewing, swallowing, or manual dexterity; the patient consumes over 50% of the daily diet ordered.

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

EC.7. Sensitivity Analysis

In this appendix, we report a series of sensitivity analyses focusing on parameter estimation, which include terminal costs, the transition matrix, and the misclassification matrix. We also explore alternative patient classification schemes and different definitions of terminal costs.

EC.7.1. Sensitivity Analysis on Terminal Costs

Note that unobserved selection bias might exist when estimating terminal costs. To make our results more robust, we conduct a sensitivity analysis to investigate the potential impact of estimation errors on terminal costs and extubation policies. Due to the potential selection bias arising from unobserved factors, the patients who were extubated in the data might be healthier in those unobserved dimensions. Consequently, our estimator $\hat{G}$ could be underestimated compared to the true parameter G. We assume a bias coefficient v, such that the true terminal cost is given by $G = (1+v)\hat{G}$.

To investigate the effect of this bias on the optimal policy, we vary v from 0 to 0.5 in steps of 0.01 and derive the optimal policy using terminal cost G. As a policy with predictive information consists of a set of action matrices at each epoch $t \in \mathcal{T}$ with binary entries, we use the entry-wise absolute difference to measure the discrepancy between two policies. We define the action deviation as

$$d(A^G, A^{\hat{G}}) = \frac{1}{T-1} \sum_{t=1}^{T-1} \sum_{i=1}^{C} \sum_{j=1}^{C} \left| A^G_{t,i,j} - A^{\hat{G}}_{t,i,j} \right|,$$

where A^G and $A^{\hat{G}}$ represent the optimal policies using G and $\hat{G}$, respectively; the first index is for time t, the second index for the current class i, and the third index for predicted class j.

Four different policies were observed among the 500 experiments, excluding the one presented in the main paper, i.e., the optimal policy under $\hat{G}$. Note that the optimal policy only changes when there are sufficient changes in the terminal costs (i.e., a significant change in v). The action deviation is 0.1 on average for the four different policies. In our case, describing a policy at an epoch requires $5 \times 5 = 25$ entries. Comparing 0.1 with 25, we conclude that the optimal policy is not sensitive to estimation bias in terminal costs.

EC.7.2. Sensitivity Analysis on Transition and Misclassification Matrices

The estimated transition and misclassification matrices could be biased due to the extubation decisions made. In this section, we perform a set of sensitivity analyses on these matrices and investigate their impact on our results.

Specifically, to construct the uncertainty set, we use the bootstrap method to estimate the 95% confidence interval for each element of the transition matrix, P. We randomly select observations with replacement from the full sample to generate a new sample of the same size. From this

bootstrapped sample, we estimate a transition matrix. We then construct an element-wise interval uncertainty set using the 50,000 estimated transition matrices. That is, for the i^{th} row of P, the uncertainty set $\mathcal{P}_i$ is defined as

$$\mathcal{P}_i = \left(\left[\underline{P}_{i,1}, \bar{P}_{i,1} \right], \cdots, \left[\underline{P}_{i,C}, \bar{P}_{i,C} \right] \right) \cap \Delta_C,$$

where $[\underline{P}_{i,j}, \bar{P}_{i,j}]$ is the confidence interval of element P_{ij}, and Δ_C represents the $(C-1)$-dimensional probability simplex. We subsequently generate 50,000 $\hat{P}$ samples based on $\mathcal{P}_i$ and derive the optimal policy with imperfect prediction, as well as the optimal cost, for each $\hat{P}$. Among the 50,000 experiments, we observe 209 distinct policies, and the action deviation is 0.5, on average, for these policies. Table EC.3 provides a summary of the performance, while Table EC.4 presents the benefits of predictive information compared to the model without predictive information. We also carry out a similar analysis for the misclassification matrix, in which we observe 205 different policies among the 50,000 experiments, and the action deviation is 0.5 on average. Hence, we conclude that the optimal policy remains relatively robust to estimation errors in both the transition and misclassification matrices.

Table EC.3 Performance of the optimal extubation policy with imperfect prediction under uncertain transition and misclassification matrices.

Initial class	Uncertain transition matrix		Uncertain misclassification matrix	
	LOS (hr)	EFR (%)	LOS (hr)	EFR (%)
1	105.4 (102.0-109.7)	19.6 (17.6-22.9)	105.3 (105.2-105.4)	19.6 (19.4-19.7)
2	102.9 (101.9-103.7)	23.6 (21.8-27.0)	102.9 (102.6-103.2)	23.2 (22.8-23.6)
3	97.1 (97.0-97.1)	20.6 (20.2-21.0)	97.1 (97.1-97.1)	21.0 (21.0-21.0)
4	94.3 (94.2-94.3)	16.9 (16.8-17.0)	94.3 (94.1-94.6)	16.9 (16.8-17.0)

LOS: length of stay; EFR: extubation failure rate.
* Numbers in parentheses indicate 95% confidence intervals.

EC.7.3. Adjusted Transition Matrix Estimation

As mentioned earlier, the extubation decision can also affect the estimation of the transition matrix. When patients are extubated, we are unable to observe their health transitions as we would if they 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 for terminal cost estimation to account for this

Table EC.4 Benefit of the predictive information under uncertain transition and misclassification matrices.

Initial class	Uncertain transition matrix		Uncertain misclassification matrix	
	LOS (hr)	EFR (%)	LOS (hr)	EFR (%)
1	0.5 (0.2-1.0)	2.0 (-1.8-4.0)	0.4 (0.3-0.5)	1.3 (1.1-1.4)
2	1.4 (0.6-2.2)	5.7 (-4.4-7.4)	1.2 (1.0-1.4)	3.5 (3.1-3.9)
3	0.0 (0.0-0.1)	0.1 (0.0-0.8)	0.0 (0.0-0.0)	0.0 (0.0-0.0)
4	0.3 (0.0-0.4)	0.8 (0.7-0.9)	0.3 (0.0-0.5)	0.8 (0.6-0.9)

LOS: length of stay; EFR: extubation failure rate.
* Numbers in parentheses indicate 95% confidence intervals.

selection bias as much as possible. In particular, we estimate the transition probabilities using the following equation:

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

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

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

The performance of our models under the adjusted transition matrix is shown in Table EC.5. The results closely resemble those obtained using the transition matrix estimated via the sample mean approach presented in the main text.

Table EC.5 Performance of different extubation policies under the propensity-adjusted transition matrix.

Initial Class	OP-B		OP-PP		OP-IP		OP-PP under $\hat{Q}$	
	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 prediction; 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.7.4. Alternative Patient Classification Schemes

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

EC.7.4.1. Six-class Stratification

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

Table EC.6 Estimated RLOS and EFR under six-class patient stratification.

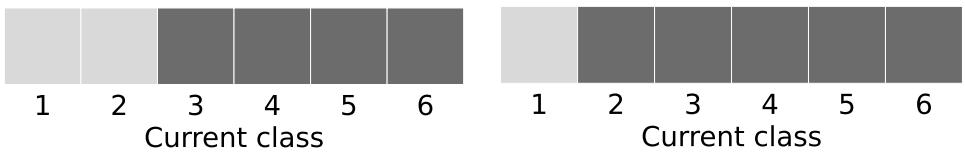
Terminal class	No. of observations	RLOS (hr)	EFR (%)
1	29	137.8	40.9
2	55	99.6	26.0
3	95	97.3	30.2
4	419	91.1	21.0
5	688	88.6	17.7
6	1,781	79.1	13.8

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

The transition probability matrix is estimated as follows:

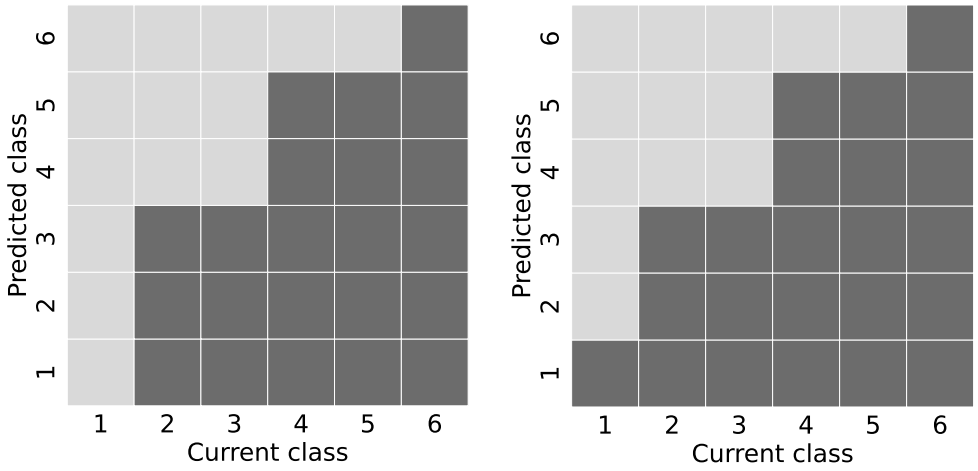
$$P=\begin{pmatrix} 0.35 & 0.13 & 0.08 & 0.20 & 0.11 & 0.13 \\ 0.13 & 0.20 & 0.20 & 0.22 & 0.08 & 0.17 \\ 0.07 & 0.11 & 0.29 & 0.31 & 0.15 & 0.07 \\ 0.02 & 0.05 & 0.09 & 0.48 & 0.24 & 0.12 \\ 0.01 & 0.02 & 0.03 & 0.23 & 0.36 & 0.35 \\ 0.01 & 0.01 & 0.01 & 0.07 & 0.18 & 0.72 \end{pmatrix}.$$

With these parameters, we derive results under our models. The optimal policies are depicted in Figures EC.1 and EC.2. We find that classifying patients into more classes enables more precise decisions. For example, under the base policy without prediction, patients with an RSBI in $[80,105)$ will not be extubated in the five-class case but will be extubated in the six-class case during the last two epochs.



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

Figure EC.1 Optimal extubation policy without prediction under six-class patient stratification. The dark gray grid indicates extubation, while the light gray grid represents continued ventilation.



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

Figure EC.2 Optimal extubation policy with perfect prediction under six-class patient stratification. The dark gray grid indicates extubation, while the light gray grid represents continued ventilation.

The performance under different extubation policies is summarized in Table EC.7. With one additional class, we can further reduce LOS and EFR using predictive information. However, the improvement is marginal. Moreover, the issues with the base model persist even with finer patient classification: both LOS and EFR can worsen compared with current practice if the model does not incorporate predictive information.

Table EC.7 Performance of different extubation policies under six-class patient stratification.

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 prediction; OP-PP: optimal extubation policy with perfect prediction; LOS: length of stay; EFR: extubation failure rate.

EC.7.4.2. Seven-class Stratification

In this section, we further refine the classification and group 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 estimated terminal costs are shown in Table EC.8.

Table EC.8 Estimated RLOS and EFR under seven-class patient stratification.

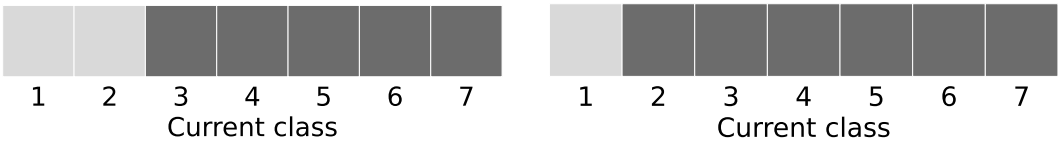
Terminal class	No. of observations	RLOS (hr)	EFR (%)
1	29	137.8	40.9
2	55	99.6	26.0
3	95	97.3	30.2
4	121	94.7	21.4
5	298	89.4	20.7
6	688	88.6	17.7
7	1,781	79.1	13.8

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

The transition probability matrix is estimated as follows:

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

The optimal policies are depicted in Figures EC.3 and EC.4, and their performance is summarized in Table EC.9. All observations and insights are consistent with the results reported in the main text with five patient classes.

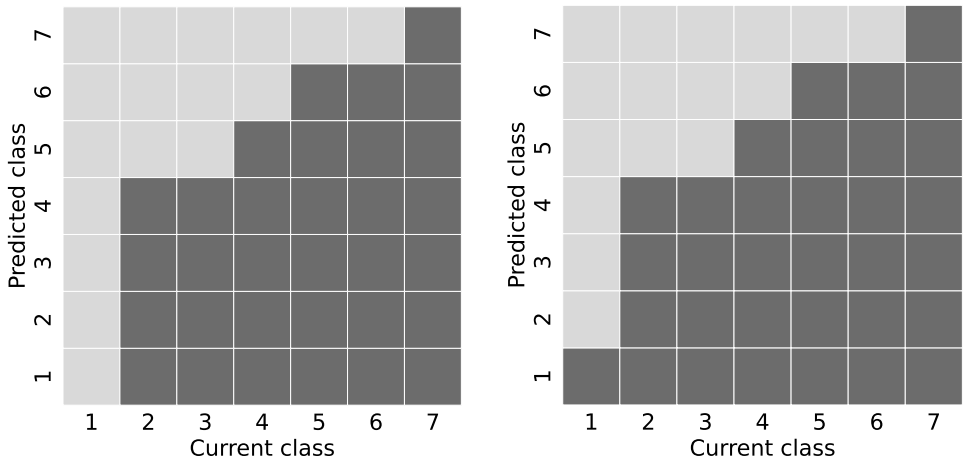


(a) $t = 1, 2, \ldots, 24$. (b) $t = 25, 26, 27$.

Figure EC.3 Optimal extubation policy without prediction under seven-class patient stratification. The dark gray grid indicates extubation, while the light gray grid represents continued ventilation.

EC.7.5. Different Weights on Terminal Cost

Physicians consider two important factors when making extubation decisions: LOS and EFR. In the main analysis, we construct the objective function using LOS as the sole criterion. We show that optimal policies failing to consider predictive information can result in more extubation failures compared with current practice. However, if predictive information is properly incorporated into the model, the optimal policies can substantially reduce extubation failures while minimizing only LOS. In this section, we conduct a sensitivity analysis by modifying the value function to directly account for EFR in the models. To this end, we introduce a tuning parameter λ to model the trade-off between ventilation duration and terminal cost.



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

Figure EC.4 Optimal extubation policy with perfect prediction under seven-class patient stratification. The dark gray grid indicates extubation, while the light gray grid represents continued ventilation.

Table EC.9 Performance of different extubation policies under seven-class patient stratification.

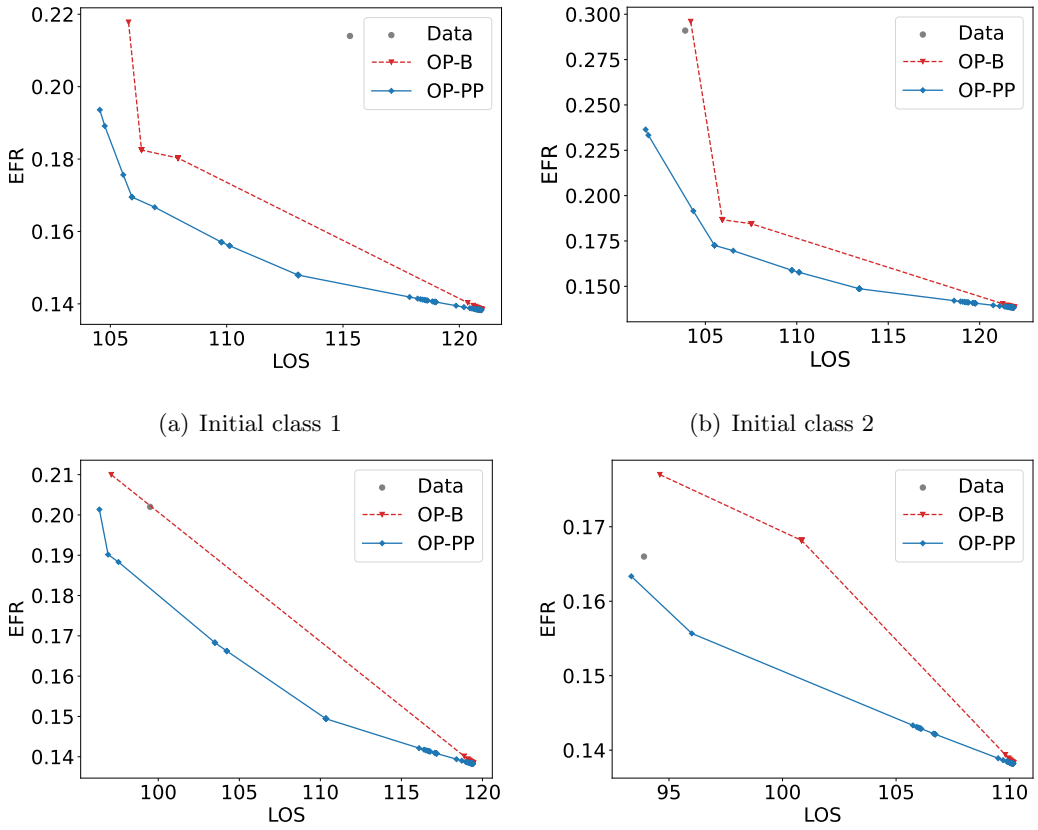
Initial Class	Data		OP-B		OP-PP	
	LOS (hr)	EFR (%)	LOS (hr)	EFR (%)	LOS (hr)	EFR (%)
1	115.3	21.4	105.9	20.2	104.7	19.2
2	104.4	27.8	105.2	21.3	102.2	21.8
3	103.7	29.9	103.3	30.2	101.5	25.6
4	104.6	26.3	100.7	21.4	99.6	20.1
5	98.3	18.7	95.4	20.7	94.9	19.9
6	93.9	16.6	94.6	17.7	93.3	16.3
7	86.7	13.9	85.0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

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

Specifically, the value function of the MDP-P model becomes:

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

Since the terminal cost for each class increases with the risk of extubation failure, λ can be viewed as the weight assigned to extubation failure relative to LOS; $\lambda = 1$ corresponds to the case reported in the main text. With a larger λ, the DM tends to keep patients ventilated for longer durations and extubates them when they are in better conditions. We vary λ from 1 to 10 at intervals of 0.01 to construct the efficient frontiers that capture the trade-off between LOS and EFR under different models. The results are presented in Figure EC.5. Since Class 5 represents the best condition, the weight has no impact on the optimal action for this class of patients; therefore, we only present results for the other four classes.



(a) Initial class 1

(b) Initial class 2

(c) Initial class 3

(d) Initial class 4

Figure EC.5 Trade-off between LOS and EFR under different weights on terminal costs.

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

In Figure EC.5, the x-axis represents the expected total LOS since intubation, and the y-axis corresponds to the EFR. Each point corresponds to a specific λ value. In these charts, points at the lower-left corners represent more effective policies with lower LOS and EFR. Since a larger λ results in extended LOS, the point moves from left to right as λ increases, thereby increasing LOS while decreasing EFR. The leftmost points on these curves correspond to cases with $\lambda = 1$, which are the results reported in the main text. These charts enable the DM to visualize the benefits of using predictive information and to identify inefficiencies in the base model without predictive information compared to current practice. The DM can use these charts to balance the trade-off between the two objectives, i.e., LOS and EFR, opting for a larger λ if EFR is considered to carry greater weight in extubation decisions.

EC.8. Impact of Utilization with Different Weighting Functions

In practice, system utilization can influence individual-level decisions in several ways (KC and Terwiesch 2009, Kim et al. 2016, Shen et al. 2021). Specific to our context, in a less-congested unit, the DM might prioritize the quality of care and aim for a lower EFR. However, as the unit becomes more congested, the DM may feel compelled to conduct early extubations to transition more patients to downstream care units. Incorporating the utilization level directly into the state would result in a high-dimensional MDP. Following the approach of Grand-Clément et al. (2021), we implicitly account for the effects of utilization by discounting the continuation cost. Specifically, we use the utilization level to discount the constant treatment cost in our model, leading to the following revised value function:

$$J_t^P(c_t, \hat{c}_{t+1}) = \min \left\{ G(c_t),\ w_\rho H + \sum_{c_{t+1} \in \mathcal{C}} \sum_{\hat{c}_{t+2} \in \mathcal{C}} \mathbb{P}\left(c_{t+1}, \hat{c}_{t+2} \mid c_t, \hat{c}_{t+1}\right) J_{t+1}^P(c_{t+1}, \hat{c}_{t+2}) \right\},$$

where w_ρ denotes the impact of utilization level on the immediate cost H. Intuitively, the weight w_ρ should increase with ρ. As the utilization level grows, the one-epoch LOS assumes heightened importance.

In our analysis, we adopt various forms of the weighting function w_ρ to represent the influence of utilization on the extubation decision. Given that Erlang loss systems are commonly applied to model healthcare service systems, including ICUs, we formulate the weighting function based on the Erlang loss system's blocking probability (van Dijk and Kortbeek 2009). In an Erlang loss system with capacity s, arrival rate λ, and service rate μ, the blocking probability is defined as $B(s,\rho) = (a^s/s!)/(\sum_{k=0}^{s} a^k/k!)$, where $a = \lambda/\mu$. In our model, $s = 15$. We define $\rho = a/s$ as the system's utilization level. Subsequently, as we vary $\rho \in (0,1)$, we calculate the associated blocking probabilities for different utilization levels. To construct the weight function, we normalize the blocking probability to get:

$$w_\rho = \frac{B(s,a)}{\max_{\rho \in (0,1)} B(s,a)}. \tag{EC.10}$$

This weighting function yields two desirable properties: (1) $w_\rho \to 1$ as $\rho \to 1$, implying that the objective becomes solely the LOS when the system is at its busiest, and (2) $w_\rho \to 0$ as $\rho \to 0$, indicating that the objective shifts to minimize the EFR when the utilization level gets lower.

Figure EC.6 depicts the benefit of perfect prediction—which is the reduction in objective values achieved through using perfect predictions—at various utilization levels. Note that we evaluate the objective values under the newly-defined value function with the weighting parameter specified in (EC.10). We observe that incorporating predictive information provides no benefit when the system's utilization level is low. This is because, in less busy scenarios, the holding cost becomes less important, allowing the DM the luxury of waiting until the patient progresses to a healthier

state. However, as congestion intensifies (i.e., when utilization level exceeds 70%), the value of integrating predictive information becomes more apparent. In our study ICU, the utilization level exceeds 70% more than 60% of the time, suggesting that predictive information is beneficial in most scenarios.

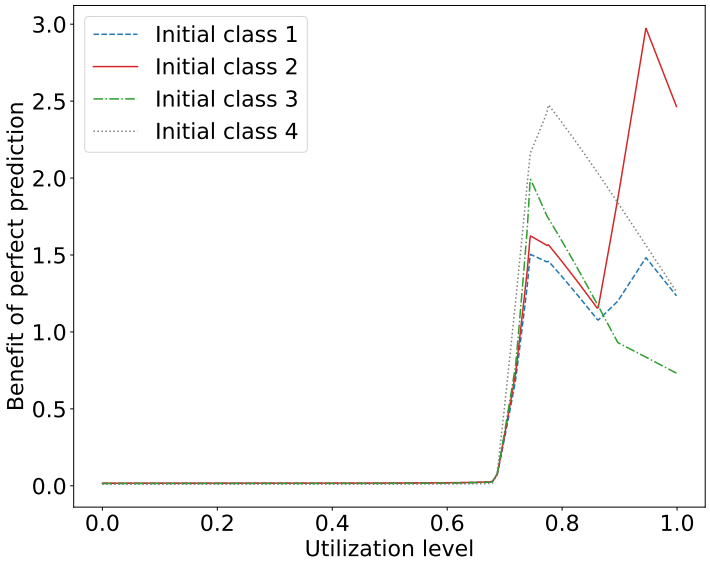


Figure EC.6 Impact of utilization levels on the benefit of perfect prediction with the holding cost weighted by Erlang blocking probability.

To more comprehensively investigate alternative weighting functions related to utilization, we also consider the Erlang delay system's waiting probability, along with a linear function and an S-shaped function, as potential utilization weights. That is,

$$w_\rho = \frac{\frac{(s\rho)^s}{s!(1-\rho)}}{\sum_{k=0}^{s-1}\frac{(s\rho)^k}{k!}+\frac{(s\rho)^s}{s!(1-\rho)}},$$

$$w_\rho = \rho, \text{ or}$$

$$w_\rho = \frac{S(\rho)-\min_{\rho\in(0,1)} S(\rho)}{\max_{\rho\in(0,1)} S(\rho)-\min_{\rho\in(0,1)} S(\rho)}, \text{ where } S(\rho)=\frac{\exp(4\rho-2)}{\exp(4\rho-2)+\exp(-4\rho+2)}.$$

Note that the values of the third function are normalized to $[0,1]$. Figures EC.7–EC.9 depict the results. We can observe that the insights are consistent with the findings based on the blocking probability of the Erlang loss system.

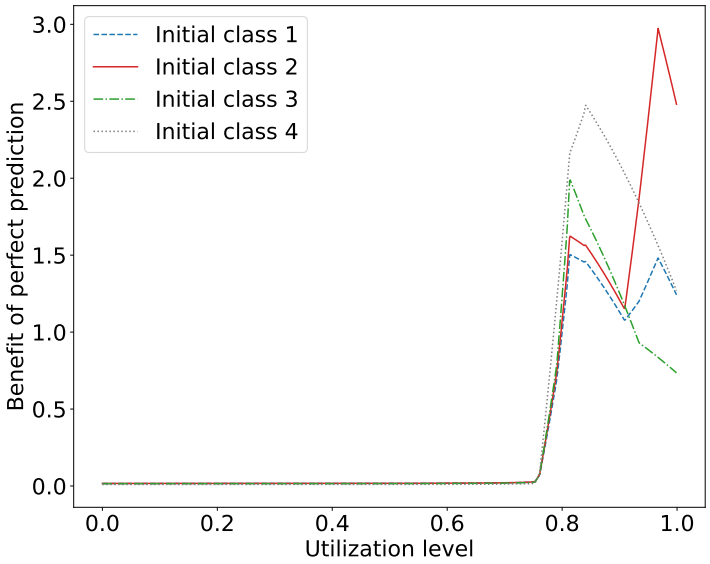


Figure EC.7 **Impact of utilization levels on the benefit of perfect prediction with the holding cost weighted by Erlang waiting probability.**

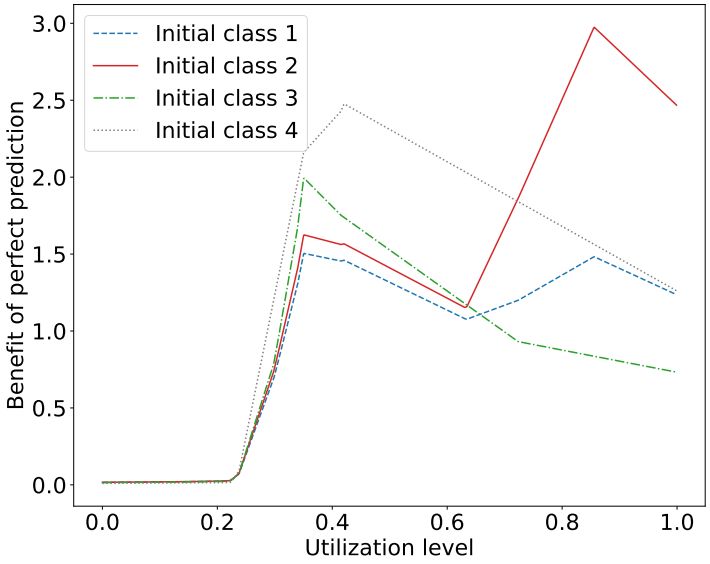


Figure EC.8 **Impact of utilization levels on the benefit of perfect prediction with the holding cost weighted by a linear function of the utilization level.**

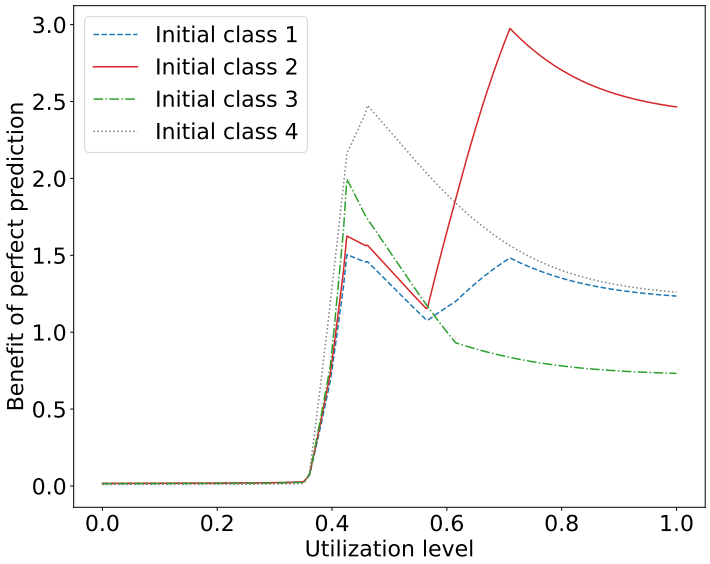


Figure EC.9 Impact of utilization levels on the benefit of perfect prediction with the holding cost weighted by an S-shaped function of the utilization level.

EC.9. Regression Models in Terminal Costs Estimation

In all regression models, we incorporate not only the original clinical variables but also transform numerical variables into categorical ones based on their normal ranges, aiming to produce more accurate regression models. For instance, for heart rate (HR), we encode $\text{HR} < 60$, $\text{HR} \in [60, 100]$, and $\text{HR} > 100$ into three categories. The regression coefficients from models that estimate RLOS, EFR, and propensity to extubate are shown in Table EC.10. As discussed in Section 4.3, we use linear regression to estimate RLOS and extubation failures. Logistic regression is used to estimate the propensity to extubate.

Table EC.10: Regression coefficients from models for parameters estimation.

Variables	RLOS	Extubation Failure	Extubation Decision
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
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
GCS	-0.037	-0.005*	0.170***
Creatinine			
45-110	1	1	1
<45	-5.949	-0.038	-0.006
>110	24.174***	0.115***	-0.456***
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
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***
WBC			
4.5-11	1	1	1
<4.5	-14.271	-0.056	0.603**

>11	6.565***	0.008	-0.018
Platelets			
150-450	1	1	1
<150	8.326***	0.045***	-0.077
>450	-1.313	0.073	0.568
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 pO2			
80-100	1	1	1
<80	2.555	0.028	0.258***
>100	0.331	0.024	-0.117
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***
Arterial SaO2			
93-97	1	1	1
<93	-3.026	-0.068***	0.055
>97	-6.222	-0.071***	0.438***
FiO2	0.326**	0.003***	-0.012***
Tidal volume	0.013	0.000	0.000
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*
SOFA	2.303***	-0.001	-0.124***
RSBI	0.228*	0.000	0.000
Arterial pO2/FiO2	-0.740	-0.032***	-0.028
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
No. of observations	3,067	3,067	7,281
(Pseudo) R-squared	0.15	0.16	0.27

Prob > Chi-squared	< 0.001

RLOS: remaining length of stay, HR: heart rate; RR: respiratory rate; GCS: Glasgow Coma Scale; WBC: white blood cell; BP: blood pressure.
Linear regression is used for predicting RLOS and extubation failures, while logistic regression is used for predicting extubation decisions.
$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

EC.10. Constrained Estimation of Transition and Misclassification Matrices

Threshold-type policies are widely used in practice for their better interpretability and ease of implementation. Our analytical results in Section 4.1—specifically, Theorem 1—guarantee the optimality of threshold-type policies, provided that both the transition probability matrix and the misclassification matrix satisfy the TP2 condition (i.e., Assumptions 1 and 2). As previously discussed, these assumptions are reasonable and have also been made in other healthcare studies (e.g., Boloori et al. 2020). However, naïve estimators for these matrices may not satisfy the TP2 condition, even if the underlying transition probability and misclassification matrices are inherently TP2. Consequently, optimal policies based on biased estimators may not be of the threshold type, complicating interpretation and implementation.

We propose an estimation procedure that incorporates the TP2 requirement into the estimation of the transition probability matrix and misclassification matrix by solving constrained maximum likelihood estimation (MLE) problems. Using these estimated matrices as inputs, we can guarantee the derivation of threshold-type policies through solving the MDP-P model. We then assess the performance gap between policies obtained using this approach and those obtained earlier via naïve MLE estimation.

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

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

where N is a matrix consisting of the counts of patient transitions. This problem aims to maximize the log-likelihood of observed transition data within the space of TP2-constrained transition probability matrices. A similar formulation can be made for the misclassification matrix, in which we replace N with the misclassification matrix estimated using the naïve approach. The resulting estimation problem is a non-convex, quadratically-constrained optimization problem. Although we attempt to solve this problem using the sequential quadratic programming (SQP) algorithm, not all constraints may be feasible with SQP. It should be noted that the primary reason for estimating a TP2 transition probability matrix is to obtain a threshold policy; thus, it is more important to guarantee feasibility than optimality when solving the estimation problem.

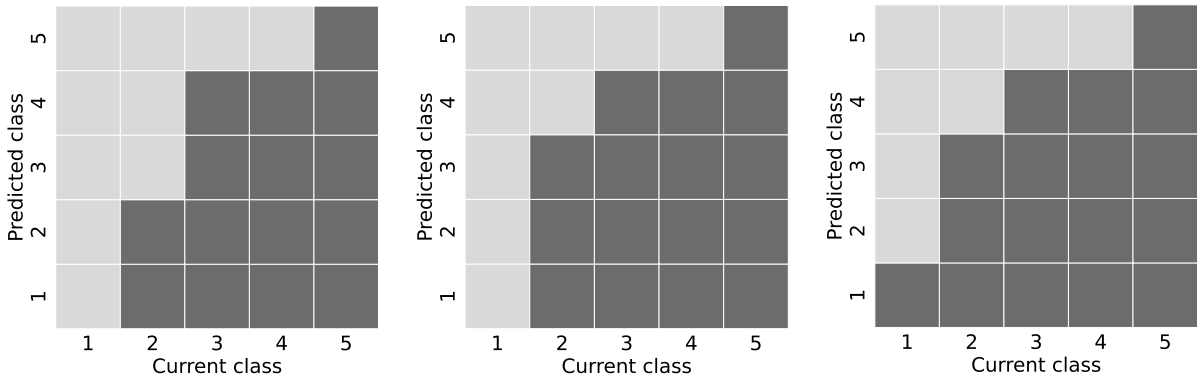
To address the aforementioned issue, we propose a heuristic that solves a series of convex optimization problems based on the equivalence between the TP2 condition and row-wise MLR dominance, as discussed in Section 4.1. Specifically, we estimate the transition probability matrix row

by row, requiring MLR dominance between consecutive rows. For the first row, the transition probabilities are estimated via conventional MLE subject to a simplex constraint. For subsequent rows, we solve a constrained MLE problem that includes an additional condition in which the second row exhibits MLR dominance over the first row. This results in a convex optimization problem that can be efficiently solved. We continue this process for each remaining row of the transition probability matrix. Although this procedure is a greedy heuristic, we find it produces higher likelihoods for both the transition and misclassification matrices compared to those estimated by the SQP algorithm. The estimated matrices are provided as follows:

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

$$\hat{Q}_{TP2} = \begin{pmatrix} 0.60 & 0.27 & 0.09 & 0.04 & 0.00 \\ 0.11 & 0.59 & 0.20 & 0.08 & 0.02 \\ 0.02 & 0.08 & 0.61 & 0.23 & 0.06 \\ 0.00 & 0.03 & 0.30 & 0.53 & 0.14 \\ 0.00 & 0.01 & 0.11 & 0.28 & 0.60 \end{pmatrix}.$$

Next, we resolve the MDP-B and MDP-P models using the above estimation. The OP-B and OP-PP policies remain the same as those depicted in Figures 3 and 4, respectively. The new OP-IP policy is presented in Figure EC.10, which is a threshold policy in contrast to Figure 5. In particular, at epoch 27, class 3 patients predicted to improve to class 5 at epoch 28 are now recommended to continue ventilation under the new policy illustrated in Figure EC.10.



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

Figure EC.10 Optimal extubation policy with imperfect prediction derived from using TP2 transition and misclassification matrices. The dark gray grid indicates extubation, while the light gray grid represents continued ventilation.

The performance of different policies under TP2 matrices is summarized in Table EC.11. In Table EC.11, the system dynamics are assumed to follow TP2 matrices. All previously observed

insights continue to hold under the new parameters. We also evaluate the performance of the optimal extubation policy with imperfect prediction, derived from using TP2 matrices, while the system dynamics follow the non-TP2 matrices estimated in Sections 4.3 and 5.2. As demonstrated in Table EC.12, even when assuming that the true system dynamics adhere to parameters estimated via naïve MLE, the performance of the threshold policy derived from using TP2 matrices closely approximates that of the optimal policy as reported in Figure 5 and Table 5. Thus, the proposed method allows us to estimate TP2 matrices and subsequently derive a threshold policy for practical implementation. While it may not be optimal, the threshold policy offers easier interpretation and implementation with minimal performance loss.

Table EC.11 Performance of different extubation policies derived from using TP2 transition and misclassification matrices in the environment governed by TP2 transition and misclassification matrices.

Initial 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 prediction; 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.12 Performance of the optimal extubation policy with imperfect prediction derived using TP2 transition and misclassification matrices in the environment governed by matrices estimated via naïve MLE

Initial 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

MLE: maximum likelihood estimation; LOS: length of stay; EFR: extubation failure rate.