# DATA-DRIVEN MANAGEMENT OF INTENSIVE CARE UNITS

by

**Mehmet Yasin Ulukuş**

B.S., Boğaziçi University, 2006

M.S., Boğaziçi University, 2009

Submitted to the Graduate Faculty of

the Swanson School of Engineering in partial fulfillment

of the requirements for the degree of

**Doctor of Philosophy**

University of Pittsburgh

2017

UNIVERSITY OF PITTSBURGH

SWANSON SCHOOL OF ENGINEERING

This dissertation was presented

by

Mehmet Yasin Ulukuş

It was defended on

July 21, 2017

and approved by

Lisa M. Maillart, PhD, Professor, Department of Industrial Engineering

Andrew J. Schaefer, PhD, Professor, Department of Computational and Applied Mathematics, Rice University

Guodong Pang, PhD, Associate Professor, Department of Industrial and Manufacturing Engineering, Penn State University

Jayant Rajgopal, PhD, Professor, Department of Industrial Engineering

Gilles Clermont, MD, MSc, Professor, Department of Critical Care Medicine

Dissertation Director: Lisa M. Maillart, PhD, Professor, Department of Industrial Engineering

**DATA-DRIVEN MANAGEMENT OF INTENSIVE CARE UNITS**

Mehmet Yasin Ulukuş, PhD

University of Pittsburgh, 2017

An Intensive Care Unit (ICU) is a specialized section of a hospital that provides comprehensive and continuous care for patients in critical conditions. Around 20% of hospital operating costs are due to ICUs, and this percentage has been increasing. Modeling patient flow through an ICU is challenging due to significant heterogeneity of patient cases and high variability of evolving patient conditions. Using a highly detailed data set of ICU patients from a single health system, we build a stochastic and dynamic model of patient physiology that can significantly improve ICU operations and predictions.

Scoring systems that assess the severity at admission or progression of severity during the stay have been used to predict the outcome (mortality or readmission). Existing scores are not sufficient to predict readmissions or mortalities after transferring to a lower level care unit. We present ICU outcome prediction models that perform better than existing models and could be used to benchmark ICU discharge policies and guide post-ICU resource needs.

We consider the transfer operations of patients to a downstream unit. In current practice, downstream beds are requested once a patient is clinically ready to be transferred. We investigate *anticipative* bed requests that can be made before a patient is ready for transfer. Patient health is described via a novel transfer readiness score created using our readmission prediction model that we incorporate into a Markov decision process model. Our numerical results indicate that an anticipative transfer request policy can significantly improve the system performance. We investigate the sensitivity of policy change upon cost parameter estimation errors by using robust models, and demonstrate that proactive strategies are more beneficial than reactive current policy in most scenarios.

We present an explicit stochastic *length of stay* model considering patient physiology modeled by the transfer readiness score as well as transfer delays. We characterize the stochastic process under certain assumptions. We show that the model demonstrates a moderate performance in fitting the underlying distribution of the length of stay, and improvements on the score will improve the predictive power of the model.

**Keywords:** Operations research, Markov decision processes, simulation, statistical data analysis, classification models, intensive care unit, hospital operations, medical decision making.

## TABLE OF CONTENTS

## LIST OF TABLES

## LIST OF FIGURES

# PREFACE

*in memory of my grandfather, Mehmet Ulukuş (1933-2017)*

This dissertation is dedicated to my parents Harun and Necla Ulukuş for all their love and unconditional support. I am also greatly thankful to my brother Emre and my sister Berru. My beautiful little nieces Esila and Elif, you are my sunshine.

I am indebted to my adviser, Dr. Andrew J. Schaefer, for his helpful advices, encouragement, and direction through my doctoral studies. To work with him has been a real pleasure to me. I am grateful to Dr. Lisa M. Maillart, for her support, feedback and graciously serving as my official dissertation director. I am also thankful to Dr. Guodong Pang for his continuous support, suggestions and inspiring discussions. I would like to express my sincere gratitude to Dr. Gilles Clermont and Dr. Jayant Rajgopal for invaluable comments and feedback.

I am also indebted to my dear friends who made this hard work enjoyable. This accomplishment would not have been possible without them.

Finally, I thank God for bringing you (B.D) in my life. You have always been there for me when I need.

# 1.0 INTRODUCTION

The healthcare industry is one of the largest and fastest growing industries in the United States. In 2015, total healthcare expenditures amounted to $3.2 trillion (17.8% of gross domestic product), which translates to $9,990 a year per person [85]. Healthcare costs are growing faster than GDP. It accounted to 4% of income in 1960 compared to 6 percent in 2013 [36], and expected to grow 5.8% on average over the period 2015-2025 with the aging population [85]. The U.S. ranks last in performance, e.g., efficiency, access, and equity, among 11 industrialized countries (Australia, Canada, France, Germany, the Netherlands, New Zealand, Norway, Sweden, Switzerland, the United Kingdom, and the United States), despite having the highest healthcare costs [36].

Increasing costs have motivated significant research to address efficiency and effectiveness issues in health care. Operations Research (OR) techniques present abundant opportunities to improve such tasks. OR techniques have been applied to a wide variety of healthcare problems. There is an extensive literature on models considering system design and planning, health care operations management and medical decision making, e.g., [14, 41, 55, 63, 100, 109, 122, 123, 126, 130, 145].

This dissertation focuses on building a stochastic and dynamic model of patient physiology to improve Intensive Care Unit (ICU) operations and predictions. To this end, we (1) build models that can make better predictions of the outcome (mortality and readmission) when a patient is transferred to a downstream unit, (2) develop a transfer readiness score and employ the score via an optimization model to make anticipative bed requests (3) and build a stochastic model of patient length of stay based on patient physiology and transfer delay dynamics.

## 1.1 BACKGROUND INFORMATION ON INTENSIVE CARE UNITS AND MODELING CHALLENGES

An ICU is a specialized hospital department that provides temporary support to critically ill patients. Patients require close monitoring and support to ensure stable health conditions. The units are staffed by specialized doctors and nurses. Generally, ICUs have higher nurse-to-patient ratios and more advanced equipment, e.g., ventilating machines, pulse oximeters, cardiac monitors are available. ICU costs have been increasing significantly [32, 111], and represent around 20% of all hospital operating costs [69]. The rate of increase in the number ICU admissions tripled the rate of increase in general hospital admissions from 2002 to 2009 [10].

ICUs are tightly connected to other hospital units such as operating rooms and emergency departments. Patients are transferred to an ICU from an emergency department after surgery or from other units if their conditions deteriorate. Thus, improved ICU workflow management should improve patient flow in the entire hospital, reduce mortalities and readmissions, and help reduce healthcare expenses. Improvements can be at the strategic level, e.g., designing and locating the unit, or at the tactical level, e.g., staff scheduling, or admission, discharge and routing control. No data-driven consensus criteria have been developed for many practical decisions in critical care, including patient discharge; see, e.g., [20, 60, 152, 149].

Although the modeling and analysis of hospital units, like surgery rooms or emergency departments, are well studied in the literature, ICU models are relatively unstudied. This disparity is due to important modeling challenges common to ICUs: (1) Unlike patient flow models for other parts of a hospital, a stochastic and dynamic model of physiology is crucial. This is very challenging, and requires novel modeling techniques, as well as enormous amounts of highly detailed data. (2) Patients in various conditions are admitted to the ICU from different departments: patient cases are highly heterogeneous. This is evidenced by the ICU length of stay (LOS) in our data set, in which the mean LOS is 171 hours and the standard deviation is 265 hours. (3) Clinical practices and policies on admissions, transfers, staff scheduling and bed availability have important impacts on patient outcomes. Models

of ICUs must incorporate heterogeneous patient populations and evolving patient health conditions, as well as interactions with other units. As a result, despite their importance, ICUs remain less amenable to traditional modeling approaches. We briefly describe the problem statements and contributions, which are later detailed in Chapters 3, 4 and 5.

## 1.2 PREDICTING INTENSIVE CARE UNIT READMISSION AND MORTALITY

ICU readmissions are typically observed in 4% to 6% of discharged patients [15, 94], and may be considerably higher in selected populations and environments [74, 80]. Patients readmitted to the ICU are generally sicker [157], and incur increased hospital mortality and length of stay [93]. Existing literature also suggests that many patients discharged from the ICU will die prior to hospital discharge [74]. Additionally, ICU survivors do not revert to a population-level risk of death [25, 26, 73]. Evidence-based ICU discharge guidelines and follow-up care could lead to fewer readmissions and generally improved outcomes if modifiable factors leading to readmission are accurately identified.

Accurate predictions are essential in providing guidelines for transfer decisions, family decisions regarding end-of-life, and allocating resources dedicated to mitigating early hospital readmission. However, the quality and performance of existing models [7, 48, 57, 93] of ICU readmission risk are generally not robust and thus ongoing evaluation in larger data sets. Moreover, current mortality prediction models have not focused on the population of patients that have survived the ICU, let alone were discharged alive from the hospital. ICU readmission rates can also be used as a benchmark of quality if appropriate allocation is made for case-mix, as different environments of care vastly influence risk of readmission [74].

In Chapter 3, we develop ICU models that can make better predictions of the outcome (mortality or readmission) when a patient is transferred to a downstream unit. To calibrate these models, we employ a database from a major medical institution, which contains both medical and bed flow information of many thousands of patients over a six-year period. We can track the change of physiological indicators through patient stays and relate this

information to the outcomes (readmission or mortality) upon their transfers. We compare the performance of the model with existing studies and show that our model predicts readmissions or mortality much better than existing models in the literature. Furthermore, we demonstrate that some commonly used scores are not sufficient in predicting readmissions or mortalities after transferring to a lower level care unit. We also show that health evolution plays a key role in patient outcomes and that there is a need for a better transfer score capable of summarizing health evolution in the ICU.

## 1.3 PHYSIOLOGY-BASED ANTICIPATIVE INTENSIVE CARE UNIT MANAGEMENT

In Chapter 4, we focus on the tactical decisions in the ICU, in which we seek to reduce transfer delays, thereby reducing congestion without increasing capacity. We employ the score described in Chapter 3 to demonstrate the operational advantages of a stochastic understanding of health progression.

Patients are typically admitted to an ICU from other locations within the hospital, such as the emergency department or post-operative recovery rooms. Upon discharge from the ICU, patients are transferred to a downstream unit, e.g., stepdown units or floors (90% of the patients in our data); see also [58]. Current practice in the ICU system we consider herein is typical: when a patient is clinically ready to be transferred to a lower level of care, a transfer request is made and the patient is usually physically moved after a considerable amount of delay (an average of 9 hours and occasionally up to 24 hours in our data). This delay may be caused by many factors, such as bed availability, personnel availability, or bed cleaning. Patients typically stay in the ICU while transfer operations are conducted, unnecessarily occupying a bed. In other words, transfer delay usually increases a patient's LOS in the relatively expensive ICU bed rather than a less costly downstream bed. This may adversely affect hospital operations, including surgery cancellations and patient diversions. Moreover, prolonged LOS may lead to nosocomial infections [34, 44, 151]. We propose an anticipative transfer request policy where transfer requests are made before the patients

are medically ready to be transferred and necessary preparations in the downstream units start after the request, so that patients can be immediately transferred to the downstream units whenever they become medically ready. Figure 1.1 depicts both the current practice and our proposed anticipative policy. To apply such an anticipative policy, we must model

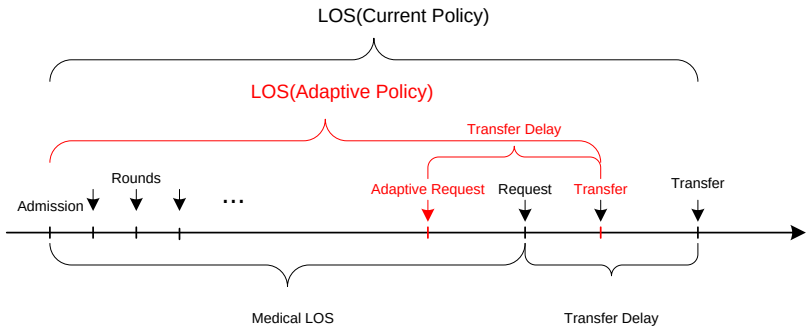



Figure 1.1: Current practice versus anticipative request policy

the patients' suitability for transfer, so that physicians can track and predict their health conditions, and thus make proactive transfer decisions. There is no unified measure of the severity of an ICU patient as it relates to transfer decisions. Patients are generally classified by intensivists as "ready to be transferred" with an emphasis on stability of their conditions. Stability can be defined as the likelihood of readmission or death upon transfer to a lower stream unit at the assessment time. Hence, we employ ICU readmission model given in Chapter 3 to generate a Transfer Readiness Score, which models patients' health condition.

In Chapter 4, we build an infinite-horizon Markov Decision Process (MDP) based on the scores to model the anticipative transfer request problem, with the objective of minimizing the expected total cost of unnecessarily occupying ICU beds plus the cost of unnecessarily reserving downstream beds. To illustrate the benefits of the anticipative transfer request policy, we conduct a numerical study based on a large data set, and compare our results with the current practice. The anticipative policy is very effective in reducing the transfer delays (up to %50 reduction) and hence reducing the congestion in the ICU. We show that the proposed policy increases the throughput of the system without decreasing the quality of care as it does not discharge patients early. The benefit of applying the anticipative

policy significantly increases as the ICU becomes more loaded. We further consider the case where bed costs are in a sense intangible and unknowable in advance, as they largely reflect the opportunity costs of beds, which may vary greatly. Thus, we build a robust dynamic programming model, which considers parameter ambiguity. Our results indicate that the anticipative policy performs better when the cost estimates are relatively accurate, and both the anticipative and the robust policies perform better than the current practice except when the cost of the downstream bed is much larger than expected.

We also explore a threshold-type policy structure for the multi-patient transfer request problem. Due to the complexity of the MDP, it remains open to prove the existence and optimality of a threshold policy. However, we conduct an experimental study where we generate problem instances with random transition probability matrices and solve them to optimality. We observe that optimal policies are of threshold type. On the other hand, for the single-patient problem, we are able to prove the existence of an optimal threshold policy under certain conditions, which states that a transfer request is made if and only if the patient's score is below a certain score.

## 1.4 AN EXPLICIT STOCHASTIC MODEL OF INTENSIVE CARE UNIT LENGTH OF STAYS

ICU LOS is a key indicator for understanding the clinical and operational characteristics of the system. An accurate LOS prediction model would enable managers and clinicians to better plan interventions and utilize resources [6]. However, LOS is a complex metric that is influenced by many factors including demographics, patient types, varying physical conditions and treatment methods as well as discharge decisions and transfer delays. Typically, the average LOS is used to quantify the performance, which is far from being sufficient for planning purposes [112]. The LOS data generally has high variability, its distribution is positively skewed and has a long tail [19, 104, 147]. This further complicates the predictions and may lead to poor planning and hence poor use of expensive resources.

In Chapter 5, we present a linear regression model to identify clinical factors influencing LOS. We demonstrate that a simple statistical model is not sufficient to predict the LOS. The classical approach of employing statistical models, which use Day 1 information, lacks the highly dynamic nature of the patient physiology, ignores the fact that the whole LOS cannot be attributed to medical reasons, considering long transfer delays. Furthermore, we build a classification model to identify longer LOS's. We show that our model presents a good discrimination.

Understanding LOS dynamics can be considered as the core of understanding the ICU. For that purpose, we present an explicit stochastic model of LOS that is a function of patient physiology, as well as the transfer delay dynamics. We characterize the LOS process for different stochastic score processes; a discrete time Markov chain (DTMC) and a continuous time Markov chain (CTMC). Next, we test the estimation power of our model on our data set. We compare the goodness of fit performance of our model with commonly employed phase-type distribution, the Coxian. We show that the stochastic model slightly underestimates the LOS, whereas it captures the shape of the LOS distribution. The Coxian distribution provides the best performance in matching the moments and capturing the long tail. Although there is room for improvement, we demonstrate that a single dimensional descriptive stochastic model can moderately predict a highly complex outcome.

The remainder of this dissertation is organized as follows. In Chapter 2, we present the relevant literature. In Chapter 3, we describe our prediction models to estimate readmission and mortality. We present an optimization model and a numerical study comparing the current practice with the anticipative transfer policy in Chapter 4 presents. We discuss an explicit LOS model in Chapter 5, and conclude the dissertation by highlighting future research directions in Chapter 6.

# 2.0 LITERATURE REVIEW

In this chapter, we review the literature related to the problems and methodologies discussed in this dissertation. In Section 2.1, we review outcome prediction models. We present a survey of OR studies related to ICUs and other hospital units and a partial review of approximate dynamic programming and robust dynamic programming techniques, in Section 2.2. We present relevant length of stay models in Section 2.3.

## 2.1 PREDICTION MODELS

There is an extensive medical literature that uses critical care scores to predict readmission and mortality as well as other demographic information. These include:

- APACHE, an initial risk classification of severely ill hospitalized ICU patients [92].
- Simplified Acute Physiology Score (SAPS), which assesses admission severity [96].
- Sequential Organ Failure Assessment (SOFA), which tracks a patient's status during the stay in an ICU [150].
- Systemic Inflammatory Response Syndrome (SIRS), which models the severity of sepsis and septic shock [12].

APACHE, SAPS and SOFA are shown to be good at predicting mortality [52, 91, 150, 152] as well as other demographic information. SAPS II and APACHE are widely used score for mortality prediction [152]. The APACHE score is updated to improve predictions [91, 92]. Vincent et al. [152] present a detailed comparison of scores. However, none of the models focus on the population of patients that have survived the ICU, let alone were discharged

alive from the hospital. Our model provides an excellent discrimination and hence good at predicting mortality. Admission prognostic score APACHE is found to be a significant variable in predicting readmission [18, 57, 121] . Frost et al. [57], dos Santos et al. [45], and Woldhek et al. [156] are examples of studies that employ SOFA score for predicting readmission.

Readmission rates are proposed a benchmark measure for quality of care [93]. Age, gender, severity of illnesses, patient type, and comorbidities have been identified as factors influencing readmission [7, 30, 57, 159]. Other factors, e.g., the influence of discharging patients at nights or weekends, or mechanical ventilation usage on the readmission rates have attracted significant interest. Discharge at nights or weekends shown to lead more readmission or death due to lower staffing [16, 28, 61, 137, 16, 121], however our results indicate that they are not significant in predicting death or readmission. Similar to [2, 56, 153], we observe that mechanically ventilated patients have higher readmission and mortality rates.

Almost all models use scores computed from the first day of ICU data, or medical information at discharge, e.g., [7, 18, 81, 83, 98, 113, 121, 124, 159]. Most of these models do not consider dynamic health except Clermont et al. [30], which employs simulation techniques to estimate mortality and LOS, and Ferreira et al. [52], which shows that repeated measurements of SOFA score improve mortality prediction. Le Gall et al. [95] and Rue et al. [136] show that current day information is the most informative in predicting mortality, hence updating information improves the evaluation of risk of death through the stay. Similarly, we show that health evolution plays a key role in patient outcomes, and risk predictions at admission are not sufficient. Our models differ from [18, 57, 121] that admission prognostic score APACHE is no longer a significant variable as the information is carried by dynamic changes of other clinical figures. Our results also verify [7, 18, 30, 121, 124, 156, 159] that SOFA by itself is not enough to predict readmission.

Multivariate logistic regression is typically used for predicting readmission and death. Our readmission model is better in discriminating readmission with existing models [7, 57, 18, 121]. We believe that the inclusion of dynamic data at least partially explains improved performance. We do not set a time limit on readmission or post-ICU mortality in our predic-

tion models, whereas [121, 156, 18] predict readmission over a predetermined duration after discharge. Other machine learning techniques, e.g., Naive-Bayes, artificial neural networks (ANN), support vector machines (SVM) are employed to predict mortality. Dybowski et al. [47] and Nimgaonkar et al. [118] compare artificial neural networks and logistic regression and conclude that ANN performs better in predicting mortality. However, Clermont et al. [29] and Doig et al. [43] show that they perform similarly. Ribas et al. [132] report that SVM performs significantly better in discriminating mortality than logistic regression. We compare our logistic regression [75] models with Naive-Bayes models [46]. We show that logistic regression performs better than Naive-Bayes approach in predicting both death and mortality. We utilize Receiver Operating Characteristics curves (ROC), where the Area Under Curve (AUC) measure is used to compare the discriminatory power of different models [71, 125, 138], which is commonly accepted in the statistical learning community.

## 2.2 HOSPITAL OPERATION MODELS

Mathematical models of ICUs and other hospital units have gained increasing attention in recent years. The dynamic, stochastic and complex nature of hospitals makes the problem challenging. Two primary approaches are utilized to estimate performance measures and answer design and policy questions: discrete event simulation and queueing models. Simulation models are the most common in the healthcare literature thanks to its ability to capture complex dynamics (see Jacobson et al. [79] and the references therein). Design alternatives and a small number of different policies can easily be evaluated via simulation. Simulation models of an ICU can estimate the performance measures (e.g., bed utilizations, expected waiting times, rejection probabilities, patient throughput, etc.), and thus, different bed allocations, admission rules, elective patient schedules, or staffing policies are explored with an objective to optimize patient flows or reduce costs while maintaining quality care [31, 33, 70, 101, 102, 133, 143, 155]. However, previous studies do not consider the dynamics of patients' physiology in ICUs. Unlike simulation models, queueing models provide simple analytical expressions that can be incorporated into optimization models, although queue-

ing models often have stronger assumptions for analytical tractability. Green [65] provides a survey of queueing models of hospital units. For the most part, while traditional queueing models are found in the literature (e.g., [38, 64, 88, 110]), no existing queueing (network) models consider physiological changes in patient flows.

Operational models considering admission, discharge or transfer decisions are less studied than nurse scheduling, unit capacity planning or configuration models. Some examples of ICU admission models are Armony et al. [89] and Shmueli et al. [144]. Discharge models are closer to our approach. Lowery et al. [102] is the first to consider early discharges in a simulation model. Dobson et al. [42] present a stochastic model of an ICU where early discharge (bumping) can be used whenever the ICU is full. They determine the probability of a patient being bumped and the expected remaining LOS for a bumped patient. In [42], bumping decisions are based on the remaining LOS, i.e., the patient with the least remaining LOS is chosen to be bumped. Instead, we employ patient health status as a proxy for taking transfer decisions, hence the LOS is not endogenously modeled. Early discharges might undesirably influence the patient's health and hence she/he might have to be readmitted to the ICU [7]. Chan et al. [22] study early discharge policies by considering the readmission phenomenon. They use dynamic programming to demonstrate that a certain index rule based on readmission risk is close to optimal. Chan et al. [24] present a fluid model of an ICU with readmissions (Erlang-R system), in which the service rate (treatment rate in the ICU) might be increased when the ICU is full, and they present the steady-state behavior of the system. Recall that our model assumes that patients cannot be transferred unless they are clinically cleared. This guarantees a certain quality of care and alleviates adverse effects of early discharges such as increased readmissions and mortalities as shown in [84] for some cardiac care units. Kim et al. [90] emphasize the importance of tracking patients' health condition as they show patients in poorer acuity at the time of discharge have higher mortality rates and post ICU LOS. Hence to accurately characterize ICU workload, patient health conditions need to be considered.

ICUs should not be considered as isolated units, since congestions in the downstream units have certain implications in the ICUs. Armony et al. [4] analyze the role of stepdown units through a queueing network model, and examine the capacity tradeoff between the

ICUs and the stepdown units. Although our approach does not explicitly model a capacity-constrained stepdown unit, we implicitly model the downstream capacity through extended delays, since the main cause of transfer delays is the congestions in the downstream units.

Although there is an extensive literature focusing on transfer delays when admitting to the ICU, e.g., [21, 23, 134], almost no study focus on ICU transfer delays when discharging patients to downstream units. Christensen [27] is the first to study transfer delay through a simulation model, offering a policy to reduce transfer delay where the transfer operations start as soon as the patient is clinically ready instead of watching whether other patients need a downstream bed, and rolling a 24-hour medical clearance process. Mathews and Long [107] also study an ICU-stepdown unit system through a simulation model, and show that significant improvement can be attained through reducing transfer delays. Hu et al. [76] examine proactive decisions where they consider transfers to the ICU from other units. They provide a predictive model to identify patients staying in lower level of care units with risk for deterioration, and employ a simulation model to study the benefits of proactive transfers.

Approximate dynamic programming (ADP) is a scheme for modeling and solving complex, large-scale dynamic and generally stochastic problems, e.g., [1, 39, 40, 128, 158]. Recently, these methods are employed in healthcare applications. Some examples include management of dialysis therapy [97], ambulance redeployment problem [108], and HIV drug allocation [87]. Our method can be considered as a state aggregation method [39, 40, 128]. However, we generate an approximate policy directly, instead of using basis functions to approximate the value function.

Robust dynamic programs are employed to model ambiguity in the parameters defining a dynamic program. There is an increasing attention of numerous researchers on robust dynamic programs since model parameters are not always accurately predictable in many real life problems. White and Eldieb [154] obtain optimal policies for different realizations of the uncertain rewards and determined non-dominated policies. Iyengar [78] and Nilim and El-Ghaoui [117] study the problems where the uncertainty is on the transition probabilities. Iyengar [78] and Nilim and El-Ghaoui [117] tackle the uncertainty by computing worst case scenarios. In this dissertation, we modify the methods provided in [117] to solve the robust problem.

## 2.3 LENGTH OF STAY MODELS

Statistical models to predict LOS have received increasing attention in the medical literature [8, 83, 124, 159]. Linear regression is one of the most commonly used techniques to identify factors influencing LOS. In most of these models, physiology is limited to Day 1 information. Such static statistical models ignore the dynamic nature and high variability of patient conditions; exceptions include [30, 98]. Data-mining techniques (e.g., logistic regression, Naive-Bayes, artificial neural networks, support vector machines) are also extensively used to identify patients experiencing prolonged LOS [17, 54, 68, 72, 98, 114, 119]. These models generally classify patients into disjoint sets of LOS ranges (e.g., short, medium, long) rather than providing point estimates. Verburg et al. [148] presents a systematic review of ICU LOS prediction models.

Various probabilistic models including classical distributions (exponential, lognormal, gamma, etc.), phase-type distributions, and Markov models are proposed to characterize LOS distribution. The exponential distribution is common in the operations research literature to facilitate the analytical tractability of queueing models. However, Griffiths et al. [67] and many other studies report that the exponential assumption is not valid for ICUs in practice. Debruin et al. [37] and Griffiths et al. [67] employ the hyper-exponential distribution to model LOS. Faddy [50] and Marazzi et al. [103] demonstrate that lognormal distribution is a good fit and superior to Gamma and Weibull for modeling LOS.

Phase-type distributions describe the time to absorption of a finite state Markov chain with an absorbing state and the process starts in a transient state [115]. The Coxian distribution [35] is special case of the phase-type distribution, in which the transient states are ordered. Faddy and McClean [51] use the Coxian distribution to model LOS. Generalized phase-type and conditional phase-type distributions are employed to model LOS for geriatric patients in hospital [62, 105, 106]. Fackrell [49] fits a generalized phase-type distribution with six transient states and compare it with other distributions, e.g., exponential, hyper-exponential, generalized Erlang and the Coxian. He shows that generalized phase-type distribution outperforms alternatives, whereas the performance of the Coxian distribution is close. Garg et al. [59] model LOS in a stroke unit as phase-type distribution with multiple

absorbing states. Although phase-type distributions provide better results, they have more parameters to estimate and the estimation methods are complex. Moreover, there is no standard package to carry estimation procedures, hence they are not extensively used. Our stochastic model can also be considered as a phase-type distribution. The main difference is that the parametrization of the Markov Chain is performed by the score model exogenously, rather than a log-likelihood algorithm fitting the data. Furthermore, switching to the absorbing state is determined by transfer delay dynamics and threshold policy. Almost all models ignore the fact that not all LOS is due to health; patients might need to wait in the ICU due to delays in the transfer process or blocking in the ICU. This might yield an overestimation of LOS and hence lead to incorrect analysis. Bountourelis et al. [13] are first to estimate medical LOS by decoupling it from LOS due to blocking delays.

# 3.0 PREDICTING INTENSIVE CARE UNIT READMISSION AND MORTALITY

## 3.1 INTRODUCTION

In this chapter, we develop and validate robust ICU readmission and mortality models based on dynamic variables extracted from an electronic health record (EHR) from the University of Pittsburgh Medical Center, which contains both medical and bed flow information of 16,059 patients over a six-year period that identifies potentially modifiable factors of readmission. The data set includes extensive medical information as extracted from a detailed electronic health record from their admissions to their discharges. We track the changes of physiological indicators dynamically through ICU stays, and leveraged this information to predict readmission or mortality upon ICU discharge. We compare the performance of the models with existing studies and show that our models predict much better than existing models in the literature. We also develop and validate ICU mortality models, which when externally validated in different settings, could contribute to decisions regarding intensity of post-ICU care, hospital discharge decisions, and home-maintenance services requirements.

All studies in the literature use static medical information about the patient to estimate the likelihood of readmission and mortality. Critical care scores (e.g., APACHE, SAPS, SOFA, and SIRS) are the essential components of many of these models. The outcome of the transfer depends on the history of the patient's health while in the ICU and that there is a need for a better transfer score capable of summarizing health evolution in the ICU. We further show that some widely used scores such as APACHE, SOFA, and SIRS alone are not sufficient to describe patients' health, and are incapable of accurately predicting readmissions or mortalities after transferring to a lower level care unit.

The remainder of the chapter is organized as follows. In Section 3.2, we discuss statistical techniques employed throughout the chapter and present a detailed description of our data set. We present our prediction models and main results in Section 3.3. We give an extensive discussion and presents some clinical insights in Section 3.4.

## 3.2 MATERIAL AND METHODS

### 3.2.1 Data Sources

We use the EHR-derived High DENsity Intensive Care (HIDENIC) database of all patients admitted to one of 12 ICUs within the University of Pittsburgh Medical Center Health System between 2001 and 2008. HIDENIC is a HIPAA compliant, limited data set that contains detailed demographic, diagnostic, physiologic, laboratory, and drug administration and outcome information on a source population of 54,811 ICU admissions [86, 99, 142, 146], linked to Social Security Death Master File (SSDMF) through 2014. The study is conducted under proper approval of the University of Pittsburgh Institutional Review Board. Vitals are typically available hourly in HIDENIC while a patient is in the ICU. We compute SOFA scores across six organ systems (respiratory, cardiovascular, hepatic, coagulation, renal and neurological) using a standard definition for every non-overlapping six hours interval throughout ICU stays [52, 139, 150]. Patients for whom there is no data to generate SOFA scores for a particular organ system are excluded from the study. Where data to generate score exist, but are missing at a specific time point, we linearly interpolate missing SOFA scores, but refrain from extrapolation. In case of multiple entries relevant to a system score over a six-hour interval, we choose the value generating the highest (worst) score. We sum the six system-specific SOFA scores in our analysis to compute total SOFA. For each patient and for each system-specific SOFA, we compute the following five values to represent the trend of the SOFA score; (1) the time average of the score through ICU stay, (2) the variance of the score through ICU stay, (3) the initial score (first 6 hours of ICU admission), (4) the highest score through ICU stay, and (5) the last score prior to discharge. We also implement

an algorithm to compute a Physiologic Severity Score (APS) based on the acute physiology components of the APACHE III score, which is updated every 8 hours through a patients ICU stay. We further compute the SIRS score of each patient 48 hours prior to their discharge. We also employ demographic, clinical and patient flow information as presented in Table 3.1. We further limit our analysis to patients with complete data (other than the interpolated SOFA scores). The composition of the source and study populations are depicted in Figure 3.1.

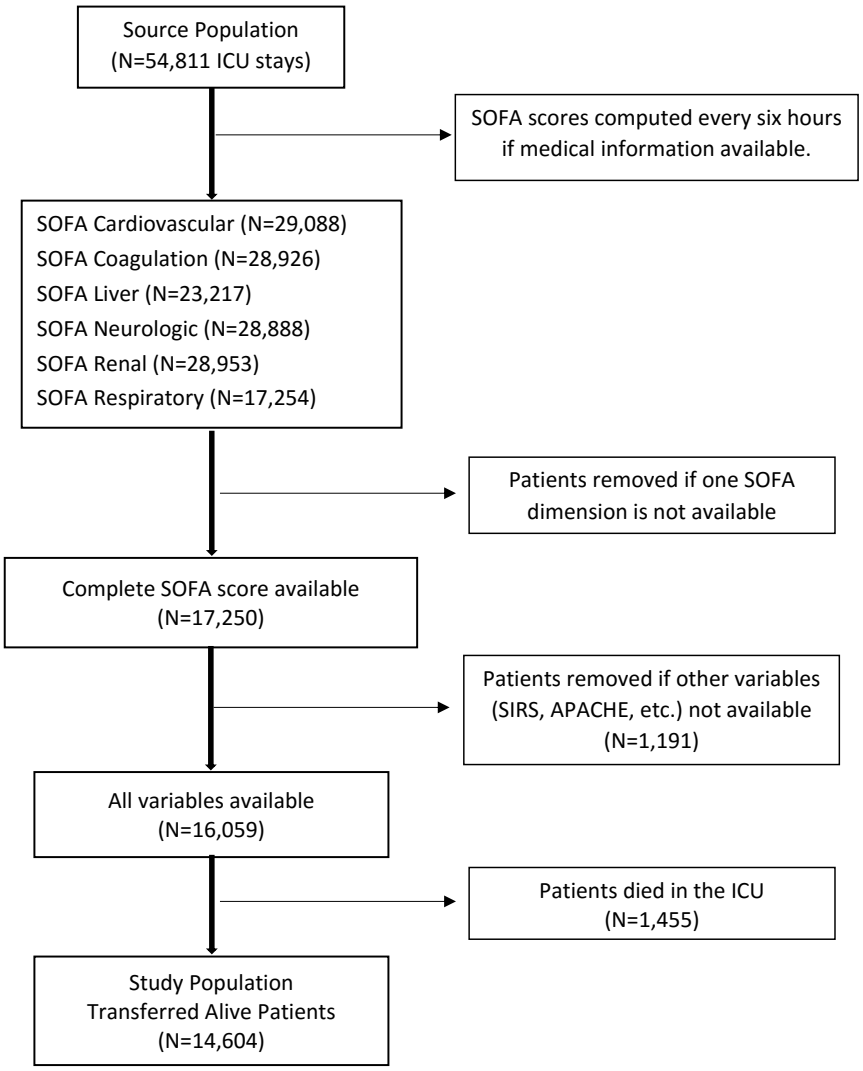



Figure 3.1: Cohort selection rules

### 3.2.2 Model Development and Validation

We employ two classification schemes to estimate ICU readmission and mortality probabilities at the time of ICU discharge: (1) multivariate logistic regression and (2) Naive-Bayes model. Multivariate logistic regression is a statistical technique for analyzing a data set, in which more than one independent variables ($X_i,\ i = 1, \ldots, n$) that determine a binary dependent variable ($Y$). Formally, logistic regression fits a multiple linear regression function defined as

$$\log\left[\frac{P(Y=1|X_1, X_2, \ldots X_n)}{1 - P(Y=1|X_1, X_2, \ldots X_n)}\right] = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n.$$

Thus, for a given input $X_i, i = 1, \ldots, n$, equation (3.1) provides the probability of dependent variable taking value 1, i.e., in our setting patient being readmitted or die.

$$P(Y=1|X_1, X_2, \ldots X_n) = \frac{e^{(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n)}}{1 + e^{(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n)}}. \tag{3.1}$$

Parameters are determined by maximum likelihood estimates. We refer the reader [75] for further details.

Naive-Bayes model is a simple classification technique, which assumes variables $X_i,\ i = 1, \ldots, n$ are conditionally independent from each other for a given dependent variable $Y$. This approach relies on Bayes' theorem. Hence, for a given input $X_i,\ i = 1, \ldots, n$, equation (3.2) provides the probability of dependent variable taking value 1,

$$P(Y=1|X_1, X_2, \ldots X_n) = \frac{P(Y=1)\prod_i P(X_i|Y=1)}{\sum_{j=0}^{1} P(Y=j)\prod_i P(X_i|Y=j)}, \tag{3.2}$$

where the distributions of $P(Y)$ and $P(X_i|Y),\ i = 1, \ldots, n$ are estimated from the training data; see [46] for details.

In this study, we assume that a readmission occurs if a patient is readmitted more than 24 hours after ICU discharge. Patients readmitted within 24 hours of discharged are not considered discharged from the ICU. We preselect candidate risk factors based on univariate statistical association with ICU readmission or post-ICU mortality using a t-test with a

significance value of 0.05 for continuous variables and chi-square test of homogeneity for proportions of categorical variables. Table 3.1 and Table 3.2 display the corresponding p-values. We note that SIRS$>$ 2 is an indicator variable as to whether the SIRS score is greater than 2 or not, and unit utilization upon discharge is the unit occupancy rate (beds occupied/bed capacity) at the time of discharge.

For variable selection, we employ a step-wise backward elimination method with a p-value threshold as 0.05. We use Walds $\chi^2$ and log-likelihood statistics to test the significance of individual variables in the logistic regression model. We further employ the Akaike information criterion (AIC) to determine the best models. We develop two ICU readmission models, one including ICU non-survivors and one excluding ICU non-survivors for each technique. We present the odd-ratios, 95% confidence intervals and p-values for the final regression models. We assess the goodness of fit of the models using Hosmer-Lemeshow test statistics. We determine the discrimination power of the models using the AUC of the ROC curves, which is a broadly accepted measure for statistical model comparison, particularly in the statistical learning community [71, 125, 138]. For validation purposes, we perform 10-fold cross-validation, where in each fold a model is developed on 90% of cohort size and tested on the remaining 10%. We repeat the 10-fold analysis 50 times, and hence the confidence intervals are determined by a sample of 500 AUC values for each model. We use the R statistical package version 2.14.1 for analysis.

### 3.2.3 Study Population

The source population consists of 46,169 patients with 54,811 ICU admissions. Of these, 5,107 died during their ICU stay and 1,849 died after ICU discharge but during the index hospitalization. The study population consists of 16,059 ICU admissions of patients with clinical and bed flow. Of those, 14,604 were discharged alive from the ICU (Figure 3.1). Of ICU admissions with live discharges, there were 2,475 readmission episodes, thus an ICU readmission rate of 15.4%. Of those discharged alive, a further 1,628 patients died in the hospital after discharge. Table 3.1 summarizes the basic characteristics of the patients for both readmitted and non-readmitted patients, and Table 3.2 summarizes the basic charac-

teristics of the patients for both surviving and non-surviving patients. Of those discharged alive, 58% were transferred to a stepdown unit, 22% of them transferred to a floor unit and the remaining patients were directly discharged from the hospital to a variety of destinations. Only 37% of admissions are medical, reflecting the local patient case-mix. A total of 3,083 patients died in the hospital, yielding a 19% hospital mortality.

Table 3.1: Comparison for readmitted and non-readmitted patients

| | Patients | Readmit. | Non-Readmit. | p-value | Non-Readmit. (Survived) |
|---|---|---|---|---|---|
| N | 16,059 | 2,475 | 13,584 | | 12,129 |
| Age, mean (sd) | 59 (17) | 61 (16) | 59 (18) | $< 0.001$ | 58 (18) |
| Male (%) | 9,187 (57.0%) | 1,420 (57.0%) | 7,767 (57.0%) | 0.847 | 6,988 (58.0%) |
| LOS (hr), mean (sd) | 171 (265) | 163 (248) | 172 (268) | 0.116 | 165 (245) |
| LOS before ICU stay (hr), mean (sd) | 29.5 (90.8) | 42.8 (117.3) | 27.1 (84.9) | $< 0.001$ | 25.7 (79.4) |
| Number of Previous ICU stays, mean | 0.14 | 0.19 | 0.13 | $< 0.001$ | 0.13 |
| Type of Patient | | | | $< 0.001$ | |
| Medical (%) | 5,956 (37.0%) | 746 (30.0%) | 5,210 (38.0%) | | 4,403 (36.0%) |
| Surgical (%) | 10,103 (63.0%) | 1,729 (70.0%) | 8,374 (62.0%) | | 7,726 (64.0%) |
| ICU Type | | | | $< 0.001$ | |
| CCU, n (%) | 1,907 (11.8%) | 307 (12.4%) | 1,600 (11.7%) | | 1,329 (10.9%) |
| Cardiothoracic, n (%) | 1,557 (9.6%) | 304 (12.2%) | 1,253 (9.2%) | | 1,216 (10.0%) |
| Medical, n (%) | 1,920 (11.9%) | 415 (16.7%) | 1,505 (11.0%) | | 1,287 (10.6%) |
| Neuro, n (%) | 3,336 (20.7%) | 349 (14.1%) | 2,987 (21.9%) | | 2,617 (21.5%) |
| Surgical, n (%) | 3,499 (21.7%) | 446 (18.0%) | 3,053 (22.4%) | | 2,848 (23.4%) |
| Transplant, n (%) | 3,387 (21.0%) | 534 (21.5%) | 2,853 (21.0%) | | 2,527 (20.8%) |
| Trauma, n (%) | 453 (2.8%) | 120 (4.8%) | 333 (2.4%) | | 305 (2.5%) |
| Origin Level of Care | | | | $< 0.001$ | |
| Stepdown Unit, n (%) | 3,962 (24.6%) | 731 (29.5%) | 3,231 (23.7%) | | 2,910 (24%) |
| Floor, n (%) | 2,389 (14.8%) | 403 (16.2%) | 1,986 (14.6%) | | 1,810 (14.9%) |
| Other, n (%) | 9,708 (60.4%) | 1,341 (54.1%) | 8,367 (61.5%) | | 7,409 (61.0%) |
| Destination Level of Care | | | | $< 0.001$ | |
| Stepdown Unit, n (%) | 9,271 (57.7%) | 1,748 (70.6%) | 7,523 (55.3%) | | 7,523 (62.0%) |
| Floor, n (%) | 3,514 (21.8%) | 727 (29.3%) | 2,787 (20.5%) | | 2,787 (22.9%) |
| Out, n (%) | 3,274 (20.3%) | 0 (0.0%) | 3,274 (24.1%) | | 1,819 (15.0%) |
| Discharge at night, n (%) | 4,160 (25.9%) | 628 (25.3%) | 3,532 (26.0%) | 0.523 | 2,994 (25.0%) |
| Discharge during weekend, n (%) | 4,293 (26.7%) | 686 (27.7%) | 3,607 (26.5%) | 0.228 | 3,190 (26.0%) |
| Procedures used during the ICU stay | | | | | |
| CVC, n (%) | 8,001 (49.8%) | 1,113 (44.9%) | 6,888 (50.7%) | $< 0.001$ | 5,929 (48.8%) |
| MV (last 24 hours), n (%) | 6,217 (38.7%) | 460 (18.5%) | 5,757 (42.3%) | $< 0.001$ | 4,455 (36.7%) |
| Charlson comorbidity index | | | 0.428 | | |
| 0, n (%) | 14,533 (90.5%) | 2,228 (90.0%) | 12,305 (90.5%) | | 10,998 (90.6%) |
| 1, n (%) | 1,474 (9.1%) | 241 (9.7%) | 1,233 (9%) | | 1,088 (8,9%) |
| $>= 2$, n (%) | 52 (0.3%) | 6 (0.2%) | 46 (0.3%) | | 43 (0.3%) |
| Unit Utilization, mean (%, sd) | 71.0 (19.0) % | 64.0 (24.0) % | 72.0 (17.0) % | $< 0.001$ | 72.0 (18.0) % |
| APACHE III (Admission), mean (sd) | 83.4 (13.6) | 82.6 (11.7) | 83. (13.9) | $< 0.001$ | 82.2 (13.0) |
| SIRS$>$ 2 last 48 hours, n (%) | 6,829 (42.5%) | 1,110 (44.8%) | 5,71 (42.1%) | 0.011 | 4,695 (38.7%) |

Table 3.2: Comparison for surviving and non-surviving patients

| | Non-surviving Patients | Surviving Patients | p-value |
|---|---|---|---|
| N | 3,083 | 12,976 | |
| Age, mean (sd) | 66 (15) | 57(17) | $< 0.001$ |
| Male (%) | 1,665 (54.0%) | 7,522 (58.0%) | $< 0.001$ |
| LOS (hr), mean (sd) | 211 (368) | 161 (233) | $< 0.001$ |
| LOS before ICU stay (hr), mean (sd) | 40.45 (124.8) | 26.94 (80.5) | $< 0.001$ |
| Number of Previous ICU stays, mean | 0.13 | 0.14 | 0.338 |
| Type of Patient | | | $< 0.001$ |
| Medical (%) | 1,601 (52.0%) | 4,355 (34.0%) | |
| Surgical (%) | 1,482 (48.0%) | 8,621 (66.0%) | |
| ICU Type | | | $< 0.001$ |
| Cardiac, n (%) | 523 (16.9%) | 1,384 (10.6%) | |
| Cardiac-T, n (%) | 165 (5.3%) | 1,392 (10.7%) | |
| Medical, n (%) | 535 (17.3%) | 1,385 (10.6%) | |
| Neuro, n (%) | 807 (26.1%) | 2,529 (19.4%) | |
| Surgical, n (%) | 423 (13.7%) | 3,076 (23.7%) | |
| Transplant, n (%) | 554 (17.9%) | 2,833 (21.8%) | |
| Trauma, n (%) | 76 (2.4%) | 377 (2.9%) | |
| Origin Level of Care | | | 0.001 |
| Stepdown Unit, n (%) | 762 (24.7%) | 3,200 (24.6%) | |
| Floor, n (%) | 395 (12.8%) | 1,994 (15.3%) | |
| Out, n (%) | 1,926 (62.4%) | 7,782 (59.9%) | |
| Destination Level of Care | | | $< 0.001$ |
| Stepdown Unit, n (%) | 744 (24.1%) | 8,527 (65.7%) | |
| Floor, n (%) | 303 (9.8%) | 3,211 (24.0%) | |
| Out, n (%) | 2,036 (66.0%) | 1,238 (9.5%) | |
| Discharge at night, n (%) | 1,124 (36.4%) | 3,036 (23.4%) | $< 0.001$ |
| Discharge during weekend, n (%) | 883 (28.6%) | 3,410 (26.2%) | 0.009 |
| Procedures used during the ICU stay | | | |
| CVC, n (%) | 1,723 (55.8%) | 6,278 (48.3%) | $< 0.001$ |
| MV (last 24 hours), n (%) | 2,157 (69.9%) | 4,060 (31.2%) | $< 0.001$ |
| Charlson co-morbidity index | | | 0.126 |
| 0, n (%) | 2,780 (90.1%) | 1,1753 (90.5%) | |
| 1, n (%) | 298 (9.6%) | 1,176 (9.0%) | |
| $>= 2$, n (%) | 5 (0.1%) | 47 (0.37%) | |
| Unit Utilization upon discharge, mean (sd) | 0.7 (0.1) | 0.7 (0.1) | 0.700 |
| APACHE III (Admission), mean (sd) | 92.1 (16.7) | 81.3 (11.9) | $< 0.001$ |
| SIRS$>$ 2 last 48 hours, n (%) | 1,906 (61.8%) | 4,923 (37.9%) | $< 0.001$ |

The mean admission APACHE-III score is 83.40, with a standard deviation of 13.64, where the score ranges from 0 to 299, with severity increasing as the score increases. Figure 3.2 presents the histogram of admission APACHE-III scores.

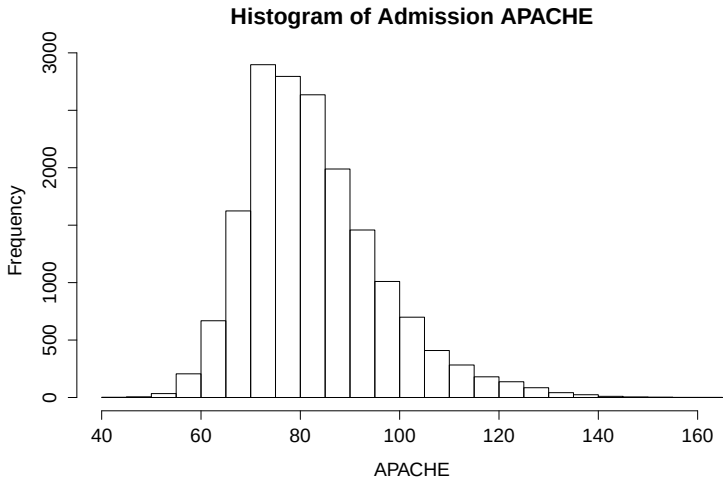



Figure 3.2: Histogram of admission APACHE-III score

Figure 3.3 presents histograms of the average SOFA score (the time average SOFA score of a patient through his/her LOS), the initial SOFA score (first 6 hours of ICU admission), the highest SOFA score (maximum SOFA score during the ICU LOS), and the discharge SOFA score (the score when the patient is transferred or discharged from the ICU).

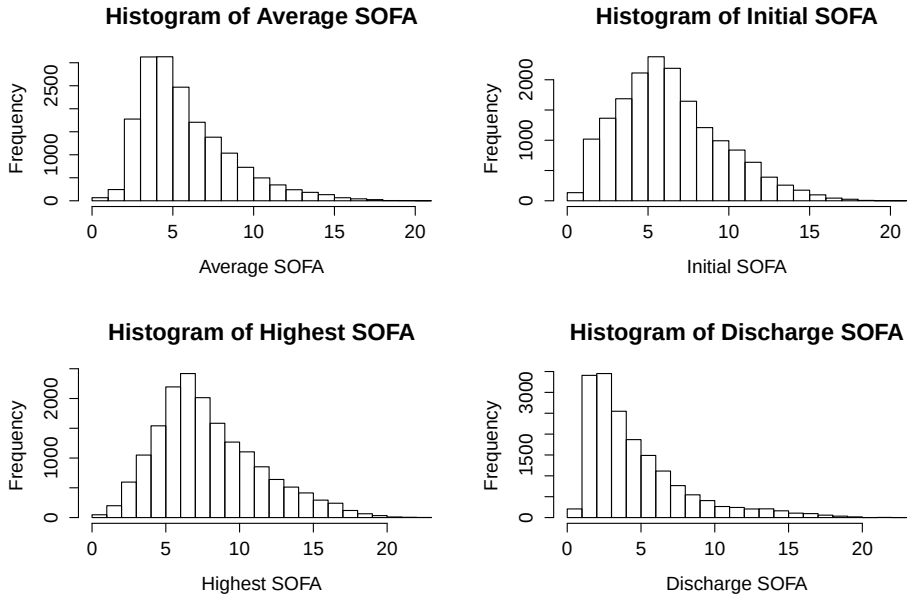



Figure 3.3: Histograms of the average, initial, maximum and discharge total SOFA scores

In the 16,059 patients with complete data, all metrics of the total SOFA were higher in non-readmitted patients, owing to the presence in this cohort of ICU non-survivors (Table 3.3), whose scores were much higher.

Table 3.3: SOFA trend values for survivors and non-survivors

| | **Surviving** | **Non-surviving** | **p-value** |
|---|---|---|---|
| Initial SOFA, mean (sd) | 6.72 (3.14) | 8.78 (3.52) | $< 0.001$ |
| Maximum SOFA, mean (sd) | 8.06 (3.29) | 12.27 (3.76) | $< 0.001$ |
| Average SOFA, mean (sd) | 5.51 (2.53) | 9.57 (3.31) | $< 0.001$ |
| Variance SOFA, mean (sd) | 2.23 (2.36) | 2.94 (3.34) | $< 0.001$ |
| Discharge SOFA, mean (sd) | 4.57 (2.75) | 10.88 (3.97) | $< 0.001$ |

Among ICU survivors, those readmitted had comparable initial and mean total SOFA scores, but higher average and discharge scores (Table 3.4).

Table 3.4: SOFA trend values for readmitted and non-readmitted patients

| | **Readmitted** | **Non-Readmitted** | **p-value** |
|---|---|---|---|
| Initial SOFA, mean (sd) | 4.96 (3.25) | 4.49 (3.12) | 0.448 |
| Maximum SOFA, mean (sd) | 8.13 (3.43) | 8.04 (3.26) | 0.239 |
| Average SOFA, mean (sd) | 5.76 (2.56) | 5.45 (2.52) | $< 0.001$ |
| Variance SOFA, mean (sd) | 1.93 (2.21) | 2.29 (2.38) | $< 0.001$ |
| Discharge SOFA, mean (sd) | 4.96 (2.54) | 4.48 (2.79) | $< 0.001$ |

## 3.3 RESULTS

### 3.3.1 Models for Predicting Intensive Care Unit Readmission

We first present logistic regression model results. Table 3.5 summarizes the 24 variables that independently predicted post-ICU readmission (N=14,604), their estimates, and p-values. The readmission regression model exhibits calibration Hosmer-Lemeshow with p-value 0.42, and good discrimination (ROC-AUC) of 0.765 with a 95% bootstrap confidence interval

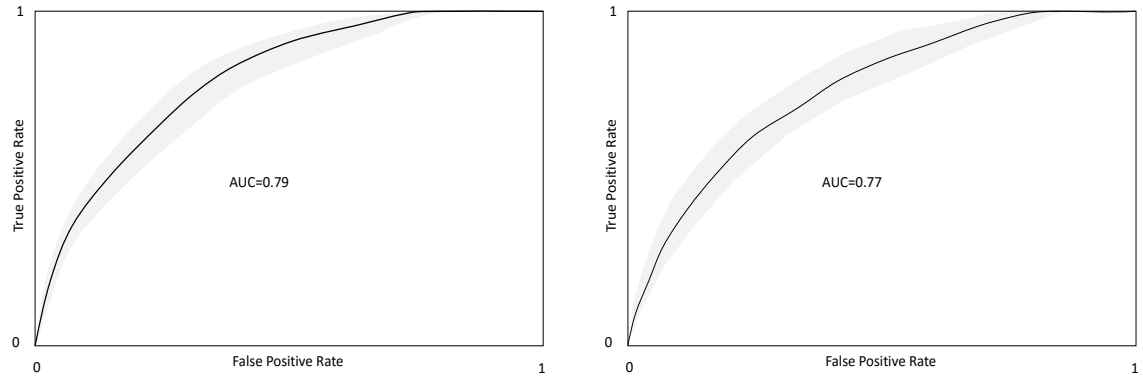
[0.766,0.764] . A similar model based on a population, which includes ICU non-survivors (N=16,059) exhibits slightly better performance with AUC of 0.79 (0.789,0.791) and good calibration $p = 0.25$. The list of predictors are same, which are presented in Table 3.6. Figure 3.4 presents ROC-AUC curves for both readmission logistic regression models.

Table 3.5: Predictors of ICU logistic regression readmission model in a cohort excluding ICU non-survivors (N=14,604)

| Variables | Odd-ratio (95% Confidence Interval) | p-value |
|---|---|---|
| (Intercept) | 0.000 (0.000 - 0.000) | 0.890 |
| ICU Type Medical | 1.378 (1.137- 1.685) | 0.001 |
| ICU Type Trauma | 1.445 (1.074 - 1.937) | 0.014 |
| Age | 1.011 (1.008 - 1.014) | $< 0.001$ |
| ICU Length of Stay | 1.001 (1.000 - 1.001) | $< 0.001$ |
| Initial SOFA Liver | 0.868 (0.766 - 0.984) | 0.027 |
| Initial SOFA Respiratory | 0.824 (0.755 - 0.899) | $< 0.001$ |
| Discharge SOFA Liver | 1.216 (1.084 - 1.366) | 0.001 |
| Discharge SOFA Neurological | 1.465 (1.357 - 1.582) | $< 0.001$ |
| Maximum SOFA Liver | 1.259 (1.087 - 1.455) | 0.002 |
| Max SOFA Neurological | 0.832 (0.788 - 0.878) | $< 0.001$ |
| Max SOFA Renal | 1.478 (1.264 - 1.727) | $< 0.001$ |
| Average SOFA Coagulation | 1.082 (1.010 - 1.159) | 0.025 |
| Average SOFA Neurological | 0.768 (0.689 - 0.855) | $< 0.001$ |
| Average SOFA Renal | 0.779 (0.665 - 0.915) | 0.002 |
| Average SOFA Respiratory | 1.510 (1.279 - 1.788) | $< 0.001$ |
| Average SOFA Cardiovascular | 1.128 (1.027 - 1.239) | 0.012 |
| Variance SOFA Renal | 0.662 (0.466 - 0.932) | 0.020 |
| Central Venous Catheter | 0.709 (0.635 - 0.791) | $< 0.001$ |
| Patient Type Surgical | 1.563 (1.405 - 1.741) | $< 0.001$ |
| LOS before ICU admission | 1.001 (1.001 - 1.001) | $< 0.001$ |
| Number of previous ICU admissions | 1.203 (1.097 - 1.318) | $< 0.001$ |
| SIRS$>$ 2 48 hours prior to discharge | 1.341 (1.218 - 1.476) | $< 0.001$ |
| Mechanical Ventilation within 24 hours of discharge | 0.544 (0.475 - 0.620) | $< 0.001$ |
| Unit Utilization upon discharge (%) | 0.184 (0.139 - 0.243) | $< 0.001$ |

Table 3.6: Predictors of ICU logistic regression readmission model in a cohort including ICU non-survivors (N=16,059)

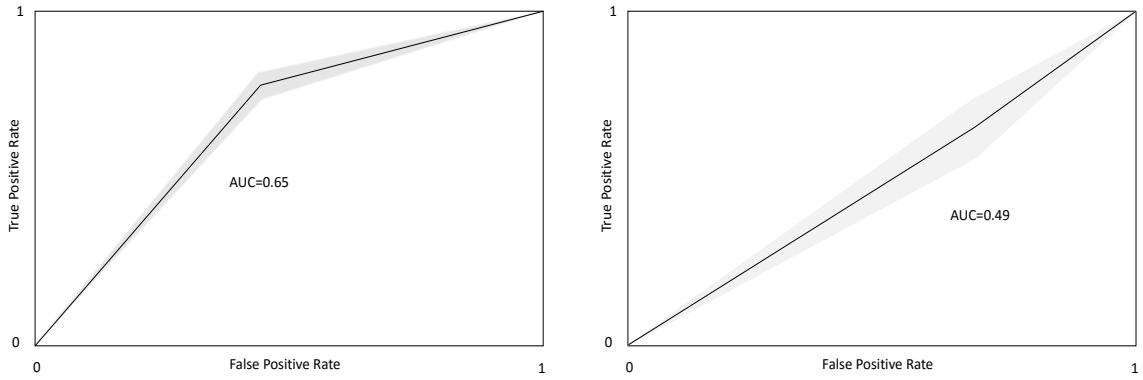
| Variables | Odd-ratio (95% Confidence Interval) | p-value |
|---|---|---|
| (Intercept) | 0.000 (0.000 - 0.000) | 0.906 |
| ICU Type Medical | 1.346 (1.098 - 1.650) | 0.001 |
| ICU Type Trauma | 1.428 (1.051 - 1.934) | 0.014 |
| Age | 1.011 (1.008 - 1.014) | $< 0.001$ |
| ICU Length of Stay | 1.000 (1.000 - 1.001) | $< 0.001$ |
| Initial SOFA Liver | 0.887 (0.778 - 1.012) | 0.027 |
| Initial SOFA Respiratory | 0.837 (0.764 - 0.917) | $< 0.001$ |
| Discharge SOFA Liver | 1.245 (1.104 - 1.407) | 0.001 |
| Discharge SOFA Neurological | 1.463 (1.350 - 1.585) | $< 0.001$ |
| Maximum SOFA Liver | 1.224 (1.049 - 1.426) | 0.002 |
| Max SOFA Neurological | 0.836 (0.790 - 0.885) | $< 0.001$ |
| Max SOFA Renal | 1.504 (1.278 - 1.769) | $< 0.001$ |
| Average SOFA Coagulation | 1.073 (0.998 - 1.153) | 0.025 |
| Average SOFA Neurological | 0.766 (0.684 - 0.858) | $< 0.001$ |
| Average SOFA Renal | 0.769 (0.652 - 0.909) | 0.002 |
| Average SOFA Respiratory | 1.473 (1.238 - 1.756) | $< 0.001$ |
| Average SOFA Cardiovascular | 1.146 (1.038 - 1.265) | 0.012 |
| Variance SOFA Renal | 0.673 (0.465 - 0.963) | 0.020 |
| Central Venous Catheter | 0.694 (0.618 - 0.779) | $< 0.001$ |
| Patient Type Surgical | 1.628 (1.455 - 1.824) | $< 0.001$ |
| LOS before ICU admission | 1.001 (1.001 - 1.002) | $< 0.001$ |
| Number of previous ICU admissions | 1.217 (1.101 - 1.342) | $< 0.001$ |
| SIRS$>$ 2 48 hours prior to discharge | 1.322 (1.195 - 1.463) | $< 0.001$ |
| Mechanical Ventilation within 24 hours of discharge | 0.544 (0.473 - 0.625) | $< 0.001$ |
| Unit Utilization upon discharge (%) | 0.207 (0.154 - 0.279) | $< 0.001$ |




(a) Model including ICU non-survivors (b) Model excluding ICU non-survivors

Figure 3.4: ROC curves for models of predicting ICU readmission using logistic regression with 95% uncertainty bands

Logistic regression significantly outperforms the Naive-Bayes approach in predicting ICU readmission: the mean AUC drops to 0.65 for the model including ICU survivors, and 0.49 for the model excluding ICU survivors. Figure 3.5 depicts ROC-AUC curves for both Naive-Bayes models. It has been shown via several data sets that logistic regression generally outperforms the Naive-Bayes approach when training data is abundant, and vice versa when training data is scarce [116]. Our results also confirm previous observations.




(a) Model including ICU non-survivors (b) Model excluding ICU non-survivors

Figure 3.5: ROC curves for models of predicting ICU readmission using Naive-Bayes technique with 95% uncertainty bands

Table 3.7 compares the AUCs of existing readmission models. The best previously reported AUC is 0.74 [121]. We perform external validation of the model presented in [121] with our data set, which yields an AUC of 0.54.

Table 3.7: Comparison with published readmission models

| **Study** | **Reference** | **Year** | **Cohort Size** | **ROC-AUC** |
|---|---|---|---|---|
| Campbell et al. | [18] | 2008 | 6,208 | 0.65 |
| Frost et al. | [57] | 2010 | 14,952 | 0.66 |
| Badawi and Breslow | [7] | 2012 | 469,976 | 0.71 |
| Ouanes et al. | [121] | 2012 | 3,462 | 0.74 |
| Readmission model including ICU non-survivors | - | - | 16,059 | 0.79 |
| Readmission model excluding ICU non-survivors | - | - | 14,604 | 0.77 |

### 3.3.2 Intensive Care Unit Discharge Mortality Models

ICU survivors in the study population has a significantly higher mortality (20.9%) compared to the source population of 46,169, where the 1-year mortality is 15.2% ($p< 0.001$), reflecting some bias towards more complete data in sicker patients. Table 3.8 presents 26 variables that independently predicts post-ICU mortality (N=14,604, of whom 1658 died in hospital). The model is well calibrated, with a Hosmer-Lemeshow p-value of 0.63, and has very good discrimination with AUC of 0.88 [0.879,0.881]. A similar model based on a population that includes ICU non-survivors (N=16,059) exhibits slightly better performance with AUC of 0.925 [0.924,0.926] and good calibration p=0.37. This shows that it is easier to predict mortality in the ICU than post ICU mortality. Table 3.9 presents the list of predictors. Figure 3.6 presents ROC-AUC curves for both logistic regression mortality models.

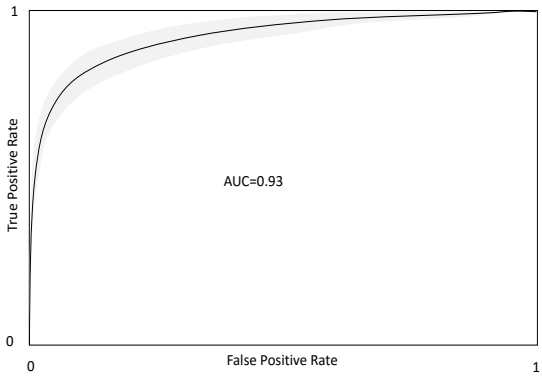
Similarly, logistic regression outperforms the Naive-Bayes approach in predicting mortality similar to readmission. Mean AUC drops from 0.93 to 0.78 and 0.88 to 0.71, for data set including ICU survivors and for data set excluding ICU survivors, respectively. Figure 3.7 presents ROC-AUC curves for both Naive-Bayes mortality models.

Table 3.8: Predictors of post-ICU logistic regression mortality model in a cohort excluding ICU non-survivors (N=14,604)

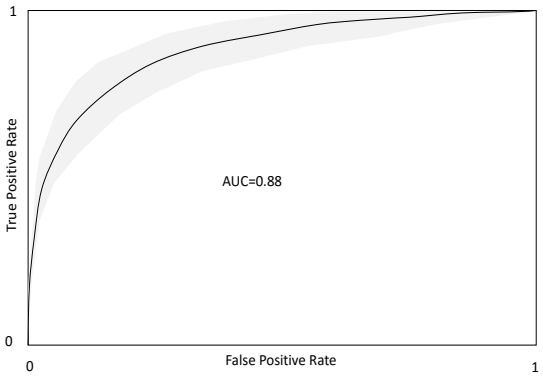
| Variables | Odd-ratio (95% Confidence Interval ) | p-value |
|---|---|---|
| (Intercept) | 0.002 (0.001 - 0.004) | $< 0.001$ |
| ICU Type Cardiac-T | 0.558 (0.414 - 0.749) | $< 0.001$ |
| ICU Type Medical | 1.338 (1.054 - 1.701) | 0.017 |
| ICU Type Neuro | 1.034 (1.030 - 1.039) | 0.001 |
| Age | 1.001 (1.000 - 1.001) | $< 0.001$ |
| LOS | 0.923 (0.866 - 0.983) | $< 0.001$ |
| Initial SOFA Neurological | 2.491 (1.903 - 3.289) | 0.012 |
| Discharge SOFA Liver | 1.245 (1.150 - 1.347) | $< 0.001$ |
| Discharge SOFA Coagulation | 2.080 (1.930 - 2.242) | $< 0.001$ |
| Discharge SOFA Neurological | 1.184 (1.114 - 1.258) | $< 0.001$ |
| Discharge SOFA Renal | 1.227 (1.062 - 1.419) | $< 0.001$ |
| Discharge SOFA Respiratory | 1.474 (1.353 - 1.606) | 0.006 |
| Discharge SOFA Cardiovascular | 0.887 (0.797 - 0.987) | $< 0.001$ |
| Maximum SOFA Neurological | 0.876 (0.773 - 0.993) | 0.028 |
| Maximum SOFA Respiratory | 0.608 (0.458 - 0.802) | 0.039 |
| Average SOFA Liver | 0.536 (0.338 - 0.834) | 0.001 |
| Variance SOFA Liver | 0.747 (0.648 - 0.858) | 0.007 |
| Variance SOFA Neurological | 2.389 (1.727 - 3.321) | $< 0.001$ |
| Variance SOFA Respiratory | 1.016 (1.010 - 1.021) | $< 0.001$ |
| APACHE III (Admission) | 0.848 (0.733 - 0.982) | $< 0.001$ |
| Central Venous Catheter | 0.796 (0.696 - 0.911) | 0.028 |
| Patient Type Surgical | 1.521 (1.325 - 1.745) | 0.001 |
| Discharge at night | 1.225 (1.062 - 1.412) | $< 0.001$ |
| Discharge during weekend | 1.002 (1.001 - 1.003) | 0.005 |
| LOS before ICU admission | 1.574 (1.382 - 1.792) | $< 0.001$ |
| SIRS> 2 48 hours prior to discharge | 0.380 (0.264 - 0.549) | $< 0.001$ |
| Unit utilization upon discharge (%) | 1.666 (1.466 - 1.865) | $< 0.001$ |

Table 3.9: Predictors of ICU logistic regression mortality model in a cohort including ICU non-survivors (N=16,059)

| Variables | Odd-ratio (95% Confidence Interval ) | p-value |
|---|---|---|
| (Intercept) | 0.008 (0.004 - 0.015) | $< 0.001$ |
| ICU Type Cardiac-T | 0.536 (0.402 - 0.711) | $< 0.001$ |
| ICU Type Medical | 1.332 (1.067 - 1.663) | 0.011 |
| ICU Type Neuro | 1.329 (1.081 - 1.636) | 0.007 |
| Age | 1.033 (1.029 - 1.038) | $< 0.001$ |
| Initial SOFA Neurological | 0.924 (0.873 - 0.979) | 0.007 |
| Discharge SOFA Liver | 2.231 (1.756 - 2.846) | $< 0.001$ |
| Discharge SOFA Coagulation | 1.278 (1.187 - 1.377) | $< 0.001$ |
| Discharge SOFA Neurological | 2.328 (2.182 - 2.486) | $< 0.001$ |
| Discharge SOFA Renal | 1.175 (1.108 - 1.246) | $< 0.001$ |
| Discharge SOFA Respiratory | 1.353 (1.166 - 1.573) | $< 0.001$ |
| Discharge SOFA Cardiovascular | 1.488 (1.374 - 1.613) | $< 0.001$ |
| Maximum SOFA Neurological | 0.836 (0.763 - 0.915) | $< 0.001$ |
| Average SOFA Liver | 0.657 (0.511 - 0.842) | $< 0.001$ |
| Average SOFA Respiratory | 0.780 (0.644 - 0.944) | 0.011 |
| Variance SOFA Liver | 0.585 (0.393 - 0.864) | 0.008 |
| Variance SOFA Neurological | 0.842 (0.743 - 0.954) | 0.007 |
| Variance SOFA Renal | 1.504 (1.162 - 1.943) | 0.002 |
| Variance SOFA Respiratory | 1.949 (1.474 - 2.583) | $< 0.001$ |
| APS (Admission) | 1.011 (1.006 - 1.016) | $< 0.001$ |
| Patient Type Surgical | 0.776 (0.686 - 0.878) | $< 0.001$ |
| Discharge at night | 1.549 (1.361 - 1.762) | $< 0.001$ |

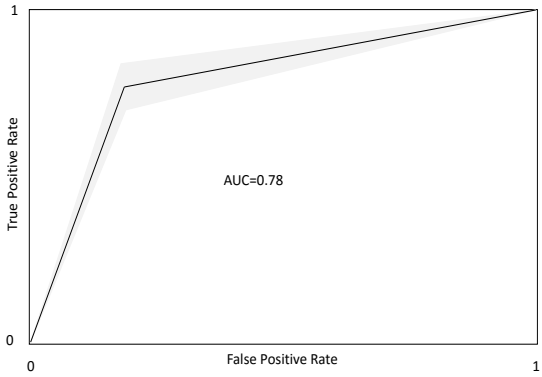



(a) Model including ICU non-survivors

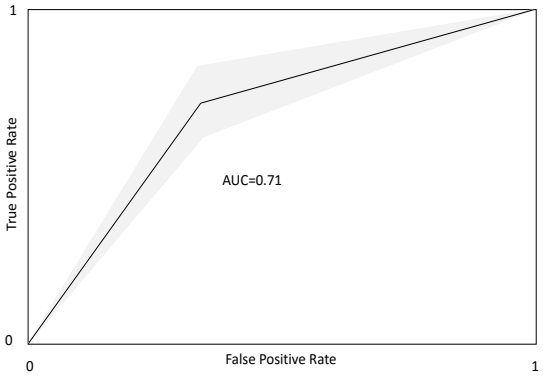



(b) Model excluding ICU non-survivors

Figure 3.6: ROC curves for models of predicting ICU mortality using logistic regression with 95% uncertainty bands




(a) Model including ICU non-survivors




(b) Model excluding ICU non-survivors

Figure 3.7: ROC curves for models of predicting ICU mortality using Naive-Bayes with 95% uncertainty bands

### 3.3.3 Incremental Predictive Value of SOFA Score Dynamics

To better understand the explanatory power associated with dynamic SOFA scores, we compare the AUCs and AICs of two models: (i) our best readmission model (N=16,509), and (ii) a model that includes only SOFA variables (SOFA only). There is a noticeable difference between AUCs and AICs of the two models, i.e., AUC= 0.67 (SOFA only) vs. 0.796 and AIC = 11,216 (SOFA only) vs. AIC = 13,025 (best model). We perform a similar analysis in the second ICU-readmission model (N=14,604) and the post-ICU mortality model. The AUC also drops to 0.66 (vs. 0.77) in the second ICU readmission model indicating that SOFA scores alone are not enough to predict ICU readmission. On the other hand, the ROC drops to only 0.85 (vs. 0.88) in the post-ICU mortality model suggesting that, as opposed to ICU-readmission models, SOFA scores dynamics by themselves explain post-discharge mortality well. We also examine the importance of score dynamics, compared to including discharge scores only. An ICU readmission model performs worse when only discharge scores are added to other risk factors with AUC of 0.77 (vs. 0.796). Hence models not including dynamic score information suffers a loss in discriminatory power. However, including only discharge SOFA score, in addition to other predictors, appears sufficient to predict post-ICU mortality with AUC of 0.874 (vs. 0.88 for dynamic SOFA information). We also verify how the ICU readmission probability changes in the 24 hours prior to actual ICU discharge (Figure 3.8). As expected, the probability decreases steadily, indicating continued patient improvement prior to discharge. Yet, the probability changes slowly, also suggesting that there should be some discretion as to the precise timing of ICU discharge.

## 3.4 DISCUSSION

Using an observational database of 12 ICUs from a large tertiary health care system, we develop and validate an ICU readmission model, as well as a post-ICU mortality model. These models offer significantly improved performance over existing readmission models and raise several interesting observations. Comparing models comprising SOFA scores and their

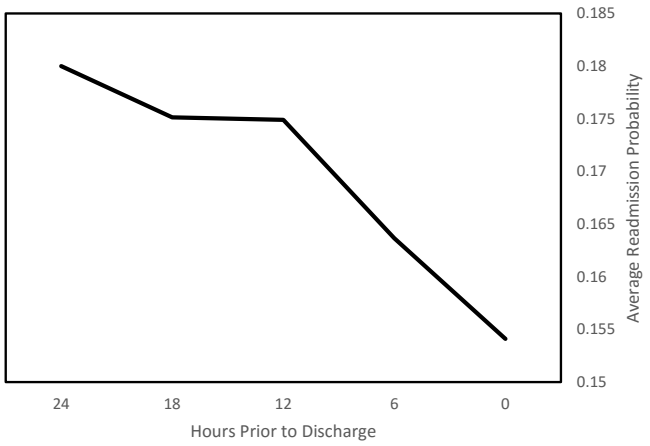



Figure 3.8: Average estimated probabilities of readmission for the entire cohort prior to discharge

trends to similar models not including these dynamic components, there is a significant improvement in model performance towards predicting ICU-readmission, but not post-ICU mortality. This finding suggests that SOFA evolution scores carry very significant explanatory power towards ICU-readmission, complementary to more conventional risk factors, and probably contribute to the incremental performance of the models we present, compared to published models. We also observe APS at the time of admission score is not a significant factor in the multivariate readmission model, unlike [18, 57, 121]. Thus, we presume that prognostic information carried by admission severity is preserved by dynamic changes in health during the ICU stay and ICU pre-discharge status. In fact, dynamic variables appear to carry additional information as suggested by improved model performance.

Others have previously reported that the type of ICU, a reflection of case-mix, was prognostic of post-discharge outcome [7] and our study extends these findings. Transplant patients have high readmission rates, as well as patients admitted to the general surgical ICU. Infection and poor organ function are frequent drivers of adverse evolution in transplant patients. Similar to [121], we show that a higher SIRS score, and persistent overt inflammation discharge increases the readmission rates. The use of central venous catheter has also ex-

planatory power in the readmission model consistent with the findings in [121]. A more detailed study of our general surgical population is required to elucidate the contributors to its increased risk beyond those already included in the models. These findings suggest that discharge decisions and acceptable readmission rates should depend on factors other than disease severity, such as case-mix. Beyond patient-based risk factors, we also note that ICU occupancy predicts both readmission and post-ICU mortality, while night and weekend discharges independently predict post-ICU mortality. Whether this is a statement on post-ICU care during weekends and nights merits further investigation; some reports [61, 121] conclude that discharging patients at night is an indicator of high risk of readmission, yet we could not verify this finding as our results are more consistent with those reported in [18].

The ICU-readmission model we present compares advantageously with existing models [7, 57, 18, 121] whereas a direct comparison with the Rothman index is not yet possible as its discrimination in predicting ICU-readmission is not yet available [127]. We believe that the inclusion of dynamic data at least partially explains improved performance. We also note that our readmission models estimate the probability of being readmitted to the ICU after discharge to a lower level of care during same hospital stay similar to [57, 81], whereas [121] estimates readmission probability within 7 days, and [18] within 2 days. Confining the time horizon of the prediction to short periods results is an easier task, explaining why these models have generally better AUCs among published models. Had we confined our event window for readmission to a shorter horizon, we anticipate we could also have obtained a more accurate model. It is appropriate to develop such a model if the purpose of prediction tool is to mitigate early readmissions.

For both the ICU readmission and late mortality models, we identify some potentially modifiable factors, including the presence of a central venous catheter, time of discharge, ICU occupancy at time of discharge, and pre-discharge SOFA scores, although it is quite plausible that the presence of a central catheter is a marker of unmeasured severity, and not a modifiable factor. A deeper analysis of the model, or restricting it to sub-populations at high risk of readmission or late mortality (e.g. transplant patients) could furnish further guidance as to the appropriate timing of discharge in these patients.

Our study has limitations. Many patients in our study population have missing data and we have evidence that data were not missing at random: we show that patients with complete data are generally sicked with higher 1-year mortality compared to overall ICU population. Thus, our models may not be as accurate in less sick patients, where event rates are lower. Although large, our population is derived from a single health system, and reflect a predominantly surgical case mix. Despite our extensive internal validation study, external validation of our models would therefore be desirable, especially in populations with a different case-mix. Our model formulates predictions based on variables measured prior to ICU discharge. Ideally, one would like to use such a model prospectively to support a disposition decision. The model is well-calibrated and therefore patients with high predicted probability of ICU readmission are indeed readmitted more often. We also find that the predicted probability of ICU readmission decreased monotonically in the days leading to recorded discharge. It can therefore be conjectured that, if patients had been discharged earlier (or later), one can compute a trade-off of an acceptable readmission rate. Operating points of such trade-off would plausibly be case-mix and environment dependent (e.g. high occupancy and high demand pressure). Flexible, yet evidence-based policies could therefore be implemented with knowledge of those trade-offs. For example, it might be better not to discharge patients at night or on week-ends if there is no demand for beds.

The most immediate impact of an accurate ICU readmission model is to use such a score as a selection mechanism as to which patients are readier for ICU discharge. Unfortunately, a head-to-head comparison of the performance of the model versus clinical judgment or alternative decision support tools, such as the Rothman index [53, 135], would require a specific evaluation. Perhaps the most attractive application is the relative of integration to modern electronic health record systems. Indeed, our model is comprised of variables that can be computed from variables automatically recorded in the EHR, excepted the neurological SOFA score. A second potential impact is the identification of potentially modifiable risk factors as described above. A third impact is that it is quite apparent that case-mix plays a major role in determining risk of ICU readmission and that policies of optimal discharge rules are therefore hardly transportable. Thus, our finding confirms a recent report that unadjusted ICU readmission rate is a poor quality indicator of system performance [156].

## 3.5 CONCLUSIONS

Optimizing ICU discharge decisions is paramount in providing cost-effective care in the increasingly outcome driven health care delivery context and promoting patient and family satisfaction. Unplanned readmissions are often associated with extreme anxiety and increased resource use, and appropriate targeting of those patients, beyond the actual timing of the discharge decision, can be helped by models such as the one we presented. We present tools that could help in such planning decision. We build logistic regression models and compare them with the published prediction models. We outperform any model in predicting readmission.

# 4.0 PHYSIOLOGY-BASED ANTICIPATIVE INTENSIVE CARE UNIT MANAGEMENT

## 4.1 INTRODUCTION

In this chapter, we build an infinite-horizon Markov Decision Process based on the scores to model the anticipative transfer request problem, with the objective of minimizing the expected total cost of unnecessarily occupying ICU beds plus the cost of unnecessarily reserving downstream beds. The model periodically determines the number of transfer requests after observing the medical conditions of the patients staying in the ICU. One major difference of our model from existing early discharge models, (e.g., Chan et al. [22] and Dobson et al. [42]), is that our model does not allow the transfer of patients to downstream units while they still need an ICU care, which more accurately reflects clinical reality. Early discharged patients have higher readmission and mortality rates [7].

Our aim is to determine the optimal number of downstream bed requests at each time step in order to minimize both the downstream bed allocation cost and unnecessary ICU occupation cost. Physicians monitor the status of each patient and determine the number of requests at each discharge round. Typically, discharge rounds are every 12 or 24 hours [77]. Our model allows state changes and discharges every 6 hours, whereas transfer decisions are taken every 12 hours. Note that if we allocate a downstream bed for each patient, then the multi-patient problem simply determines the transfer request times of each patient independently, i.e., solving many single-patient problems. Instead, we choose the number of requests so that the allocated beds in the downstream units can be used interchangeably, depending on the outcome of the patients in the next discharge round.

Bed costs are in a sense intangible and unknowable in advance, as they largely reflect the opportunity costs of beds, which may vary greatly. Hence, it is not easy to define and estimate these figures. While bed managers might have some estimates on these quantities, which are prone to high estimation errors. In Section 4.5 we subsequently consider robust dynamic programming approaches, which model ambiguity in the parameters defining a dynamic program.

Realistic formulations of the optimal transfer request problem are difficult to solve in practice even for small sized ICUs. We use an ADP approach to solve the problem. The approximation uses state-aggregation-based policy approximation [128], which partitions the state into solvable sets and then solves these smaller aggregated problems. Rather than generating an approximate value function via basis functions, we directly concatenate the policies of these small scale problems, i.e., summing the number of requests of each to determine a suboptimal policy for the original problem, which we call *an approximate policy* throughout the chapter.

The remainder of the chapter is organized as follows. In Section 4.2 we present the formulation of the model. In Section 4.3, we establish some structural properties. We discuss the approximation algorithm in Section 4.4. In Section 4.5 we present the cost ambiguity model formulation. In Section 4.6, we discuss some managerial insights through a set of numerical experiments.

## 4.2 MARKOV DECISION PROCESS FORMULATION

Transfer decisions require a clinically based measure of patients' medical conditions in the ICU. In particular, it is vital to have a measure that can predict the probability of readmission or death upon transfer to a downstream unit. While APACHE score and SOFA scores are the clinically accepted measures of disease severity in critical care patients. In Chapter 3 we show that these scores are less successful in predicting the readmission or death probabilities upon transfer to a downstream unit. Thus our objective is to develop a generalizable and dynamic transfer readiness score that can be used to provide better transfer decisions.

As we have discussed in Chapter 3, predictive ICU management decisions should be based on readmission or mortality, and thus a good estimation/prediction of readmission or mortality is essential. For that, we have developed logistic regression models that incorporates patient demographics, patient type (medical, surgical), medical scores (APACHE-III, SOFA, SIRS), Charlson index for comorbidities, procedures applied, mechanical ventilation status, unit information (current unit type, origin unit), and unit utilization. We employ the readmission logistic regression model presented in Chapter 3. To generate the score, we first partition the probability intervals and map them into point scores. Hence, the score of a patient is determined the probability of readmission, which is generated by logistic regression.

We define a score set $\mathcal{S} = \{1, \ldots, S\}$, where score 1 is the transfer state that has the probability of readmission below 0.1, and split the remaining interval $(0.1, 1]$ into equally spaced and disjoint $S-1$ intervals. Score $i$ maps into the $(i-1)^{\text{th}}$ interval. In other words, every score represents the probability of readmission being in the corresponding interval. For our numerical studies we define the range of the TRS from 1 to 5.

Now, we present an infinite-horizon discrete-time MDP model for the anticipative transfer request problem. We assume that ICU is a loss system for the MDP model; whenever the unit is full, arriving patients are diverted to other hospitals or admitted to other units.

### 4.2.1 States and Actions

The state is defined as the transfer readiness scores of all patients. We further define state $\Delta$ to represent the state that the corresponding bed is empty. We assume that if $k$ beds in downstream units are available and more than $k$ patients are ready to be transferred, the $k$ patients who move are selected in increasing order of bed numbers. Arriving patients' initial health states are drawn from an initial state distribution $Q$ that is estimated through the historical distribution of health scores upon admission to the ICU. Consider an $N$-bed ICU and denote $\mathcal{N} = \{1, \ldots, N\}$. Each patient's health in the ICU is monitored via their TRS process. We assume that the TRS process is a DTMC and each patient's health evolves independently. Let $\mathcal{S} = \{\Delta, 1, \ldots, S\}$ be the state set for each ICU patient, and $\mathbf{H}$ be the

transition probability matrix of the health evolution of a patient. Hence the state description $\mathbf{s} \in \mathscr{S}$ is an $N$-dimensional vector where $s_i \in \mathcal{S}$ is the state of bed $i$, and $\mathscr{S}$ is the state space. For a given state $\mathbf{s}$, each $a(\mathbf{s}) \in \mathcal{A} = \{0, 1, \ldots, N\}$ represents the number of transfer requests. Recall that $\Delta$ is the state where the bed is empty. A patient can only be transferred if her status is in a transferable state $\mathcal{S}_0 \subset \mathcal{S}$ at the beginning of next period. We also define $\mathcal{S}_0^c = \mathcal{S} \setminus \mathcal{S}_0$ to be non-transferable states.

### 4.2.2 Transition Probabilities

Let $Z$ be the number of patients arriving to the system at a period. Hence the probability distribution of $Z$ is given by

$$P(Z = z) = \frac{e^{-\eta}\eta^z}{z!}, \quad \text{for } z = 0, 1, \ldots$$

For computational tractability, we assume that new arriving patients are assigned to empty beds first and patients are selected to transfer in increasing order of the index set of beds of transferable patients. We assume that a requested downstream bed becomes available in the next round with probability $\alpha > 0$. Now, define set $\mathscr{E}(\mathbf{s})$ as the index set of empty beds for a given state $\mathbf{s}$. Let $P(\mathbf{s}'|\mathbf{s}, a)$ be the probability of moving from state $\mathbf{s}$ to state $\mathbf{s}'$ given action $a \in \mathcal{A}$. When $a = 0$, the probability of being in state $\mathbf{s}'$ is simply the multiplication of independent health evolution probabilities and new arrivals with initial distributions, i.e.,

$$P(\mathbf{s}'|\mathbf{s}, 0) = \prod_{i \in \mathscr{E}^c(\mathbf{s})} H(s_i'|s_i) + \begin{cases} P\left(Z = [|\mathscr{E}(\mathbf{s})| - |\mathscr{E}(\mathbf{s}')|]\right) \displaystyle\prod_{i \in \mathscr{E}(\mathbf{s}) \setminus \mathscr{E}(\mathbf{s}')} Q(s_i'), & \text{if } \mathscr{E}(\mathbf{s}') \neq \emptyset \\ P\left(Z \geq [|\mathscr{E}(\mathbf{s})|]\right) \displaystyle\prod_{i \in \mathscr{E}(\mathbf{s}) \setminus \mathscr{E}(\mathbf{s}')} Q(s_i'), & \text{if } \mathscr{E}(\mathbf{s}') = \emptyset. \end{cases} \tag{4.1}$$

When $a > 0$, the calculation of the transition probabilities is somewhat complicated. For a given $a > 0$ requests, in the next stage $k$ beds can be available where $k \leq a$. Let $K_a$ be a random variable for the number of beds available when $a$ beds are requested. Define $B_a(k)$ as the probability of having $k$ beds available when $a$ beds are requested. Note that $K_a$ is binomial distributed with success probability $\alpha > 0$ and $a$ trials. Then for each $a > 0$ and $k = 1, ..., a$, $B_a(k) = P(K_a = k) = \binom{a}{k}\alpha^k(1-\alpha)^{a-k}$. Now, let $\mathcal{I}(\mathbf{s}) \subset \mathcal{N}$ be the index set of non-transferable patients for a given $\mathbf{s}$, i.e., $\mathcal{I}(\mathbf{s}) = \{i \in \mathcal{N} : s_i \in \mathcal{S}_0^c\}$.

Recall that there are many different ways to approach state $\mathbf{s}'$ from state $\mathbf{s}$. First, we should note that some transitions are not feasible due to aforementioned technical assumption, e.g., if bed 1 and bed 2 are empty at $\mathbf{s}$, we cannot transition to a state $\mathbf{s}'$ where bed 1 is empty but bed 2 is occupied, since new patients are allocated in increasing order of the bed numbers. Hence the probability for such transitions are zero. Second, if a bed is empty in the next period, but occupied in the current period, then the patient must have been discharged, hence at least a bed must be available for such state transition to happen; if $k < |\mathscr{E}(\mathbf{s}') \setminus \mathscr{E}(\mathbf{s})|$, then the probability of transitioning from state $\mathbf{s}$ to $\mathbf{s}'$ is zero. Similarly, if a bed is occupied in the next period but empty in the current period, there must be at least an arriving patient, i.e., if $z < |\mathscr{E}(\mathbf{s}) \setminus \mathscr{E}(\mathbf{s}')|$, then the probability of transitioning from state $\mathbf{s}$ to $\mathbf{s}'$ is zero. We will therefore condition on the number of available beds k.

Consider the case, where $|\mathcal{I}(\mathbf{s}')|$ patients are non-transferable, and $N - |\mathcal{I}(\mathbf{s}')| > 0$ patients are transferable and $|\mathscr{E}(\mathbf{s}')| > 0$ beds are empty in the next period. Because there are transferable patients in the ICU in the next period, $k$ patients were transferred. Patients in the set $\mathscr{E}(\mathbf{s}') \setminus \mathscr{E}(\mathbf{s})$ are discharged, hence their states must have transitioned to a transferable state in the next period and transferred, which occurs with probability

$$\mathscr{P}_1 = \prod_{i \in \mathscr{E}(\mathbf{s}') \setminus \mathscr{E}(\mathbf{s})} \sum_{j \in \mathcal{S}_0} H(j|s_i).$$

As mentioned earlier, new patients in the set $\mathscr{E}(\mathbf{s}) = \setminus \mathscr{E}(\mathbf{s}')$ arrived with initial states to transition from $\mathbf{s}$ to $\mathbf{s}'$, which occurs with probability

$$\mathscr{P}_2 = \prod_{i \in \mathscr{E}(\mathbf{s}) \setminus \mathscr{E}(\mathbf{s}')} Q(s_i').$$

Further, $k - |\mathscr{E}(\mathbf{s}') \setminus \mathscr{E}(\mathbf{s})|$ patients that are in non-transferable states in the current period may have evolved to a transferable state and transferred, and new patients arrived with initial states. These patients can be any among in non-transferable but non-empty state in the next period and the current period. Thus, define a subset $\mathcal{T} \subset \mathcal{I}(\mathbf{s}') \setminus (\mathscr{E}(\mathbf{s}) \cup \mathscr{E}(\mathbf{s}'))$, with cardinality $|\mathcal{T}| = k - |\mathscr{E}(\mathbf{s}') \setminus \mathscr{E}(\mathbf{s})|$. Hence, the probability of such an event is

$$\mathscr{P}_3 = \sum_{\forall \mathcal{T}} \prod_{i \in \mathcal{T}} Q(s_i') \sum_{j \in \mathcal{S}_0} H(j|s_i).$$

All remaining patients should simply make the transitions through the score process, which has probability

$$\mathscr{P}_4 = \prod_{i \in \mathcal{T}^c \setminus (\mathscr{E}(\mathbf{s}) \cup \mathscr{E}(\mathbf{s}'))} H(s_i'|s_i).$$

Finally, $k - |\mathscr{E}(\mathbf{s}') \setminus \mathscr{E}(\mathbf{s})| + |\mathscr{E}(\mathbf{s})|$ must have arrived, which occurs with probability

$$\mathscr{P}_5 = P\big(Z = k - |\mathscr{E}(\mathbf{s}') \setminus \mathscr{E}(\mathbf{s})| + |\mathscr{E}(\mathbf{s}) \setminus \mathscr{E}(\mathbf{s}')|\big).$$

Hence the transition probability of going from $\mathbf{s}$ to $\mathbf{s}'$ (defined above) when $a$ beds are requested is:

$$P(\mathbf{s}'|\mathbf{s}, a) = \sum_{k=|\mathscr{E}(\mathbf{s}') \setminus \mathscr{E}(\mathbf{s})|}^{a} B_a(k) \prod_{\ell=1}^{5} \mathscr{P}_\ell.$$

The case where no patients are transferable ($\mathcal{I}(\mathbf{s}') = \mathcal{N}$) is slightly different since if $k > 0$ beds become available in the next stage, patients will stay in the ICU, because less patients may become transferable. Hence,

$$P(\mathbf{s}'|\mathbf{s}, a) = \prod_{i \in N \setminus (\mathscr{E}(\mathbf{s}) \cup \mathscr{E}(\mathbf{s}'))} H(s_i'|s_i) + \sum_{k=|\mathscr{E}(\mathbf{s}') \setminus \mathscr{E}(\mathbf{s})|}^{a} B_a(k) \prod_{\ell=1}^{5} \mathscr{P}_\ell.$$

For the case where the unit is full in the next period ($\mathscr{E}(\mathbf{s}') = 0$), the equality in $\mathscr{P}_5$ is replaced with greater than or equal to as new arrivals are lost since the ICU is full, i.e., the quantity $\mathscr{P}_5$ is modified to be equal to

$$\mathscr{P}_5 = P(Z \geq k - |\mathscr{E}(\mathbf{s}') \setminus \mathscr{E}(\mathbf{s})| + |\mathscr{E}(\mathbf{s}) \setminus \mathscr{E}(\mathbf{s}')|).$$

### 4.2.3 Cost Structure

We impose costs on allocated downstream beds that are not used, and on ICU stay for transferable patients. Specifically, we pay $c_B \geq 0$ per stage per downstream bed requested, but not allocated. Further, we pay $c_I \geq 0$ per stage for each transferable patient who stays in the ICU, since they will be allocated to the ICU unnecessarily. We should specifically note that these costs are not necessarily tangible bed operating costs, but rather represent the opportunity cost of a bed system-wide, hence we might have cases where downstream bed cost $c_B$ exceeds ICU bed cost $c_I$, if the downstream is the bottleneck. These figures are neither constant nor easy to estimate, thus we employ a robust model in Section 4.6.3.

Define $p_i$ to be the probability that patient $i$ is discharged in the next period given that she is in state $s_i$. It is easy to see that

$$p_i = \sum_{j \in \mathcal{S}_0} H(j|s_i).$$

Let $d(\mathbf{s})$ be the random number of transferable patients in the next period given state $\mathbf{s}$. The probability mass functions of $d(\mathbf{s})$ is given by

$$P(d(\mathbf{s}) = d) = \sum_{\mathcal{T} \subset \mathcal{N}, |\mathcal{T}| = d} \prod_{i \in \mathcal{T}} p_i \prod_{i \in \mathcal{T}^c} (1 - p_i).$$

Hence, the total cost can be written as

$$\begin{aligned} c(\mathbf{s}, a) &= c_I |\mathcal{I}^c(\mathbf{s})| + c_B E[(d(\mathbf{s}) - K_a)^-] \\ &= c_I |\mathcal{I}^c(\mathbf{s})| + \sum_{k=0}^{a} B_a(k) \sum_{d=0}^{k} c_B (k - d) P(d(\mathbf{s}) = d). \end{aligned}$$

### 4.2.4 Optimality Criterion

The objective is to minimize the total expected discounted extra ICU holding cost and downstream bed allocation cost. Let $\lambda \in (0, 1)$ be the discount rate, and $V(\mathbf{s})$ be the minimum expected discounted total cost starting from state $\mathbf{s}$. Recall that we allow patients to be transferred before the next decision round. To model such a case, define a dummy value function $\mathscr{V}(\mathbf{s})$ for the next round following a discharge round, where we take no actions but allow the state to change and patients to transfer.

### 4.2.5 Bellman Equations

The optimality equation for each $\mathbf{s} \in \mathscr{S}$ can be written as

$$V(\mathbf{s}) = \min_{a \in \{0,1,\ldots,N\}} \left\{ c(\mathbf{s}, a) + \lambda \sum_{\mathbf{s}' \in \mathscr{S}} \mathscr{V}(\mathbf{s}') P(\mathbf{s}'|s, a) \right\}, \tag{4.2}$$

where

$$\mathscr{V}(\mathbf{s}) = c(\mathbf{s}, 0) + \lambda \sum_{\mathbf{s}' \in \mathscr{S}} V(\mathbf{s}') P(\mathbf{s}'|s, 0). \tag{4.3}$$

Before presenting structural properties, we summarize our main assumptions as follows:

- The Transfer readiness process follows a discrete time Markov Chain.
- Patients arrive to the unit according to a Poisson distribution with mean $\eta$ in each period.
- Arriving patients' initial scores are drawn from a distribution $Q$.
- The ICU unit is a loss system, i.e., newly arriving patients are lost, if the ICU is full.
- Requested downstream beds become available according to a Bernoulli process. Note that we are not specifically modeling the downstream unit due to tractability.
- Patients are transferred first and arriving patients are admitted next at the beginning of each period.

## 4.3 STRUCTURAL PROPERTIES

In this section, we prove the existence of an optimal threshold policy for a special case of the problem. Threshold type of policies are easy to implement, appealing to decision makers and can improve the computational efficiency of dynamic programming algorithms [129]. We show the existence of an optimal threshold policy for a special case of the problem where there is a single patient in the ICU. In that case we define the threshold policy as follows: A transfer request should be made if and only if the patient's score is below a certain score.

### 4.3.1 Single-Patient Model

Let us start by describing the single-patient model. The state description $s$ is the score of the single patient, and the action set is $a(s) \in \{0, 1\}$. At each decision epoch the physician decides whether to make a transfer request. The transition probabilities are

$$P(s'|s,a) = \begin{cases} H(s'|s) & \text{if } a = 0, \\ (1-\alpha)H(s'|s) & \text{if } a = 1, \text{ and } s' \in \mathcal{S}_0, \\ H(s'|s) + \alpha Q(s') \sum_{i \in \mathcal{S}_0} H(i|s) & \text{if } a = 1, \text{ and } s' \in \mathcal{S}_0^c. \end{cases}$$

The cost function $c(s,a)$ is

$$c(s,a) = \begin{cases} 0 & \text{if } a = 0 \text{ and } s \in \mathcal{S}_0^c, \\ c_I & \text{if } a = 0 \text{ and } s \in \mathcal{S}_0, \\ c_B \alpha \left(1 - \sum_{i \in \mathcal{S}_0} H(i|s)\right) & \text{if } a = 1 \text{ and } s \in \mathcal{S}_0^c, \\ c_I + c_B \alpha \left(1 - \sum_{i \in \mathcal{S}_0} H(i|s)\right) & \text{if } a = 1 \text{ and } s \in \mathcal{S}_0. \end{cases}$$

Define $V(s)$ as the expected discounted total cost starting from state $s$. Hence the optimality equation for each $s \in \mathcal{S}$ can be written as

$$V(s) = \min_{a \in \{0,1\}} \left\{ c(s,a) + \lambda \sum_{s' \in \mathcal{S}} V(s') P(s'|s,a) \right\}. \tag{4.4}$$

We assume that there exists a threshold $\psi \in \mathcal{S}$ such that the patient is transferable if $s \leq \psi$, and non-transferable if $s > \psi$, i.e., $\mathcal{S}_0 = \{s \in \mathcal{S} : s \leq \psi\}$. Further, we assume that new arriving patients are in a non-transferable state, thus $\sum_{s > \psi} Q(s) = 1$. Moreover an ICU bed is more valuable than a downstream bed, i.e., $c_I > c_B$. We present some assumptions that are sufficient conditions for the existence of an optimal threshold policy.

### 4.3.2 Structural Assumptions

The first assumption is on the probability distribution matrix $P(s'|s,1)$.

**Assumption 4.1.** *For a given $1 \leq \psi \leq S$, $P(i|s,1) \geq P(i|j,1)$ for any $s \leq \psi$, $i \leq \psi$ and $j > \psi$, and $P(j|k,1) \geq P(j|s,1)$ for any $s \leq \psi$, $j > \psi$, and $k > \psi$.*

Figure 4.1 illustrates the structure of the matrix, which is divided into four parts from $\psi^{\text{th}}$ diagonal entry. Then, column-wise, all the entries on the upper left corner should be greater than or equal to all the entries from the lower left corner and all the entries on the upper right corner should be less than or equal to all the entries in the lower right corner.

$$\left[\begin{array}{ccc|ccc} p_{11} & \cdots & p_{1\psi} & p_{1\psi+1} & \cdots & p_{1S} \\ \vdots & & \vdots & \vdots & & \vdots \\ p_{\psi 1} & \cdots & p_{\psi\psi} & p_{\psi\psi+1} & \cdots & p_{\psi S} \\ \hline p_{\psi+11} & \cdots & p_{\psi+1\psi} & p_{\psi+1\psi+1} & \cdots & p_{\psi+1S} \\ \vdots & & \vdots & \vdots & & \vdots \\ p_{S1} & \cdots & p_{S\psi} & p_{S\psi+1} & \cdots & p_{SS} \end{array}\right]$$

Figure 4.1: Illustration of matrix structure in Assumption 4.1

Specifically, the assumption says that the likelihood of moving to a transferable state is higher when a patient is already in a transferable state rather than coming from a non-transferable state. Similarly, the likelihood of moving to non-transferable states is higher when the patient is in a non-transferable state rather than going from a transferable state.

**Remark 4.1.** *Recall Assumption 1 on the structure of the probability distribution matrix $P(s'|s,1)$. Typical probability matrix structures that have been extensively studied in the literature are: increasing failure rate (IFR) or decreasing failure rate (DFR) (cf., [9]), totally positive of order 2 (TP2) (cf., [82]) and upper-triangular. For example if a matrix is TP2, then it is IFR (see [82]).*

*We use a new matrix structure to get the monotonicity and threshold results, and this assumption is not stronger than any other in the literature. Now, for $\psi = 1$, the probability matrix $P$ satisfies Assumption 1.*

$$P = \begin{bmatrix} 0.7 & 0.1 & 0.2 \\ 0.3 & 0.3 & 0.4 \\ 0.1 & 0.6 & 0.3 \end{bmatrix}.$$

*On the other hand it is neither IFR nor DFR, hence not TP2. Also, clearly it is not upper triangular. This shows that Assumption 1 is no stronger than IFR, and hence no stronger than TP2.*

**Remark 4.2.** The conditions in Assumption 4.1 are equivalent to the following: $(1-\alpha)H(i|s) \geq H(i|j)$ for any $s \leq \psi$, $i \leq \psi$ and $j > \psi$. Further, $H(j|s) \geq H(j|k) + \alpha Q(j) \sum_{\ell \leq \psi} H(\ell|k)$ for any $s \leq \psi$, $j > \psi$, and $k > \psi$.

The second structural assumption is on the transition probability matrix $\mathbf{H}$.

**Assumption 4.2.** *$H[i|s]$ is a non-increasing function of $s$ for $i \leq \psi$.*

Intuitively, we assume that the probability of transitioning to a transferable state decreases as the severity level of the patient increases.

### 4.3.3 A Threshold-Type Transfer Request Policy

We present some structural properties that enable us to prove the existence of a threshold-type policy. We first prove that the cost of starting from a transferable state is always higher than starting from a non-transferable state (monotonicity), as summarized in Proposition 4.1.

**Proposition 4.1.** *Under Assumption 4.1, we have*

$$V(s) \geq V(s')$$

*for any $s \leq \psi$, and $s' > \psi$.*

*Proof.* Let $V_n(s)$ be the value function corresponding to the $n$th iteration of the value iteration algorithm. Similarly, let $W_n(s)$ and $R_n(s)$ denote the cost of waiting and requesting a bed at stage $n$, respectively. We prove the proposition by induction on $V_n(s) = \min\{W_n(s), R_n(s)\}$. Without loss of generality, assume that $V_0(s) = 0$ for all $s \in \mathcal{S}$. It is easy to observe that $R_1(s) \geq W_1(s) \geq R_1(s') \geq W_1(s')$, for any $s \leq \psi$, and $s' > \psi$, since $c_I > c_B$. Hence $V_1(s) \geq V_1(s')$ for any $s \leq \psi$, and $s' > \psi$. Assume that $V_{n-1}(s) \geq V_{n-1}(s')$, for any $s \leq \psi$, and $s' > \psi$. In order to show $V_n(s) \geq V_n(s')$, we need to show (i) $R_n(s) \geq W_n(s')$, (ii) $R_n(s) \geq R_n(s')$, (iii) $W_n(s) \geq W_n(s')$, and (iv) $W_n(s) \geq R_n(s')$, for any $s \leq \psi$, and $s' > \psi$. Now fix $s \leq \psi$, and $s' > \psi$. Also, let $x \in \underset{\{s:s\leq\psi\}}{\operatorname{argmin}} V_{n-1}(s)$, note that $V_{n-1}(x) \geq V_{n-1}(s')$, for $s' > \psi$.

We start by proving case (i). Now, we have

$$
\begin{aligned}
R_n(s) &= c(s,1) + \lambda \left[ \sum_{i\leq\psi} V_{n-1}(i)P(i|s,1) + \sum_{j>\psi} V_{n-1}(j)P(j|s,1) \right] \\
&\geq \lambda \left[ \sum_{i\leq\psi} V_{n-1}(i)P(i|s,1) + \sum_{j>\psi} V_{n-1}(j)P(j|s,1) \right] \\
&= \lambda \left[ \sum_{i\leq\psi} V_{n-1}(i)P(i|s',0) + \sum_{i\leq\psi} V_{n-1}(i)[P(i|s,1) - P(i|s',0)] + \sum_{j>\psi} V_{n-1}(j)P(j|s,1) \right] \\
&\geq \lambda \left[ \sum_{i\leq\psi} V_{n-1}(i)P(i|s',0) + \sum_{i\leq\psi} V_{n-1}(x)[P(i|s,1) - P(i|s',0)] + \sum_{j>\psi} V_{n-1}(j)P(j|s,1) \right] \qquad (4.5) \\
&= \lambda \left[ \sum_{i\leq\psi} V_{n-1}(i)P(i|s',0) + \sum_{j>\psi} V_{n-1}(x)[P(j|s',0) - P(j|s,1)] + \sum_{j>\psi} V_{n-1}(j)P(j|s,1) \right] \qquad (4.6) \\
&\geq \lambda \left[ \sum_{i\leq\psi} V_{n-1}(i)P(i|s',0) + \sum_{j>\psi} V_{n-1}(j)[P(j|s',0) - P(j|s,1)] + \sum_{j>\psi} V_{n-1}(j)P(j|s,1) \right] \qquad (4.7) \\
&= \lambda \left[ \sum_{i\leq\psi} V_{n-1}(i)P(i|s',0) + \sum_{j>\psi} V_{n-1}(j)P(j|s',0) \right] \\
&= W_n(s').
\end{aligned}
$$

Inequality (4.5) follows from definition of $x$, and the assumption that $(1-\alpha)H(i|s) \geq H(i|s')$ for $i \leq \psi$, hence, $[P(i|s,1) - P(i|s',0)] \geq 0$, for $i \leq \psi$. Equality (4.6) follows from the fact that both $P(\cdot|s,1)$ and $P(\cdot|s',0)$ are probability distributions. Thus,

$$
\sum_{i\leq\psi} P(i|s',0) + \sum_{j>\psi} P(j|s',0) = \sum_{i\leq\psi} P(i|s,1) + \sum_{j>\psi} P(j|s,1),
$$

which implies

$$\sum_{i\leq\psi}[P(i|s,1)-P(i|s',0)]=\sum_{j>\psi}[P(j|s',0)-P(j|s,1)].$$

Finally, inequality (4.7) follows from the assumption that $H(j|s')\geq H(j|s)+\alpha Q(j)\sum_{\ell\leq\psi}H(\ell|s)$, i.e., $[P(j|s',0)-P(j|s,1)]\geq 0$, and $V_{n-1}(x)\geq V_{n-1}(j)$ for $j>\psi$.

The same method can be repeated for other cases by making necessary arrangements in the probabilities. For case (ii), we need $P(i|s,1)\geq P(i|s',1)$, for $i\leq\psi$, $P(j|s',1)\geq P(j|s,1)$ for $j>\psi$, and $c(s,1)\geq c(s',1)$, then we can easily obtain $R_n(s)\geq R_n(s')$ by replacing $P(i|s',0)$'s and $P(j|s',0)$'s with $P(i|s',1)$'s and $P(j|s',1)$'s, respectively. Note that $c(s,1)\geq c(s',1)$ is trivial from $c(s,0)=c_I>c_B\geq c(s',1)$. From assumption 4.1, we have $P(i|s,1)=(1-\alpha)H(i|s)\geq H(i|s')\geq(1-\alpha)H(i|s')=P(i|s',1)$ for $i\leq\psi$, and $P(j|s,1)=H(j|s)+\alpha Q(j)\sum_{\ell\leq\psi}H(\ell|s)\leq H(j|s')\leq H(j|s')+\alpha Q(j)\sum_{\ell\leq\psi}H(\ell|s')=P(j|s',1)$ for $j>\psi$, thus $R_n(s)\geq R_n(s')$ for $s\leq\psi$ and $s'>\psi$. For case (iii), we have

$$\begin{aligned}
W_n(s)&=c(s,0)+\lambda\left[\sum_{i\leq\psi}V_{n-1}(i)P(i|s,0)+\sum_{j>\psi}V_{n-1}(j)P(j|s,0)\right]\\
&\geq\lambda\left[\sum_{i\leq\psi}V_{n-1}(i)P(i|s,0)+\sum_{j>\psi}V_{n-1}(j)P(j|s,0)\right]\\
&=\lambda\left[\sum_{i\leq\psi}V_{n-1}(i)P(i|s',0)+\sum_{i\leq\psi}V_{n-1}(i)[P(i|s,0)-P(i|s',0)]+\sum_{j>\psi}V_{n-1}(j)P(j|s,0)\right]\\
&\geq\lambda\left[\sum_{i\leq\psi}V_{n-1}(i)P(i|s',0)+\sum_{i\leq\psi}V_{n-1}(x)[P(i|s,0)-P(i|s',0)]+\sum_{j>\psi}V_{n-1}(j)P(j|s,0)\right] \qquad (4.8)\\
&=\lambda\left[\sum_{i\leq\psi}V_{n-1}(i)P(i|s',0)+\sum_{j>\psi}V_{n-1}(x)[P(j|s',0)-P(j|s,0)]+\sum_{j>\psi}V_{n-1}(j)P(j|s,0)\right] \qquad (4.9)\\
&\geq\lambda\left[\sum_{i\leq\psi}V_{n-1}(i)P(i|s',0)+\sum_{j>\psi}V_{n-1}(j)[P(j|s',0)-P(j|s,0)]+\sum_{j>\psi}V_{n-1}(j)P(j|s,0)\right] \qquad (4.10)\\
&=\lambda\left[\sum_{i\leq\psi}V_{n-1}(i)P(i|s',0)+\sum_{j>\psi}V_{n-1}(j)P(j|s',0)\right]\\
&=W_n(s').
\end{aligned}$$

Inequality (4.8) follows from $P(i|s,0) \geq P(i|s',0)$, which comes from Assumption 4.1, i.e., $P(i|s,0) = H(i|s) \geq (1-\alpha)H(i|s) \geq H(i|s') = P(i|s',0)$ for $i \leq \psi$, and the definition of $V_{n-1}(x)$. Inequality (4.10) follows from $P(j|s',0) \geq P(j|s,0)$ for $j > \psi$, which also comes from Assumption 4.1 that $P(j|s,0) = H(j|s) \leq H(j|s) + \alpha Q(j) \sum_{\ell \leq \psi} H(\ell|s) \leq H(j|s') = P(j|s',0)$ for $j > \psi$, and $V_{n-1}(x) \geq V_{n-1}(j)$ for $j > \psi$. Note that $c(s,1) \geq c(s',0)$ is trivial from $c_I > 0$.

For case (iv), we only need $P(i|s,0) \geq P(i|s',1)$, for $i \leq \psi$, and $P(j|s',1) \geq P(j|s,0)$ for all $j > \psi$, and further $c(s,0) \geq c(s',1)$. We can repeat the similar algebra above by replacing $P(i|s',0)$'s and $P(j|s',0)$'s with $P(i|s',1)$'s and $P(j|s',1)$'s. Note that $c(s,0) = c_I > c_B \geq c(s',1)$. From Assumption 4.1, we have $P(i|s,0) = H(i|s) \geq (1-\alpha)H(i|s) \geq H(i|s') \geq (1-\alpha)H(i|s') = P(i|s',1)$ for $i \leq \psi$, and $P(j|s,0) = H(j|s) \leq H(j|s) + \alpha Q(j) \sum_{\ell \leq \psi} H(\ell|s) \leq H(j|s') \leq H(j|s') + \alpha Q(j) \sum_{\ell \leq \psi} H(\ell|s') = P(j|s',1)$ for $j > \psi$. This completes the proof of $V_n(s) \geq V_n(s')$, for any $s \leq \psi$, and $s' > \psi$, and hence the result follows by induction. □

Let $L(s)$ be the difference between the cost of requesting a bed and the cost of waiting for a given state $s \in \mathcal{S}$, i.e.,

$$L(s) = c(s,1) - c(s,0) + \lambda \left[ \sum_{i \in \mathcal{S}} V(i) P(i|s,1) - \sum_{i \in \mathcal{S}} V(i) P(i|s,0) \right].$$

Proposition 4.2 presents the monotonicity of the cost difference function $L(s)$.

**Proposition 4.2.** *If Assumptions 4.1 and 4.2 hold, $L(s)$ is non-decreasing in $s$.*

*Proof.* Define $\theta(s) = \sum_{i \in \mathcal{S}} V(i) P(i|s,1) - \sum_{i \in \mathcal{S}} V(i) P(i|s,0)$, and rewrite $L(s) = c(s,1) - c(s,0) + \lambda\theta(s)$. We first show that $c(s,1) - c(s,0)$ is non-decreasing in $s$, and then $\theta(s)$ is non-decreasing. One can easily see that,

$$c(s,1) - c(s,0) = c_B \alpha \left( 1 - \sum_{i \leq \psi} H(i|s) \right), \tag{4.11}$$

for any $s \in \mathcal{S}$. From Assumption 4.2, it is easy to observe that $c(s,1) - c(s,0)$ is non-decreasing in $s$. Now, for $\theta(s)$, we have

$$\theta(s) = \sum_{i \in \mathcal{S}} V(i) P(i|s,1) - \sum_{i \in \mathcal{S}} V(i) P(i|s,0)$$

$$
\begin{aligned}
&= \sum_{i\in\mathcal{S}} V(i)[P(i|s,1) - H(i|s)] \\
&= \sum_{i\leq\psi} V(i)[(1-\alpha)H(i|s) - H(i|s)] + \sum_{j>\psi} V(j)[H(j|s) + \alpha Q(j)\sum_{\ell\leq\psi} H(\ell|s) - H(j|s)] \\
&= \alpha \sum_{i\leq\psi} -V(i)H(i|s) + \sum_{j>\psi} V(j)[\alpha Q(j)\sum_{\ell\leq\psi} H(\ell|s)] \\
&= \alpha \sum_{i\leq\psi} -V(i)H(i|s) + \alpha \sum_{\ell\leq\psi} H(\ell|s)\sum_{j>\psi} V(j)Q(j) \\
&= \alpha \sum_{i\leq\psi} H(i|s)[-V(i) + \sum_{j>\psi} V(j)Q(j)].
\end{aligned}
$$

Because $V(i) \geq V(j)$ for any $i \leq \psi$, and $j > \psi$, then $V(i) \geq \sum_{j>\psi} V(j)Q(j)$, because the right hand-side of the inequality is a convex combination. Thus $[-V(i) + \sum_{j>\psi} V(j)Q(j)]$ is non-positive. Then $\theta(s)$ is non-decreasing in $s$, since $H(i|s)$ is non-increasing in $s$. This completes the proof of $L(s)$ being non-decreasing in $s$. □

Our main result is given in Theorem 4.1, which presents the existence of a threshold type optimal policy for single-patient problem.

**Theorem 4.1.** *If Assumptions 4.1 and 4.2 hold, there exists a threshold $s^*$, where it is optimal to request a bed for all $s \leq s^*$, and wait for all $s > s^*$.*

**Proof of Theorem 4.1:** The result follows directly from Proposition 4.2 as follows. The fact that the cost difference is non-decreasing in $s$ indicates that whenever the cost of transfer is less than the cost of wait for some $s^*$, i.e., $L(s) \leq 0$, it will always be smaller than zero for $s \leq s^*$. Hence the cost of requesting will be always less than the cost of waiting for all $s \leq s^*$. Therefore, there exists a threshold type of optimal policy. □

For the multi-patient problem, we expect that a decrease in the severity of one patient with others being the same means more patients are likely to be in a transferable state in the next period, and, hence leads to a higher number of transfer requests, i.e., $a(\mathbf{s}) \geq a(\mathbf{s}')$, such that $s_j \geq s'_j$ for some $j \in \mathcal{N}$ and $s_i = s'_i$ for all $i \in \mathcal{N} \setminus j$. Conjecture 4.1 summarizes a lexicographic ordering of the states with regards to number of transfer requests.

**Conjecture 4.1.** *If $N > 1$ and Assumptions 4.1 and 4.2 hold, then the optimal number of requests in a state is less than or equal to the one if instead we have one patient with lower score., i.e., $a^*(\mathbf{s}) \geq a^*(\mathbf{s}')$, such that $s_j \geq s'_j$ for some $j \in \mathcal{N}$ and $s_i = s'_i$ for all $i \in \mathcal{N} \setminus j$.*

Although this conjecture remains open to be proven, we run an experimental analysis, where we randomly generate 100 5-dimensional problem instances. We solve all problem instances to optimality by employing policy iteration algorithm. We observe that in all the instances an optimal threshold policy exists. Furthermore, in all the numerical experiments, where we employ our data set, we observe a similar structure although Assumptions 4.1 and 4.2 do not hold.

**Remark 4.3.** *Note that Assumptions 4.1 and 4.2 may not hold in practice. We define the following metric for the average violation of Assumption 4.1, $\kappa_P$, as the sum of all entries violating the order of the inequalities given in Assumption 4.1 divided by the number of entries, i.e.,*

$$\kappa_P = \frac{1}{|\mathscr{S}|^2}\left[\sum_{i\leq\psi}\sum_{s\leq\psi}\sum_{j>\psi}\max\{P(i|j,1)-P(i|s,1),0\}+\sum_{k>\psi}\sum_{s\leq\psi}\sum_{j>\psi}\max\{P(j|s,1)-P(j|k,1),0\}\right].$$

*For the transition probability matrix generated through our data set, violation $\kappa_P$ is 0.032. However, in all the instances in Section 4.6 we observe that threshold policy holds when solved to optimality even when the transition probability matrix generated violates Assumption 4.1.*

## 4.4 STATE-AGGREGATION BASED POLICY APPROXIMATION ALGORITHM

The exact solution of the model is intractable for medium size ICUs (10-25 beds), as the size of the state space for a 10-bed ICU with five severity levels is approximately $10^7$. Instead, we employ an approximation heuristic that partitions the state into solvable sets, then solves the small scale problems, and finally sums the solutions of the small scale problems to determine an approximate solution to the original problem.

Recall that $N$ is the number of beds, hence the dimension of the state of the original problem, and define $w$ as the dimension of the state of a solvable problem in a reasonable time. Further define $\mathbf{s}^n$ as a state and $\mathscr{S}^n$ as the state space of an $n$-dimensional problem. Hence, $\mathbf{s}^N$ corresponds to the original state $\mathbf{s}$ and $\mathscr{S}^N$ corresponds to the original state

space $\mathscr{S}$. First, solve all small scale problems with dimensions less than or equal to $w$ to full optimality via policy iteration. Note that for every $n \leq w$-dimensional state $\mathbf{s}^n$, we have an optimal policy $a^*(\mathbf{s}^n)$. The second part is to determine an approximate optimal policy $\tilde{a}(\mathbf{s})$ for every state $\mathbf{s} = \{s_1, s_2, \ldots, s_N\} \in \mathscr{S}$, for the original problem. We partition the state $\mathbf{s}$ into disjoint solvable sets, i.e., $\mathbf{s} = \bigcup_i \mathbf{s}_i$, where each set has at most cardinality $w$, $|\mathbf{s}_i| \leq w$, in other words dimension of each $\mathbf{s}_i$ is less than or equal to $w$. For a given partition the approximate solution for the original problem is $\tilde{a}(\mathbf{s}) = \sum_i a^*(\mathbf{s}_i^{|\mathbf{s}_i|})$. Note that we only need to store optimal policies for small scale problems, and an approximate policy can be generated for every state for the original problem through summation. This solves the problem of storing the optimal policy. Algorithm 1 summarizes the approximation method.

**Algorithm 1:** Approximation algorithm

1 Determine exact optimal solutions $V^*(\mathbf{s}^n)$, for all $\mathbf{s}^n \in \mathscr{S}^n$, for $n \leq w$

2 Determine the optimal action $a^*(\mathbf{s}^n)$ for all $\mathbf{s}^n \in \mathscr{S}^n$, for $n \leq w$

3 For each $\mathbf{s} \in \mathscr{S}$

(I) Partition the state $\mathbf{s}$ into sets $\mathbf{s} = \bigcup_i \mathbf{s}_i$, of each having cardinality at most $w$

(II) Return approximate policy for state $\mathbf{s}$ as $\tilde{a}(\mathbf{s}) = \sum_i a^*(\mathbf{s}_i^{|\mathbf{s}_i|})$

We should also note that the approximate solution depends on the partitions, as the best partitioning is not obvious. We have employed two different methods: (1) sort states in ascending order and (2) random order, and picked the one with less cost in our algorithm. The random order approach generated approximately 10% less cost than the ascending order approach.

The approximation method can be considered as a state aggregation algorithm [11, 128]. In particular, the $N$-dimensional state space is aggregated to a $w$-dimensional state space. However, one of the main differences in our algorithm is that we do not approximate the value functions through basis functions. Instead, we simply generate an approximate policy by summing the policies of aggregated problems.

While our algorithm does not have a theoretical performance guarantee, we provide an experimental analysis to illustrate the performance of the algorithm in relatively small problem sizes. As we have stated, we generate 100 5-dimensional instances with random

transition probability matrices. We set the ICU bed cost $c_I = 2$, the downstream bed cost $c_B = 1$, the downstream bed availability probability $\alpha = 0.63$, and the initial state probability distribution $Q = [0, 0.2, 0.33, 0.32, 0.15]$. We determine approximate policies through our approximation algorithm by varying the solvable problem size $w$ from 1 to 4. Finally we compare the values of the approximate policies with the optimal value. Let $\pi^w$ be the approximate policy when the solvable problem size is set to $w$. Define $V^{\pi^w}(\mathbf{s})$ to be the value of policy $\pi^w$ for a given state $\mathbf{s}$. Further, define $\gamma^w$ as the mean optimality gap across all states, which is written as

$$\gamma^w = \frac{1}{|\mathscr{S}|} \sum_{\mathbf{s} \in \mathscr{S}} \left[ \frac{V^{\pi^w}(\mathbf{s})}{V^*(\mathbf{s})} - 1 \right].$$

Also let $\rho^w$ be the maximum optimality gap across all states, which is written as

$$\rho^w = \max_{\mathbf{s} \in \mathscr{S}} \left[ \frac{V^{\pi^w}(\mathbf{s})}{V^*(\mathbf{s})} - 1 \right].$$

Table 4.1 presents average mean optimality gaps $\gamma^w$ and average maximum optimality gaps $\rho^w$ among all 100 instances for $w \in \{1, 2, 3, 4\}$ in percentages.

Table 4.1: Optimality bounds for 100 random problem instances

| | $w$ | | | |
|---|---|---|---|---|
| | 1 | 2 | 3 | 4 |
| $\gamma^w$ | $50.62\% \pm 5.48\%$[1] | $12.03\% \pm 1.28\%$ | $3.63\% \pm 0.49\%$ | $3.01\% \pm 0.19\%$ |
| $\rho^w$ | $51.17\% \pm 5.48\%$ | $12.37\% \pm 1.28\%$ | $3.95\% \pm 0.49\%$ | $3.14\% \pm 0.19\%$ |

[1] 99% confidence intervals of sample averages

Note that the results of the experimental analysis are promising since the optimality gap is around 50%, when we solve 1-dimensional problems to determine an approximate policy instead of solving the problem to optimality. Furthermore, this gap reduces even up to 3% when we solve 4-dimensional problems. Notably, the optimality gap increases as states are further aggregated.

## 4.5 COST AMBIGUITY MODEL FORMULATION

Hospitals can be seen as a network of ICUs and downstream units, where downstream units can receive patients from multiple ICUs. Reducing the congestion of one ICU by making earlier requests might increase the congestion in the downstream unit, and hence affect other ICUs sharing the downstream unit. Due to inter-dependencies among units, a holistic approach needs to be pursued to decrease the congestion of the whole system. Although a model considering multiple ICUs and downstream units is not tractable, we can model this problem by adjusting the downstream bed cost to consider the congestion added to other ICUs sharing the downstream unit. Unfortunately, this complicates the estimation of the cost. This motivates us to build a robust model where bed cost parameters are ambiguous.

We adopt the max-min formulation scheme presented in Iyengar [78] and Nilim and El-Ghaoui [117], which mimics playing a game with nature. We assume that nature selects the worst possible probability distribution from the uncertainty set in each state and time after observing the controlled action.

We assume that $c_B$ follows some unknown distribution $\mu$ in an uncertainty set $\mathcal{C}$. Further define $E^{\mu}$ as the expectation with respect to the fixed measure $\mu \in \mathcal{C}$. We assume that $c_B$ is bounded. Hence the robust value function $\tilde{V}(\mathbf{s})$ for each $\mathbf{s} \in \mathscr{S}$, and $\tilde{\mathscr{V}}(\mathbf{s})$ the dummy robust value function to allow transfer one period prior to next decision round, given by

$$\tilde{V}(\mathbf{s}) = \min_{a \in \{0,1,\ldots,N\}} \sup_{\mu \in \mathcal{C}} E^{\mu} \left[ c^{\mu}(\mathbf{s}, a) + \lambda \sum_{\mathbf{s}' \in \mathscr{S}} \tilde{\mathscr{V}}(\mathbf{s}') P(\mathbf{s}'|\mathbf{s}, a) \right], \tag{4.12}$$

where

$$\tilde{\mathscr{V}}(\mathbf{s}) = \sup_{\mu \in \mathcal{C}} E^{\mu} \left[ c^{\mu}(\mathbf{s}, 0) + \lambda \sum_{\mathbf{s}' \in \mathscr{S}} \tilde{V}(\mathbf{s}') P(\mathbf{s}'|\mathbf{s}, 0) \right]. \tag{4.13}$$

We modify techniques provided in [117] to solve the above robust problem. We choose two different uncertainty set schemes in our study: (1) the interval model, and (2) the entropy model.

(i) *Interval model:* In the interval model we assume that $c_B$ can take any value between an upper bound $\overline{c}_B$ and a lower bound $\underline{c}_B$. Then the uncertainty set becomes

$$\mathcal{C} = \{\mu : P^{\mu}(c_B \leq \overline{c}_B) = 1, P^{\mu}(c_B \geq \underline{c}_B) = 1\}.$$

We assume that the decision maker has a point estimate on downstream bed cost, say $\widehat{c}_B$. Then the upper bound is $\overline{c}_B = \widehat{c}_B + \omega$, and the lower bound is $\overline{c}_B = \widehat{c}_B - \omega$. Hence the inner optimization problem is to determine the probability distribution of $c_B$, which maximizes the expectation in (4.12). We should note that the interval model is risk averse, and hence does not assume any further information other than $c_B$ being in an interval.

(ii) *Entropy model:* In the entropy model, we assume that the decision maker has an underlying probability distribution on $c_B$ and the uncertainty set is defined by an entropy model as described in [117]. Specifically, let $q$ be the empirical distribution on the parameter $c_B$. Then the description of the uncertainty set becomes

$$\mathcal{C} = \{\mu : \mu \geq 0, D(\mu||q) \leq \beta\},$$

where $\beta > 0$ is fixed, and $D(\mu||q)$ is the Kullback-Leibler divergence from $q$ to $\mu$,

$$D(\mu||q) = \sum_j \mu(j) \log \frac{\mu(j)}{q(j)}.$$

We use the bisection algorithm of [117] to solve the inner maximization problem, and the policy iteration algorithm for the outer minimization problem.

## 4.6 NUMERICAL EXPERIMENTS

In this section, we present numerical results to compare the benefits of the anticipative policy calculated via our approximation algorithm with two different policies: (1) current practice, in which a bed is requested when a patient is clinically transferable; and (2) expected transfer policy, in which the number of beds requested is simply the expected number of transferable patients in the next period that is rounded closest integer. We compare the policies in a simulation model described in Section 4.6.1. We present the performance of the non-robust anticipative policy in Section 4.6.2, and the robust policies in Section 4.6.3.

### 4.6.1 Intensive Care Unit Simulation Model

We assume that the patient arrival process is Poisson, and there is infinite waiting room capacity (e.g., [66] and [88]). We model patient transfers to other hospitals or ambulance diversions as abandonments from the waiting line and assume that patience is Bernoulli distributed. New arriving patients or patients waiting in line are admitted to the ICU whenever there is an empty bed. The initial health of each admitted patient follows the score process and is independent of other patients. We specifically model the downstream unit, which we assume has a finite capacity and there is no queue between ICU and downstream unit, thus a patient waiting for the downstream bed stays in the ICU. We assume that LOS in the downstream unit is exponential with mean 3 days. We also assume that downstream bed preparation times (includes cleaning, bed assignment, paperwork, etc.) are exponential distributed with rate 6 per day. The ICU LOS, i.e., the service process in the ICU, is dictated by the patient score process, the given transfer request policy, and the downstream bed capacity. Thus the service process is endogenously determined, which differentiates our simulation model from many other existing studies that assume a predetermined LOS distribution. Recall that the transfer policy provided by the optimization model is an input for the simulation model. We should note that the purpose of the simulation model is to compare different policies under a more realistic setting. Hence, we relaxed some assumptions of the optimization model in the simulation model. In particular:

- ICU has a waiting line, thus patients are not lost if the unit is full, but patients may abandon the queue.
- Downstream unit is specifically modeled, instead of assuming a Bernoulli bed availability process.

The length of each simulation is 6 years with a warm up period of 6 months. We perform 2,500 replications. Figure 4.2 illustrates main flow of the simulation model.

In our base scenario, we consider a medium-scale medical ICU with 12 beds. The ICU has 0.85 utilization, the mean LOS is 4.5 days and the mean transfer delay is 9.5 hours. We should note that transfer policy affects utilization, LOS and transfer delay, hence the preceding numbers are an outcome of the current transfer request policy. We estimate the

score transition probability matrix from all the patients who stayed in the unit. We assume that the ratio of the cost of ICU bed to the cost of downstream bed is 2 and the downstream unit has 10 beds in the base scenario. We set discount parameter $\lambda$ to 0.9999 in all the experiments, which corresponds to approximately a 1% monthly discount.

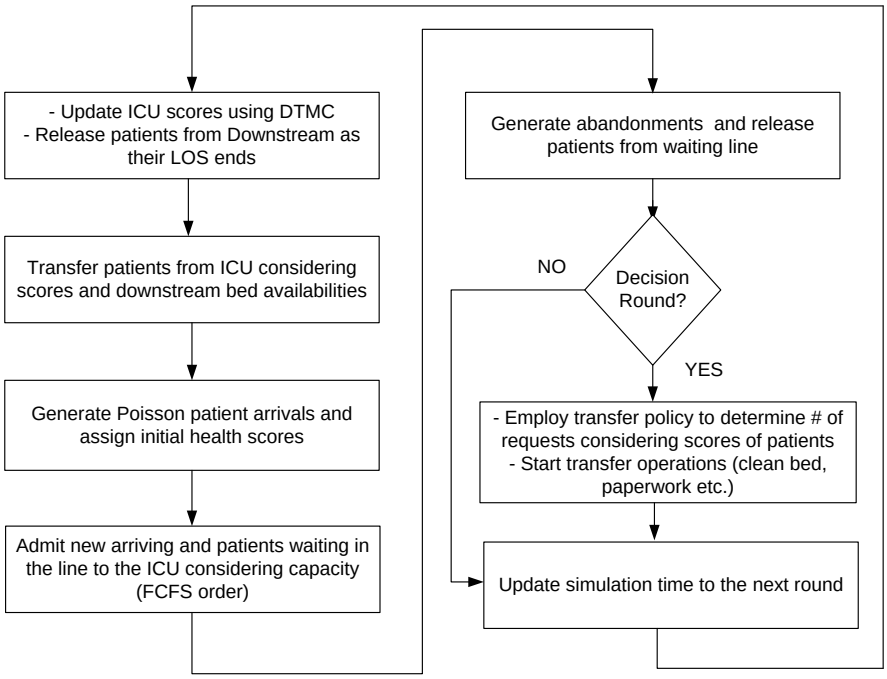



Figure 4.2: Illustration of the flow of the simulation model

### 4.6.2 Numerical Results for the Non-robust Anticipative Model

Table 4.2 shows that under all scenarios, the anticipative policy performs better in terms of costs, utilization, throughput, and transfer delay. In fact, a 11-bed unit under the anticipative policy performs as well as a 12-bed unit under current practice in terms of utilization. Thus, by simply changing the policy one might reduce congestion without increasing the unit size. We note the significant reduction in the mean transfer delay by almost 50%, which reduces the total LOS and thus helps to improve system efficiency. Table 4.2 also demonstrates that even a naive policy such as an expected request policy performs better than the current practice.

Table 4.2: Performance of policies in terms of the mean utilization, mean throughput and mean transfer delay with different unit sizes

| # of beds | Policy | Utilization | Throughput (per day) | Delay (hours) |
|---|---|---|---|---|
| 10 | Current Policy | $0.9144 \pm 0.0009$[1] | $2.0441 \pm 0.0020$ | $8.1020 \pm 0.0247$ |
| 10 | Expected Transfer policy | $0.8926 \pm 0.0010$ | $2.1191 \pm 0.0021$ | $4.8172 \pm 0.0283$ |
| 10 | Anticipative Policy | $0.8820 \pm 0.0010$ | $2.1378 \pm 0.0021$ | $3.7234 \pm 0.0293$ |
| 11 | Current Policy | $0.8831 \pm 0.0010$ | $2.1433 \pm 0.0021$ | $8.8298 \pm 0.0307$ |
| 11 | Expected Transfer Policy | $0.8557 \pm 0.0011$ | $2.2082 \pm 0.0023$ | $5.5009 \pm 0.0346$ |
| 11 | Anticipative Policy | $0.8444 \pm 0.0012$ | $2.2176 \pm 0.0023$ | $4.4954 \pm 0.0357$ |
| 12 | Current Policy | $0.8502 \pm 0.0011$ | $2.2178 \pm 0.0023$ | $9.6105 \pm 0.0384$ |
| 12 | Expected Transfer Policy | $0.8175 \pm 0.0012$ | $2.2684 \pm 0.0023$ | $6.2085 \pm 0.0425$ |
| 12 | Anticipative Policy | $0.8067 \pm 0.0013$ | $2.2779 \pm 0.0024$ | $5.3131 \pm 0.0443$ |

[1] 99% confidence intervals of sample means

We test the performance of the anticipative policy by changing the load of the ICU. Figure 4.3 depicts the change in the expected number of patients waiting in line to be admitted to the ICU as the arrival rate changes. Not surprisingly, the anticipative policy performs best in terms of the expected queue length. Furthermore, one can see from Figure 4.3 that the benefit of the anticipative policy increases as the ICU gets more loaded, which can be observed from the fact that the gap between lines gets wider.

The ratio of the ICU bed cost to the downstream bed cost, $c_I/c_B$, is the key to define the anticipative transfer policy, since it captures the relative importance of an ICU bed. We investigate the sensitivity of the benefit to the change in the cost ratio. We define the benefit of applying the anticipative policy as the percentage decrease in unit utilization, throughput, cost and transfer delay when the anticipative policy is applied instead of the current practice. Figure 4.4 shows the change of benefit when the cost ratio $c_I/c_B$ varies. We should note that the benefit of the anticipative policy compared to the current practice depends on the ratio of the costs of the ICU bed and the downstream bed. It is unsurprising that the anticipative policy provides the highest benefit when ICU beds are more expensive. The anticipative policy makes earlier requests, and transfers patients rapidly from the ICU. The benefit of the policy decreases as the cost of the downstream bed increases. The anticipative policy becomes more conservative in making earlier requests as the early allocation cost of downstream bed increases compared to the ICU unit bed cost, i.e., the anticipative policy

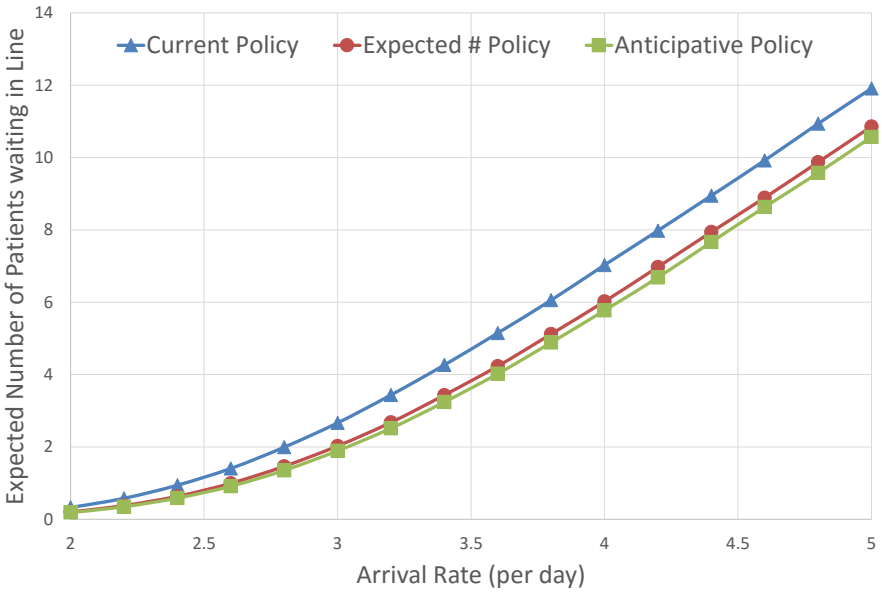



Figure 4.3: Expected number of patients waiting in line under different policies as the arrival rate changes

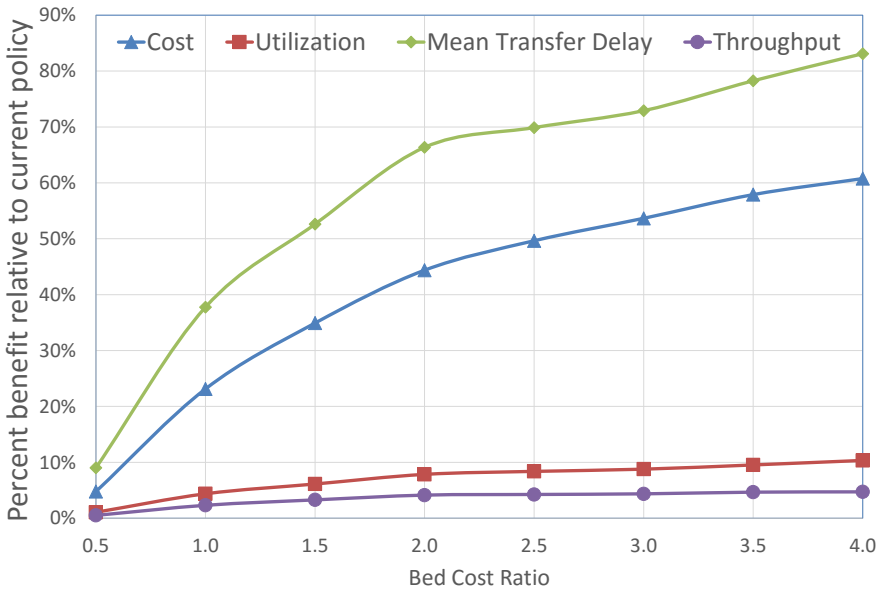



Figure 4.4: Benefit of applying the anticipative policy compared to the current practice as the ICU bed cost to the downstream bed cost ratio changes

tends to act the same as the current practice by delaying transfer requests. As mentioned earlier, the cost might stem from either high downstream costs or the lack of available beds from downstream beds, and much of the congestion is due to downstream units.

We also examine the change in the benefit of applying the anticipative policy as the downstream capacity changes, in other words, varying the downstream unit congestion. From Figure 4.5 we can see that the anticipative policy performs better as the downstream unit has more capacity. In particular, as the downstream unit gets more congested, patients have to wait longer in the ICU to find a bed in the downstream unit regardless of the transfer request policy.

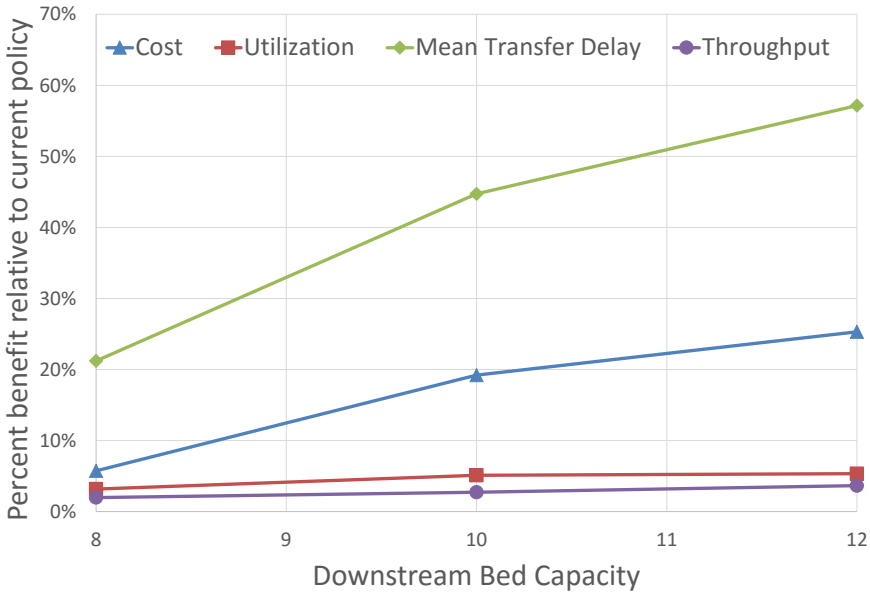



Figure 4.5: Benefit of applying the anticipative policy compared to the current practice as downstream capacity changes

### 4.6.3 Numerical Results for the Robust Methods

We provide numerical results for the two robust models: the interval model and the entropy model. We test the performance of the robust policies as well as the non-robust anticipative policy and compare them with the performance of current practice.

In the interval model, we choose the interval to be $[\widehat{c}_B - 0.5, \widehat{c}_B + 0.5]$. For the entropy model, we assume that the empirical distribution $q$ of $c_B$ is uniform, taking discrete values, with $P(c_B = \widehat{c}_B - 0.5) = 1/3, P(c_B = \widehat{c}_B) = 1/3, P(c_B = \widehat{c}_B + 0.5) = 1/3$. We choose $\beta$ to be 0.05, 0.25. Note that increasing values of $\beta$ corresponds to increasing levels of uncertainty.

We first determine optimal policies using our point estimates $\widehat{c}_B \in \{0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5\}$ and uncertainty sets as described, while fixing $c_I = 1$. Next, we simulate the performance of the policies under different realizations of downstream cost $c_B \in \{0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5\}$. Table 4.3 displays the benefit of applying various policies in terms of total cost relative to the current policy under all scenarios. We highlight the best performing policy by presenting the largest percentages in bold for each scenario. Also note that we choose not to present the percentage of the best performing alternative policy in bold for the scenarios on the lower diagonal of Table 4.3, because the current practice is the best (all percentages are negative).

From Table 4.3 we see that the benefits of the robust and the anticipative policies diminish as the downstream bed cost estimate $\widehat{c}_B$ increases. This is due to the fact that as the cost of downstream bed increases, the anticipative and the robust policies will delay making transfer requests. At the extreme, where the downstream bed cost is infinite and the ICU bed cost is finite, all proactive policies and current practice are the same.

We further make the following observations from the numerical experiments.

(i) The anticipative policy performs best when the downstream bed cost is accurately predicted, i.e., $c_B$ is close to $\widehat{c}_B$. This is due to the fact that the robust policies delay transfer requests by overestimating the downstream bed cost, and hence perform worse than the anticipative policy.

(ii) The anticipative policy is the best policy when the downstream bed cost is overestimated, i.e., $\widehat{c}_B > c_B$, which corresponds to the upper diagonal scenarios of Table 4.3. Note that the robust models assign even higher downstream bed costs than point estimate $\widehat{c}_B$, since they are designed to protect the downstream unit. The anticipative model uses an estimate of downstream bed cost closer to reality than the robust models under these scenarios. Thus, the anticipative policy performs better than the robust models.

Table 4.3: Expected discounted cost gain in percentages for best performing policies compared to the current policy

| Estimated downstream bed cost $\widehat{\mathbf{c}}_{\mathbf{B}}$ | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Policy | 0.5 | 1 | 1.5 | 2 | 2.5 | 3 | 3.5 | 4 | 4.5 | 5 |
| Actual downstream bed cost $\mathbf{c}_{\mathbf{B}}$ | 0.5 | Ant[1] | **19.21*** | **15.12** | **6.62** | **4.32** | **1.02** | **0.51** | **0.24** | **0.24** | 0.00 | 0.00 |
| | | RI[2] | 15.12 | 6.62 | 4.32 | 1.02 | 0.51 | 0.24 | 0.24 | 0.00 | 0.00 | 0.00 |
| | | RE005[3] | 17.30 | 5.27 | 5.63 | 1.83 | 0.73 | 0.51 | 0.22 | 0.21 | 0.00 | 0.00 |
| | | RE025[4] | 14.12 | 4.67 | 4.97 | 1.29 | 0.51 | 0.18 | 0.22 | 0.21 | 0.00 | 0.00 |
| | 1 | Ant | 4.95 | **5.31** | **4.23** | **3.21** | **2.67** | **1.21** | **0.53** | **0.22** | **0.21** | 0.00 |
| | | RI | 5.31 | 4.23 | 3.21 | 2.67 | 1.21 | 0.53 | 0.22 | 0.21 | 0.00 | 0.00 |
| | | RE005 | **10.85** | 5.27 | 4.01 | 1.83 | 0.73 | 0.51 | 0.22 | 0.21 | 0.00 | 0.00 |
| | | RE025 | 7.82 | 4.67 | 3.55 | 1.29 | 0.51 | 0.18 | 0.22 | 0.21 | 0.00 | 0.00 |
| | 1.5 | Ant | -4.90 | 0.45 | **3.33** | **2.03** | **1.58** | **1.14** | **0.47** | **0.21** | **0.21** | 0.00 |
| | | RI | 0.45 | 3.33 | 2.03 | 1.58 | 1.14 | 0.47 | 0.21 | 0.21 | 0.00 | 0.00 |
| | | RE005 | -1.40 | **4.95** | 3.20 | 1.95 | 1.24 | 0.66 | 0.47 | 0.21 | 0.00 | 0.00 |
| | | RE025 | **5.43** | 3.67 | 2.55 | 1.91 | 1.04 | 0.47 | 0.21 | 0.21 | 0.00 | 0.00 |
| | 2 | Ant | -10.09 | -3.20 | 2.10 | **3.14** | **0.99** | **0.44** | **0.01** | 0.00 | 0.00 | 0.00 |
| | | RI | -3.20 | 2.10 | 3.14 | 0.99 | 0.44 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 |
| | | RE005 | -5.12 | -0.54 | **3.78** | 2.52 | 0.81 | 0.23 | 0.00 | 0.00 | 0.00 | 0.00 |
| | | RE025 | -4.32 | **2.35** | 3.26 | 1.43 | 0.76 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 |
| | 2.5 | Ant | -20.11 | -14.63 | -7.54 | -2.32 | **0.44** | **0.12** | 0.00 | 0.00 | 0.00 | 0.00 |
| | | RI | -14.63 | -7.54 | -2.32 | 0.44 | 0.12 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| | | RE005 | -18.79 | -4.30 | -0.97 | **1.33** | 0.24 | 0.05 | 0.00 | 0.00 | 0.00 | 0.00 |
| | | RE025 | -15.75 | -7.12 | **1.06** | 0.97 | 0.13 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| | 3 | Ant | -33.29 | -29.31 | -14.42 | -3.98 | -1.12 | **0.11** | **0.11** | 0.00 | 0.00 | 0.00 |
| | | RI | -29.31 | -14.42 | -3.98 | -1.12 | **0.11** | 0.11 | 0.11 | 0.00 | 0.00 | 0.00 |
| | | RE005 | -31.24 | -19.23 | -6.73 | -2.99 | -0.34 | 0.11 | 0.11 | 0.00 | 0.00 | 0.00 |
| | | RE025 | -30.05 | -12.21 | -3.56 | -0.33 | 0.27 | 0.11 | 0.11 | 0.00 | 0.00 | 0.00 |
| | 3.5 | Ant | -44.21 | -36.39 | -17.73 | -4.78 | -1.45 | -0.98 | **0.09** | **0.03** | 0.00 | 0.00 |
| | | RI | -36.39 | -17.73 | -4.78 | -1.45 | -0.98 | 0.09 | 0.03 | 0.00 | 0.00 | 0.00 |
| | | RE005 | -40.87 | -25.49 | -5.24 | -2.34 | -1.23 | **0.24** | 0.06 | 0.00 | 0.00 | 0.00 |
| | | RE025 | -37.15 | -21.31 | -3.17 | -1.51 | -1.11 | 0.10 | 0.03 | 0.00 | 0.00 | 0.00 |
| | 4 | Ant | -52.31 | -43.28 | -20.96 | -5.56 | -1.77 | -1.05 | 0.07 | 0.09 | 0.00 | 0.00 |
| | | RI | -43.28 | -20.96 | -5.56 | -1.77 | -1.05 | **0.07** | 0.09 | 0.00 | 0.00 | 0.00 |
| | | RE005 | -48.75 | -27.43 | -12.37 | -2.55 | -1.54 | 0.05 | **0.10** | 0.00 | 0.00 | 0.00 |
| | | RE025 | -45.64 | -25.19 | -10.87 | -2.12 | -1.19 | 0.06 | 0.09 | 0.00 | 0.00 | 0.00 |
| | 4.5 | Ant | -65.12 | -50.01 | -24.12 | -6.33 | -2.09 | -1.12 | -0.05 | -0.03 | 0.00 | 0.00 |
| | | RI | -50.01 | -24.12 | -6.33 | -2.09 | -1.12 | -0.05 | -0.03 | 0.00 | 0.00 | 0.00 |
| | | RE005 | -61.32 | -29.89 | -14.23 | -3.07 | -1.33 | -1.01 | 0.00 | 0.00 | 0.00 | 0.00 |
| | | RE025 | -54.44 | -26.45 | -12.11 | -2.51 | -1.45 | -0.55 | 0.00 | 0.00 | 0.00 | 0.00 |
| | 5 | Ant | -74.22 | -56.58 | -27.20 | -7.07 | -2.39 | -1.18 | -0.02 | -0.02 | 0.00 | 0.00 |
| | | RI | -56.58 | -27.20 | -7.07 | -2.39 | -1.18 | -0.02 | -0.02 | 0.00 | 0.00 | 0.00 |
| | | RE005 | -65.55 | -34.31 | -23.21 | -4.20 | -1.45 | -0.78 | -0.02 | 0.00 | 0.00 | 0.00 |
| | | RE025 | -58.74 | -28.67 | -15.43 | -2.71 | -2.09 | -1.10 | -0.02 | 0.00 | 0.00 | 0.00 |

[1] Anticipative policy
[2] Robust policy with interval uncertainty set
[3] Robust policy with entropy uncertainty set where $\beta = 0.05$
[4] Robust policy with entropy uncertainty set where $\beta = 0.25$
* All numbers are in percentages

(iii) Current practice outperforms other policies when the downstream bed cost is greatly underestimated. Such scenarios are located at the lower diagonal of Table 4.3. These are the scenarios where the downstream unit is the bottleneck rather than the ICU unit. This is intuitive, since in such cases the optimal strategy should be to keep patients in the ICU and not take the risk of making early allocations. However, proactive policies cannot capture this as $c_B$ is underestimated.

(iv) Note that the range of the benefit for the robust policies are smaller than the range of the benefit for the non-robust anticipative policy as expected. For instance, the benefit of the anticipative policy varies between -74.79% and 19.21%, while it is between -56.58% and 15.12% for the RI policy.

(v) It is clear from Table 4.3 that the robust policies might perform better than the anticipative policy under some scenarios where the downstream bed cost is underestimated. Interestingly, the performance of the robust policies depends on the underestimation error, and the uncertainty level implied by the model. In particular, when the underestimation error is high, RI performs best, as it already implies more uncertainty than the entropy models, hence is better prepared for the worst scenarios. Similarly, we can also observe that the robust policy with entropy uncertainty set with $\beta = 0.25$ (RE025) performs better than the robust policy with entropy uncertainty set with $\beta = 0.05$ (RE005), under higher underestimation errors, due to the fact that higher $\beta$ means higher uncertainty level. Evidently, RE005 performs best when the underestimation error is relatively lower.

## 4.7 CONCLUSIONS

Understanding of day-to-day ICU operations and altering the existing policies and procedures can significantly improve patient flow as well as patient outcomes. In this chapter, we quantify the value of switching from a reactive transfer policy to a proactive transfer policy. To this end, we build an MDP model, and solve it via an approximation method. Our results indicate that a proactive strategy can be very effective, for example, a 11-bed unit under our proposed policy performs as well as a 12-bed unit under current practice. Effective

use of clinical markers to define patient status through LOS in the ICU is the key to our success. We also provide a proof of threshold-structured transfer request policy for single patient problem. We provide numerical support for the existence of threshold policy for multi-patient problem through an experimental analysis. In addition, we provide a robust model to evaluate the sensitivity of the anticipative policy to cost parameter estimation errors. The results show that proactive strategies perform better than current practice in most of the settings, with the exception of cases where the downstream unit cost is extremely underestimated.

Some patients (20% in our data set) die while staying in the ICU and were discharged directly from the ICU without being transferred to a downstream unit. We should note that our model does not consider these patients and focuses only those who are transferred to downstream unit. However, we can tackle the problem by adjusting the model in the following way. (1) Enlarge the state space to contain a death state $\Delta$, where the transfer score does not change through time whenever patient dies, i.e., the patient stays in the death state with probability one. (2) Extend the notion of downstream unit to contain the morgue unit. (3) Assume that a transfer request is necessary for dead patients to transfer them to the morgue, similar to patients in transferable state. Hence, same transfer request mechanism now applies for the patients who died in the ICU. Another limitation of our study is that, we assume that all physicians employed the current practice all the time. Some physicians might have had a sense of the situation in the downstream unit and acted proactively when giving transfer decisions. This will reduce the reported benefits of the anticipative policy. However, expert opinion suggests that acting proactively is rare. Finally, employing anticipative policies may also influence the staff behavior. In particular, they may start acting slower in carrying transfer operations, knowing the fact that the transfers are not imminent. But, we still believe that an early transfer request will enable staff to better plan their operations, although they tend to act slower, that will eventually improve the efficiency.

# 5.0 AN EXPLICIT STOCHASTIC MODEL OF INTENSIVE CARE UNIT LENGTH OF STAYS

## 5.1 INTRODUCTION

Better LOS modeling can greatly enhance the quality of care as well as operational efficiency. In particular, more accurate predictions help hospital managers to plan the number of beds and staff required, or enables physicians to identify possible prolonged stays and prepare a better treatment plan. Many statistical or classification models use Day 1 information to predict a point estimate or an LOS range. To the best of our knowledge, this is the first attempt to model ICU LOS explicitly through patient health progression. This model allows us to distinguish between medically indicated LOS, and congestion LOS.

First, we present statistical and data-mining models to identify potential variables influencing LOS. We should note that our aim is not to employ these models as prediction tools. We utilize them to give us a brief understanding on factors inducing prolonged LOS's, as it is more difficult to model longer LOS's via probabilistic models presented in this study. We build a linear regression model with a poor fit ($R^2 = 0.1$). We also build a logistic regression model to identify patients with prolonged LOS's. We show that our model presents a good discrimination. The discriminative power increases 4% as the threshold increases from 10 days to 30 days, i.e., it is easier to distinguish patients staying longer.

Second, we provide an explicit ICU LOS model by (1) considering the health progression of the patients via the score process, (2) modeling the transfer decision via a threshold policy, and (3) considering the bed availability and delay dynamics. We analyze the model by assuming two score processes (DTMC and CTMC), and conclude that the LOS process is a phase-type distribution under these assumptions. A non-negative random variable is

said to have a phase-type distribution if it represents the time to absorption of some Markov chain [115]. Moreover, we compare the performance of the stochastic model with the Coxian distribution. We show that our model slightly underestimates the LOS (12.5 % smaller mean value), although it matches the shape of the distribution. The Coxian distribution fits the best as the model is flexible to fit the LOS data but the phases do not have physical meaning. Furthermore, the stochastic model demonstrates a poorer performance in predicting the tail of the distribution compared to the Coxian. We discuss some drawbacks of our model and discuss future extensions that may improve the goodness of fit.

The remainder of the chapter is organized as follows. In Section 5.2, we present preliminary analysis on the LOS data and the transfer delay data. We discuss statistical models to identify factors affecting LOS in Section 5.3. We present a detailed description of the stochastic model and analysis in Section 5.4. In Section 5.5, we discuss performance of the model while comparing with the Coxian distribution.

## 5.2 PRELIMINARY ANALYSIS

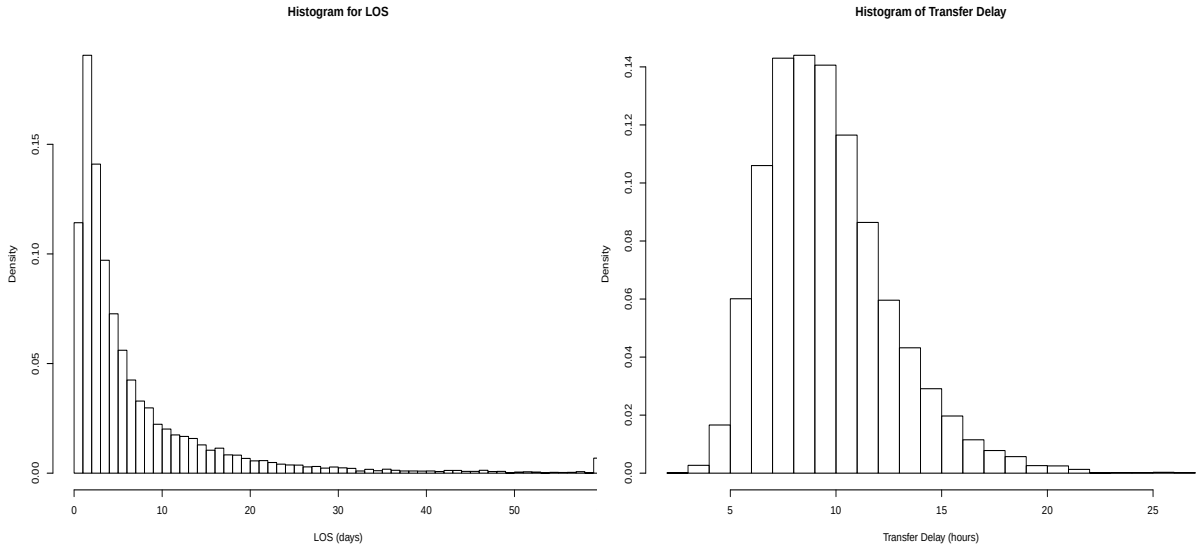
In this section, we present a preliminary statistical analysis. This will allow us to gain insights on the underlying structure of the LOS distribution. We use the same population discussed in Chapter 3 that consists of 16,059 ICU patients whose medical conditions can be tracked and transfer scores can be created from admission to discharge. Recall that our aim is to build an LOS model incorporating medical progression. LOS represents the number of days a patient stayed in one of the ICUs from admission to discharge. We choose not to remove patients who died in the ICU from analysis as deaths are not known at admission and dead patients must be considered in resource management. We define transfer delay as the duration between the time of transfer request and the time of discharge, in which transfer operations (e.g., assigning bed, cleaning bed, assigning staff) are performed.

Table 5.1 gives the basic descriptive statistics. The LOS distribution is positively skewed as the mean doubles the median, it has high variability with a coefficient of variation 1.55. The maximum LOS is around 234 days.

Table 5.1: LOS and transfer delay descriptive statistics

| | Mean | Median | Std. Dev. | Minimum | Maximum | Skewness |
|---|---|---|---|---|---|---|
| LOS (days) | 7.12 | 3.54 | 11.04 | 0.21 | 234.80 | 5.82 |
| Transfer Delay (hours) | 9.59 | 9.18 | 0.53 | 2.973 | 26.16 | 0.92 |

Figure 5.1 depicts the distributions of the LOS and transfer delay. Figure 5.1 demonstrates that the LOS distribution has a significantly long tail.

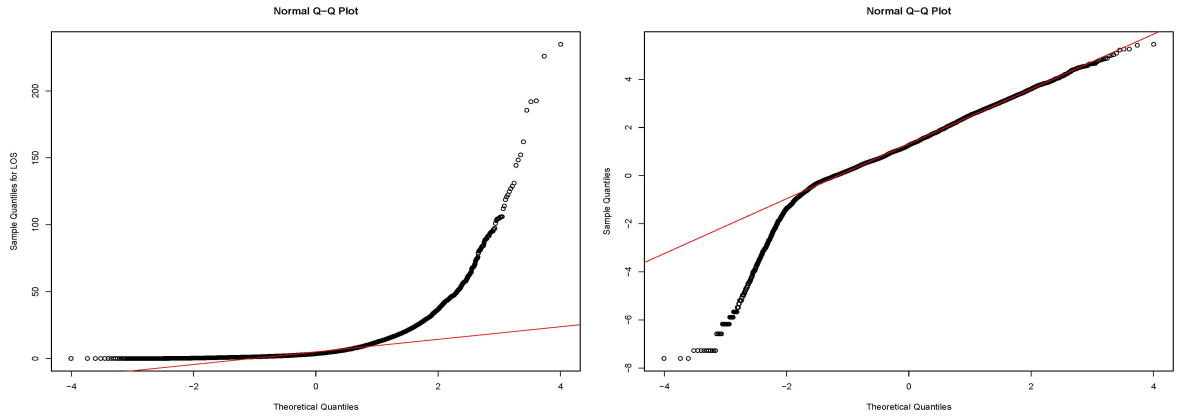



(a) LOS Histogram

(b) Transfer Delay Histogram

Figure 5.1: Histograms of LOS and transfer delay

Although it is clear visually that the LOS data is not normally distributed, we still provide the quantile-quantile (Q-Q) plot to show the normality assumption does not hold for the LOS distribution. Figure 5.2 presents Q-Q plot for the LOS data and the Q-Q plot for the log-transformed LOS data. We choose not the remove outliers from analysis, since they represent prolonged LOS's. Recall that our aim is to model LOS to include longer stays so that it can be used in resource management decisions.




(a) LOS (b) Log Transformed LOS

Figure 5.2: Q-Q Plots for LOS

As a starting point for modeling LOS, we fit three distributions; Weibull, Lognormal and Gamma. Table 5.2 summarizes the estimated parameters of the distributions.

Table 5.2: Distributions and estimated parameters

| Distribution | Notation | Parameters (standard error) |
|---|---|---|
| Weibull | $W(\lambda, k)$ | $k = 0.857\,(0.005)$<br>$\lambda = 6.488\,(0.063)$ |
| Lognormal | $\ln(N(\mu, \sigma))$ | $\mu = 1.266\,(0.010)$<br>$\sigma = 1.284\,(0.007)$ |
| Gamma | $G(\alpha, \beta)$ | $\alpha = 0.846\,(0.008)$<br>$\beta = 0.119\,(0.001)$ |

We also give log-likelihood, Akaike information criterion and Bayesian information criterion (BIC) values [140] in Table 5.3 for model selection.

Table 5.3: Results with different measures of goodness of fit

| Distribution | Log-Likelihood | AIC | BIC |
|---|---|---|---|
| Weibull | -47,160 | 94,324 | 94,340 |
| Lognormal* | -47,134 | 94,271 | 94,287 |
| Gamma | -47,409 | 94,822 | 94,837 |

Figure 5.3 presents the histogram of the LOS data and theoretical densities of the fitted distributions. Lognormal distribution matches the best visually the shape of the underlying LOS distribution. Our results confirm previous observations [50, 103] that Lognormal distribution is a good fit and superior to Gamma and Weibull to model LOS. Figures A1, A2 and A3 also present Q-Q plots and p-p plots for the fitted distributions. One can see from the Q-Q plots in the figures A1, A2 and A3 that all distributions are incapable of modeling prolonged stays.

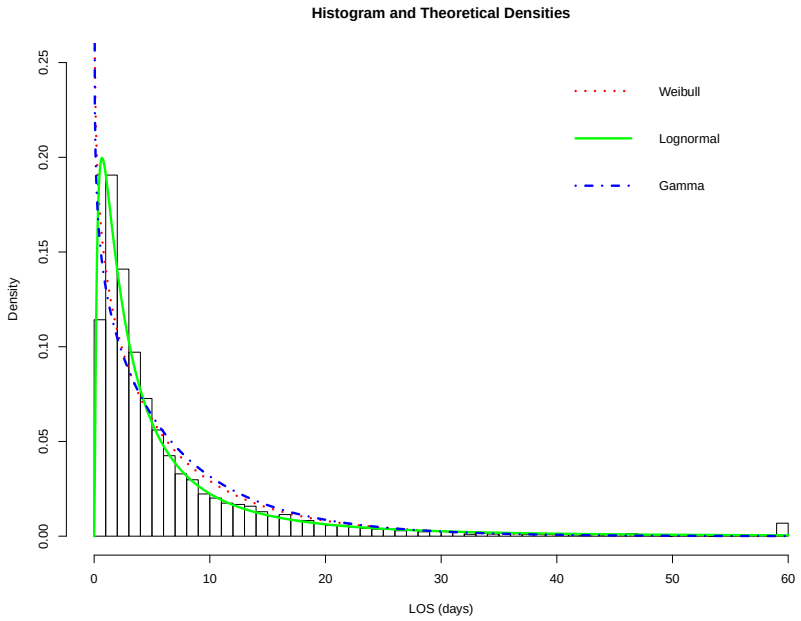



Figure 5.3: Histogram of LOS and theoretical densities of fitted distributions

## 5.3 STATISTICAL ANALYSIS OF INTENSIVE CARE UNIT LENGTH OF STAY

In this section, we present a statistical analysis to identify factors affecting LOS. We first employ a linear regression model, and discuss variables contributing positively to LOS. Next, we build two logistic regression models, where binary variable is defined as LOS being greater than a selected threshold. Similar to the analysis in Chapter 3, we use the step-wise backward selection method to remove variables with a p-value threshold 0.05. Some variables

used in the readmission models in Chapter 3 are unknown prior to discharge, thus we fit our models with variables available at the admission. We provide estimates of regression coefficients as well as 95% confidence intervals. We also present ROC-AUC curves to show the discriminatory power of the logistic regression models.

Table 5.4 presents significant variables in the linear regression model that independently predicted LOS. Moreover, Table 5.5 demonstrates summary statistics of the linear regression model.

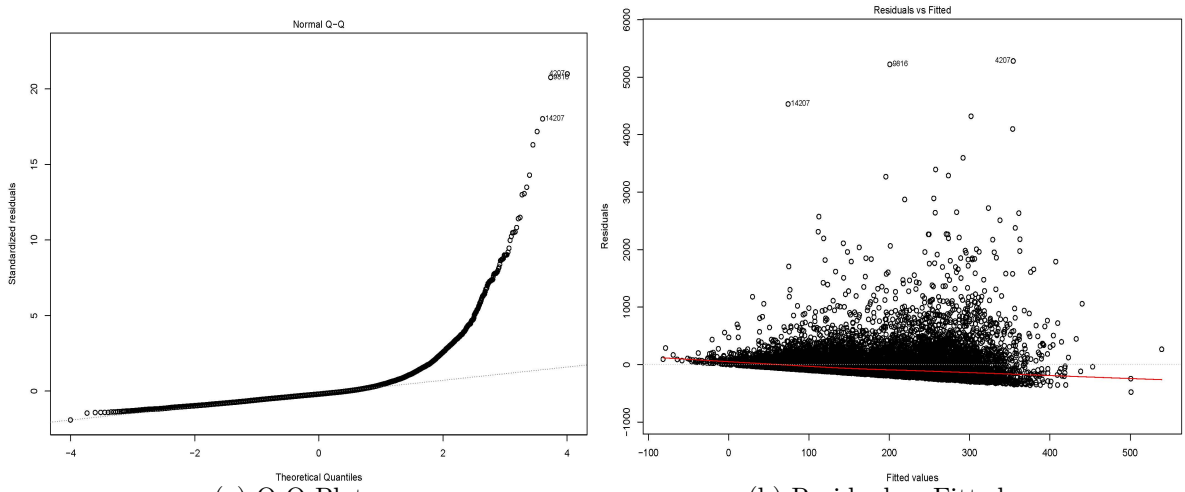
Table 5.4: Predictors of ICU linear regression LOS model

| Variables | Odd-ratio (95% Confidence Interval) | p-value |
|---|---|---|
| (Intercept) | -44.238 ( -77.003 - -11.472 ) | 0.008 |
| ICU Type Cardiac-T | -73.513 ( -90.890 - -56.136 ) | $< 0.001$ |
| ICU Type Medical | 42.948 ( 26.807 - 59.089 ) | $< 0.001$ |
| ICU Type Neurological | 32.877 ( 18.389 - 47.364 ) | $< 0.001$ |
| ICU Type Surgical | -9.906 ( -24.246 - 4.433 ) | 0.176 |
| ICU Type Trauma | 42.419 ( 16.484 - 68.353 ) | 0.001 |
| Origin Level of Care Regular | -12.479 ( -23.998 - -0.960 ) | 0.034 |
| Origin Level of Care SDU | -20.575 ( -30.133 - -11.017 ) | $< 0.001$ |
| Age | -0.282 ( -0.512 - -0.052 ) | 0.016 |
| Initial SOFA Liver | 23.242 ( 18.692 - 27.791 ) | $< 0.001$ |
| Initial SOFA Respiratory | 10.800 ( 5.298 - 16.301 ) | $< 0.001$ |
| APACHE III (Admission) | 1.224 ( 0.915 - 1.533 ) | $< 0.001$ |
| Charlson Index | -25.014 ( -37.741 - -12.287 ) | $< 0.001$ |
| Central Venous Catheter | 115.593 ( 107.070 - 124.115 ) | $< 0.001$ |
| Patient Type Surgical | 52.155 ( 43.682 - 60.628 ) | $< 0.001$ |
| Number of previous ICU admissions | 36.858 ( 28.747 - 44.969 ) | $< 0.001$ |

Table 5.5: Summary statistics ICU linear regression LOS model

| | |
|---|---|
| Residual standard error | 251.7 |
| Degrees of freedom | 16,037 |
| $R^2$ | 0.099 |
| Adjusted $R^2$ | 0.098 |
| F-statistic | 111.1 |
| p-value | $< 0.001$ |

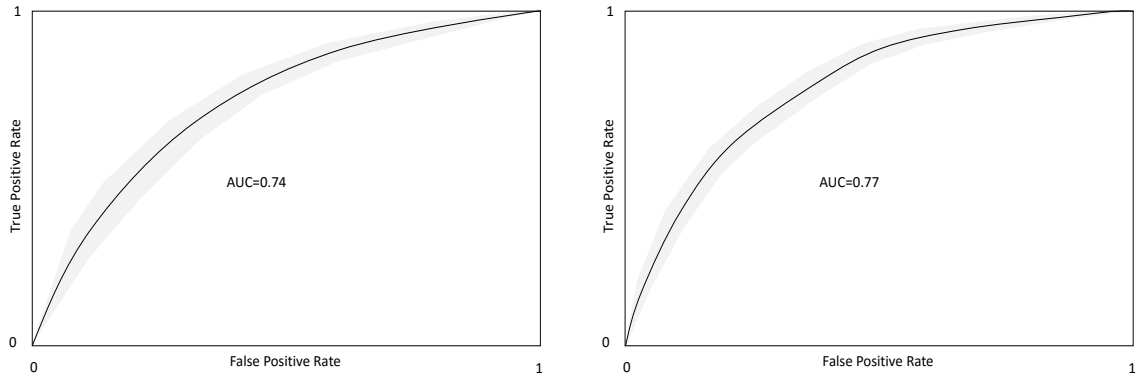
The model exhibits a poor fit with $R^2 = 0.1$. Verburg et al. [148] report that $R^2$ typically ranges 0.05-0.28 among published studies. Figure 5.4 also shows how the model fits the data poorly. However, recall that our goal is not to provide a prediction model to be used for planning but identify a selection of variables, which are significant to prolonged LOS. Verburg et al. [148] review several LOS prediction models and conclude that none of the models satisfy the requirements given in the study for planning purposes.




(a) Q-Q Plot (b) Residual vs Fitted

Figure 5.4: Diagnostic plots for LOS linear regression model

We build multivariate logistic regression models to predict unexpectedly longer LOS's. The aim of the models is to distinguish patients staying longer than a given threshold. We choose two thresholds; 10 days and 30 days as in [19, 120]. In our data set, 3,225 patients stayed longer than 10 days and 955 patients stayed longer than 30 days out of total 16,059 patients. We employ a similar 10-fold analysis for validation described in Chapter 3. 10-day model exhibits an AUC of 0.736 with a 95% confidence interval 0.735-0.737, and 30-day model exhibits an AUC of 0.770 with a 95% confidence interval 0.769-0.771. Figure 5.5 presents ROC curves for the logistic regression models. Interestingly, the discriminatory power increases almost 4% as the threshold moves from 10 days to 30 days. Although a more detailed analysis is necessary, we can argue that it is relatively easier to distinguish patients with very long LOS's. Tables 5.6 and 5.7 present predictors of logistic regression models.




(a) LOS longer than 10 days

(b) LOS longer than 30 days

Figure 5.5: ROC curves for logistic regression models for predicting prolonged LOS with 95% uncertainty bands

From the coefficients in Table 5.4 and odd-ratios in Tables 5.6 and 5.7, we observe that patients in relatively severe conditions at admission are more likely to stay longer in the ICU. Both initial SOFA scores and Admission APACHE III score have positive correlation with LOS similar to many studies [3, 30, 131, 141]. Additionally, patients who are previously admitted to the ICU tend to stay longer, as the corresponding coefficient is positive in the regression model (36.858), and the odd-ratios are greater than one in both logistic regression models (1.309-1.404). Surprisingly, age has a negative coefficient in the linear regression model (-0.282) and it is not a significant variable in logistic regression model. Contrary to the literature, increasing age is not a contributing factor to LOS. The model predicts surgical patients to stay longer than medical patients. Furthermore, patients who are admitted to the ICU from out of hospital have longer LOS's compared to the ones admitted from in-hospital units (stepdown unit or regular unit). Finally, trauma patients are more likely to have prolonged ICU stays.

Table 5.6: Predictors of ICU logistic regression model for predicting $LOS > 10$ days

| Variables | Odd-ratio (95% Confidence Interval) | p-value |
|---|---|---|
| (Intercept) | 0.015 ( 0.011 - 0.021 ) | $< 0.001$ |
| ICU Type Cardiac-T | 0.430 ( 0.351 - 0.526 ) | $< 0.001$ |
| ICU Type Medical | 1.544 ( 1.304 - 1.828 ) | $< 0.001$ |
| ICU Type Neurological | 1.760 ( 1.505 - 2.060 ) | $< 0.001$ |
| ICU Type Trauma | 2.103 ( 1.630 - 2.706 ) | $< 0.001$ |
| Origin Level of Care Regular | 0.792 ( 0.698 - 0.897 ) | $< 0.001$ |
| Origin Level of Care SDU | 0.699 ( 0.625 - 0.780 ) | $< 0.001$ |
| Initial SOFA Liver | 1.186 ( 1.134 - 1.239 ) | $< 0.001$ |
| Initial SOFA Renal | 1.065 ( 1.023 1.107 ) | 0.002 |
| Initial SOFA Respiratory | 1.178 ( 1.111 - 1.249 ) | $< 0.001$ |
| Initial SOFA Cardiovascular | 0.945 ( 0.901 0.990 ) | 0.017 |
| APACHE III (Admission) | 1.013 ( 1.010 - 1.016 ) | $< 0.001$ |
| Charlson Index | 0.762 ( 0.657 - 0.880 ) | $< 0.001$ |
| Central Venous Catheter | 4.003 ( 3.639 - 4.407 ) | $< 0.001$ |
| Patient Type Surgical | 1.736 ( 1.581 - 1.908 ) | $< 0.001$ |
| LOS prior to ICU admission | 1.001 ( 1.000 1.001 ) | 0.022 |
| Number of previous ICU admissions | 1.309 ( 1.209 - 1.416 ) | $< 0.001$ |

Table 5.7: Predictors of ICU logistic regression model for predicting $LOS > 30$ days

| Variables | Odd-ratio (95% Confidence Interval) | p-value |
|---|---|---|
| (Intercept) | 0.002 ( 0.001 - 0.003 ) | $< 0.001$ |
| ICU Type Cardiac-T | 0.692 ( 0.500 - 0.953 ) | 0.025 |
| ICU Type Medical | 1.907 ( 1.451 - 2.516 ) | $< 0.001$ |
| ICU Type Neurological | 1.458 ( 1.113 - 1.919 ) | 0.007 |
| ICU Type Trauma | 1.872 ( 1.236 - 2.788 ) | 0.002 |
| Origin Level of Care SDU | 0.828 ( 0.700 - 0.977 ) | 0.027 |
| Initial SOFA Liver | 1.233 ( 1.154 - 1.317 ) | $< 0.001$ |
| Initial SOFA Renal | 1.071 ( 1.006 1.139 ) | 0.031 |
| APACHE III (Admission) | 1.018 ( 1.013 - 1.023 ) | $< 0.001$ |
| Charlson Index | 0.623 ( 0.471 - 0.807 ) | $< 0.001$ |
| Central Venous Catheter | 4.444 ( 3.710 - 5.354 ) | $< 0.001$ |
| Patient Type Surgical | 2.992 ( 2.511 - 3.583 ) | $< 0.001$ |
| Number of previous ICU admissions | 1.404 ( 1.253 - 1.567 ) | $< 0.001$ |

## 5.4 STOCHASTIC INTENSIVE CARE UNIT LENGTH OF STAY MODEL

We start by assuming the score follows a scalar stochastic process $\{S_t : t \geq 0\}$ as long as the patient stays in the ICU. Define a random variable $\mathscr{L}$ as the LOS in the ICU. We will express $\mathscr{L}$ based on the score process and bed availability dynamics. We assume that a patient is medically non-transferable to a downstream unit if her score at time $t$, $S_t$, is above a certain threshold $\psi$. We further define a request threshold $\kappa \geq \psi$, where a bed request is given and the bed becomes available after a random amount of time $\delta$.

Let $\tau_r^1$ be the first time to reach the request threshold from admission, i.e., $\tau_r^1 = \inf\{t > 0 : S_t \leq \kappa\}$. Further define $\tau_d^1 = \tau_r^1 + \delta$ be the sum of the first time to reach the request threshold and the time when the target bed is available. Now, if the patient's status is transferable at time $\tau_d^1$, i.e., $S_{\tau_d^1} \leq \psi$, then the patient is transferred, and hence the LOS is simply $\mathscr{L} = \tau_d^1$; otherwise we need to wait until a patient's score reaches transfer threshold $\kappa$ again and a request is given again, or if the score is already below the request threshold but above the medical threshold, then a new request is given immediately.

Let $i$ be the number of times the described cycle is repeated, and then define $\tau_d^i = \tau_r^i + \delta$, and $\tau_r^i = \inf\{t \geq \tau_d^{i-1} : S_t \leq \kappa\}$ recursively. Finally, let $\Omega$ be the random total number of times the cycle is repeated. Then LOS is $\mathscr{L} = \sum_{i=1}^{\Omega} \tau_d^i - \tau_d^{i-1}$. Figure 5.6 illustrates a sample path of the LOS process.

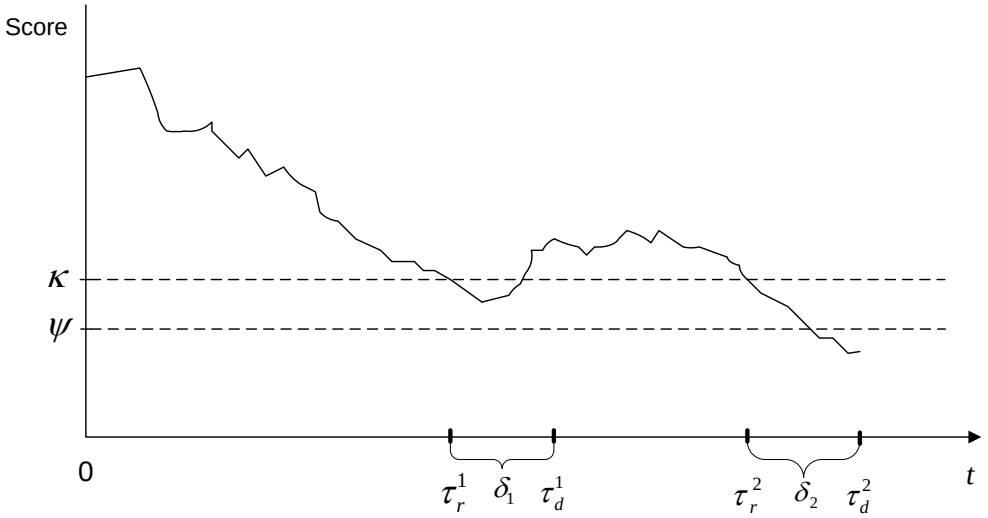



Figure 5.6: Illustration of a sample path of the LOS process: The patient was first healthier than $\kappa$ at time $\tau_r^1$, so a downstream bed was requested. However, when the bed was ready $\delta^1$ time units later (at time $\tau_d^1$), the patient was above $\psi$, so too sick to be transferred. However, at time $\tau_r^2$, the next time the patient was healthier than the request threshold $\kappa$, the transfer occurred $\delta_2$ time units later (at time $\tau_d^2$)

Next, we characterize the LOS process under certain assumptions. In particular, we show that if (1) the score process follows a DTMC (CTMC) and (2) the transfer delay is geometrically (exponentially) distributed, then the LOS process is a phase-type distribution.

### 5.4.1 The Score Process as a Discrete Time Markov Chain

We assume that the score process is a DTMC and each patient's health evolves independently. Let $\mathcal{S} = \{1, \ldots, S\}$ be the score set, and $\mathbf{P}$ be the transition probability matrix of the score evolution of a patient.

$$\mathbf{P} = \begin{bmatrix} p_{11} & \cdots & p_{1\psi} & \cdots & p_{1\kappa} & \cdots & p_{1S} \\ \vdots & & \vdots & & \vdots & \cdots & \vdots \\ p_{\psi 1} & \cdots & p_{\psi\psi} & \cdots & p_{\psi\kappa} & \cdots & p_{\psi S} \\ \vdots & & \vdots & & \vdots & \cdots & \vdots \\ p_{\kappa 1} & \cdots & p_{\kappa\psi} & \cdots & p_{\kappa\kappa} & \cdots & p_{\kappa S} \\ \vdots & & \vdots & & \vdots & & \vdots \\ p_{S1} & \cdots & p_{S\psi} & \cdots & p_{S\kappa} & \cdots & p_{SS} \end{bmatrix}.$$

Further assume that a requested downstream bed becomes available in the next period with probability $\alpha > 0$. Now, we will model LOS as the time to absorption of a DTMC by employing the score process defined by $\mathbf{P}$ and the transfer request phenomena. First, define $\Delta$ as an absorbing state where the patient is discharged. Further, define $(s, R)$ for $\psi < s \leq \kappa$ to represent the states where the downstream bed is ready and the patient has a requestable score. We also define $\psi < s \leq \kappa$ as the states where the patient has a requestable score but the downstream bed is not available. We should note that, if the patient has a transferable score and the downstream bed is ready then she is discharged, hence $s \leq \psi$ can be defined as the states where the patient is transferable and the downstream bed is not available. Finally, define $s > \kappa$ as non-requestable states as the patient has a non-requestable score. The patient stays in the ICU if her score is non-requestable. The probability of going to any other state $s'$ from a non-requestable state is $p_{ss'}$ (5.1a), where $s > \kappa$, and $s' \in \mathcal{S}$. Once the process reaches to a requestable state $s \leq \kappa$, as patient's score reaches a requestable score, a transfer request is given. The patient can be discharged and the LOS process terminates if the patient's score switches to one of the transferable scores $s' \leq \psi$ and the downstream bed becomes available. Thus, the probability of going to the absorbing state is $\alpha \sum_{i=1}^{\psi} p_{si}$ where $s \leq \kappa$, given that the process is in $s \leq \kappa$ presented in (5.1b). The process stays in a requestable state $s' \leq \kappa$, if the downstream bed does not become available, which has probability $(1-\alpha)p_{ss'}$, $s \leq \kappa$, $s' \leq \kappa$ given in (5.1c). The state switches to $(s', R)$, if the downstream bed becomes available but the patient score becomes requestable but non-transferable $\psi < s' \leq \kappa$, which has probability $\alpha p_{sj}$, $s \leq \kappa$, $\psi < j \leq \kappa$ given in (5.1d). Recall that, if the state switches to a non-requestable state, we release the downstream bed and wait until it reaches to a

requestable state to make a request again. The probability of going to a non-requestable state from a requestable state is $p_{ss'}$, $s \leq \kappa$, and $s' > \kappa$ presented in (5.1e). The process can switch to other non-requestable states with probability $p_{is'}$, where $\psi < i \leq \kappa$, $s' > \kappa$ given in (5.1f). In addition, the process may stay in a requestable and ready bed states with probability $p_{ij}$, where $\psi < i \leq \kappa$, $\psi < j \leq \kappa$ presented in (5.1g). Finally, if the process is in state $(i, R)$ where $\psi < i \leq \kappa$, the process switches to the absorbing state only if the patient's score becomes transferable, which has probability $\sum_{k=1}^{\psi} p_{ik}$ given in (5.1h), since the downstream bed is already available. All the remaining transition probabilities are zero, except the process stays in $\Delta$ with probability one. Let $\mathbf{H}$ be the transition probability matrix of the described process, where each entry is summarized.

$$
H(s'|s) = \begin{cases}
p_{ss'}, & \text{if } s > \kappa \text{ and } s' \in \mathcal{S} \quad (5.1a)\\
\alpha \sum_{i=1}^{\psi} p_{si}, & \text{if } s \leq \kappa \text{ and } s' = \Delta \quad (5.1b)\\
(1-\alpha) p_{ss'}, & \text{if } s \leq \kappa \text{ and } s' \leq \kappa \quad (5.1c)\\
\alpha p_{sj}, & \text{if } s \leq \kappa \text{ and } s' \in \{(j, R) : \psi < j \leq \kappa\} \quad (5.1d)\\
p_{ss'}, & \text{if } s \leq \kappa \text{ and } s' > \kappa \quad (5.1e)\\
p_{is'}, & \text{if } s \in \{(i, R) : \psi < i \leq \kappa\} \text{ and } s' > \kappa \quad (5.1f)\\
p_{ij}, & \text{if } s \in \{(i, R) : \psi < i \leq \kappa\} \text{ and } s' \in \{(j, R) : \psi < j \leq \kappa\} \quad (5.1g)\\
\sum_{k=1}^{\psi} p_{ik}, & \text{if } s \in \{(i, R) : \psi < i \leq \kappa\} \text{ and } s' = \Delta \quad (5.1h)\\
0, & \text{if } s > \kappa \text{ and } s' \in \Delta \cup \{(s', R) : \psi < s' \leq \kappa\} \quad (5.1i)\\
0, & \text{if } s \in \{(i, R) : \psi < i \leq \kappa\} \text{ and } s' \leq \kappa \quad (5.1j)\\
0, & \text{if } s = \Delta \text{ and } s' \neq \Delta \quad (5.1k)\\
1, & \text{if } s = \Delta \text{ and } s' = \Delta. \quad (5.1l)
\end{cases}
$$

$\mathbf{H}$ can be rewritten as

$$
\mathbf{H} = \begin{bmatrix} & & | \\ & \mathbf{T} & T^0 \\ & & | \\ - & 0 & - \quad 1 \end{bmatrix},
$$

where $T^0$ is the column vector of the transition probabilities to the absorbing state $\Delta$ from transient states. Further define $Q$ as the initial state distribution. Proposition 5.1 gives the complete characterization of $\mathscr{L}$.

**Proposition 5.1.** *If the score process follows a DTMC determined by* $\mathbf{P}$*, and transfer request policy is threshold-type and a requested downstream bed becomes available in the next period with probability* $\alpha > 0$*, then the CDF of* $\mathscr{L}$ *is given by:*

$$F(\mathscr{L} = k) = 1 - Q\mathbf{T}^k\mathbf{I},$$

*the pdf of* $\mathscr{L}$ *is given by:*

$$f(\mathscr{L} = k) = Q\mathbf{T}^{k-1}\,T^0,$$

*and the* $n^{th}$ *factorial moment is given by:*

$$E[\mathscr{L}(\mathscr{L}-1)\dots(\mathscr{L}-n+1)] = n!\,Q(\mathbf{I}-\mathbf{T})^{-n}\mathbf{T}^{n-1}\mathbf{I},$$

*where* $\mathbf{I}$ *is the identity matrix.*

*Proof.* The result follows from the fact that $\mathscr{L}$ represents the time to absorption of discrete time Markov chain $\mathbf{H}$ given that the process starts from an initial state with distribution $Q$, hence it is a phase-type distribution. See [115] for characterization results. □

### 5.4.2 The Score Process as a Continuous Time Markov Chain

Now, we consider the case of a CTMC score process. We assume that the TRS process is a CTMC and each patient's health evolves independently. Let $\mathcal{S} = \{1, \dots, S\}$ be the score set, and $\mathbf{P}$ be the infinitesimal generator of the health evolution of a patient.

$$\mathbf{P} = \begin{bmatrix} -\sum_i^S \mu_{1i} & \cdots & \mu_{1\psi} & \cdots & \mu_{1\kappa} & \cdots & \mu_{1S} \\ \vdots & \ddots & \vdots & & \vdots & & \vdots \\ \mu_{\psi 1} & \cdots & -\sum_i^S \mu_{\psi i} & \cdots & \mu_{\psi\kappa} & \cdots & \mu_{\psi S} \\ \vdots & & \vdots & \ddots & \vdots & & \vdots \\ \mu_{\kappa 1} & \cdots & \mu_{\kappa\psi} & \cdots & -\sum_i^S \mu_{\kappa i} & \cdots & \mu_{\kappa S} \\ \vdots & & \vdots & & \vdots & \ddots & \vdots \\ \mu_{S1} & \cdots & \mu_{S\psi} & \cdots & \mu_{S\kappa} & \cdots & -\sum_i^S \mu_{Si} \end{bmatrix}.$$

Further assume that transfer delay is exponentially distributed with rate $\lambda > 0$. Similarly, we will model LOS as the time to absorption of a CTMC by employing the score process defined by $\mathbf{P}$ and the transfer request phenomena. The state definitions of the CTMC is the same as the DTMC described in Section 5.4.1. The patient stays in the ICU the patient has a non-requestable score. The rate of going to any other states $s'$ given that the state is non-requestable is $\mu_{ss'}$, where $s > \kappa$, and $s' \in \mathcal{S}$ given in (5.2a). Once the process reaches to a requestable state $s \leq \kappa$, as patient's score reaches a requestable score, a transfer request is given. The patient can be discharged or the process transitions to the absorbing state $\Delta$, if the downstream bed becomes available and the patient's score is transferable. The rate of going to discharge state given that the process is in a state where the patient is transferable and the downstream bed is not available is $\lambda$ presented in (5.2b). The process can also switch to a requestable state or a non-requestable state $s' \leq \kappa$ with rate $\mu_{ss'}$, $s \leq \psi$, $s' \in$ given in (5.2c). Given that the process is in a state that patient has a requestable but not transferable score, the process may switch to $(s', R)$ with rate $\lambda$, where $\psi < s' \leq \kappa$ presented in (5.2d). Recall that, if the state switches to a non-requestable state, we release the downstream bed and wait until it reaches to a requestable state to make a request again. The process switches to a non-requestable state with rate $\mu_{is'}$, given that the patient has a requestable $\psi < i \leq \kappa$ but not transferable score and the downstream bed is ready presented in (5.2e).

If the process is in state $(i, R)$, where $\psi < i \leq \kappa$, the process can go to the absorbing state if the score process switches to a transferable state, which has rate of $\sum_{k=1}^{\psi} \mu_{ik}$, since the downstream bed is already available presented in (5.2f). In addition, the process can also switch to other requestable and ready bed states with rate $\mu_{ij}$, where $\psi < i \leq \kappa$, $\psi < j \leq \kappa$ given in (5.2g). Finally, the process can switch to other non-requestable states with rate $\mu_{ss'}$, where $\psi < s \leq \kappa$, $\psi < s' \leq \kappa$ presented in (5.2h). All the remaining transition rates are zero. Let $\mathbf{H}$ be the infinitesimal generator matrix of the described process, where each entry is summarized.

$$H(s'|s) = \begin{cases} \mu_{ss'}, & \text{if } s > \kappa \text{ and } s' \in \mathcal{S} & (5.2a) \\ \lambda, & \text{if } s \leq \psi \text{ and } s' = \Delta & (5.2b) \\ \mu_{ss'}, & \text{if } s \leq \psi \text{ and } s' \leq \mathcal{S} & (5.2c) \\ \lambda, & \text{if } \psi < s \leq \kappa \text{ and } s' = (s, R) & (5.2d) \\ \mu_{is'}, & \text{if } s \in \{(i, R) : \psi < i \leq \kappa\} \text{ and } s' > \kappa & (5.2e) \\ \sum_{k=1}^{\psi} \mu_{ik}, & \text{if } s \in \{(i, R) : \psi < i \leq \kappa\} \text{ and } s' = \Delta & (5.2f) \\ \mu_{ij}, & \text{if } s \in \{(i, R) : \psi < i \leq \kappa\} \text{ and } s' \in \{(j, R) : \psi < j \leq \kappa\} & (5.2g) \\ \mu_{ss'}, & \text{if } \psi < s \leq \kappa \text{ and } s' \in \mathcal{S} & (5.2h) \\ 0, & \text{if } s > \kappa \text{ and } s' \in \Delta \cup \{(s', R) : \psi < s' \leq \kappa\} & (5.2i) \\ 0, & \text{if } \psi < s \leq \kappa \text{ and } s' = \Delta & (5.2j) \\ 0, & \text{if } s = \Delta \text{ and } s' = \Delta & (5.2k) \\ 0, & \text{if } s = \Delta \text{ and } s' \neq \Delta. & (5.2l) \end{cases}$$

We skip to present self-loops, which are negative sum of out rates. $\mathbf{H}$ can be rewritten as

$$\mathbf{H} = \begin{bmatrix} & & | \\ & \mathbf{T} & T^0 \\ & & | \\ - & 0 & - \quad 0 \end{bmatrix},$$

where $T^0$ is the column vector of the transition rates to the absorbing state $\Delta$ from transient states. Further define $Q$ as the initial state distribution. Proposition 5.2 presents the complete characterization of $\mathscr{L}$.

**Proposition 5.2.** *If the score process follows a CTMC determined by* $\mathbf{P}$*, the transfer request policy is threshold-type and a downstream bed becomes available after exponentially distributed transfer delay with rate* $\lambda > 0$*, then the CDF of* $\mathscr{L}$ *is given by:*

$$F(\mathscr{L} \leq \ell) = 1 - Qe^{(\mathbf{T}\ell)}\mathbf{I},$$

*the pdf of* $\mathscr{L}$ *is given by:*

$$f(\ell) = Qe^{(\mathbf{T}\ell)}T^0,$$

*and the* $n^{th}$ *moment is is given by:*

$$E[\mathscr{L}^n] = (-1)^n n! Q\mathbf{T}^{-n}\mathbf{I},$$

*where* $\mathbf{I}$ *is the identity matrix.*

*Proof.* The result follows from the fact that $\mathscr{L}$ represents the time to absorption of continuous time Markov chain $\mathbf{H}$ given that the process starts from an initial state with distribution $Q$, hence it is a phase-type distribution. See [115] for characterization results. □

## 5.5 GOODNESS OF FIT PERFORMANCE OF THE MODEL AND THE COXIAN DISTRIBUTION

In this section, we compare the goodness of fit of the proposed stochastic model with the Coxian distribution, a special type phase-type distribution. The Coxian distribution differs from general phase-type distributions in that the transient states (or phases) are ordered. The chain has $k$ phases and the time spent in phase $i$ is exponentially distributed with rate $\mu_i$, and the process may switch to phase $i+1$, after phase $i$, with probability $p_i$ or be absorbed with probability $1 - p_i$. The process starts from phase 1 with probability 1. The representation of the Coxian distribution is given in (5.3).

$$\mathbf{T} = \begin{bmatrix} -\mu_1 & p_1\mu_1 & 0 & 0 & \cdots & 0 \\ 0 & -\mu_2 & p_2\mu_2 & 0 & \cdots & 0 \\ 0 & 0 & -\mu_3 & p_3\mu_3 & & 0 \\ 0 & 0 & 0 & \ddots & \ddots & 0 \\ \vdots & \vdots & \ddots & \ddots & -\mu_{k-1} & p_{k-1}\mu_{k-1} \\ 0 & 0 & \cdots & 0 & 0 & -\mu_k \end{bmatrix}. \tag{5.3}$$

Fitting the Coxian distribution is relatively easier than fitting general phase-type distributions, since it requires fewer parameter estimations. It has also been shown that the Coxian distribution is a good fit for the LOS of geriatric patients [49, 62, 105, 106]. Recall that the stochastic LOS model is also a phase-type distribution. However, the main difference between the stochastic LOS model and the Coxian distribution is that the phases of the Coxian distribution do not represent physical conditions, besides the difference on the structure of the transitions. In fact, the stochastic model is more of a descriptive model of the LOS based on the evolution of physical conditions of patients. Hence, the parametrization methods are completely different. We estimate the transition probabilities of the LOS chain using the score transition probabilities that we have estimated from the data set, whereas the parameters of the Coxian is estimated via a log-likelihood maximization algorithm.

For the goodness of fit analysis, we build two LOS models with: 10 score states $S_1 = \{1, 2, \ldots, 10\}$ and 20 score states $S_2 = \{1, 2, \ldots, 20\}$. As in Chapter 4, we estimate the readmission probabilities for all 16,059 patients from admission to discharge every six hours and map the probabilities to corresponding scores by assuming equally spaced intervals, e.g.,score $i$ maps into the $(i-1)^{\text{th}}$ probability interval. Next, we estimate the transition probability matrix $\mathbf{P}$ using the data set. We further choose 0.1 as the probability threshold of transfer, which corresponds to score-1 for 10 score state LOS model and score-2 for 20 score state model. We set the transferable score $\psi$ and requestable score $\kappa$ same, since in practice physicians are not utilizing anticipative transfer request policy. We should note that the simulation is discrete time as the score changes every six hours. Finally, we set the probability of downstream bed availability probability to 0.63, which corresponds to on

average of 9.5 hours delay, because the step length of the simulation is six hours. We perform 100,000 replications. For the Coxian distribution, we set the number of phases to 5 and use the EMPht algorithm developed by Asmussen et al. [5].

Table 5.8 summarizes the descriptive statistics of the models. Figure 5.7 presents the histogram of empirical LOS and density functions of three models. Moreover, Figure 5.8 depicts the CDFs of the models compared with the empirical LOS CDF.

Table 5.8: Descriptive statistics of the stochastic LOS models and the Coxian distribution

| | Mean | Median | Std. Dev. | Skewness | 99% Quantile |
|---|---|---|---|---|---|
| Empirical LOS | 7.12 | 3.54 | 11.05 | 5.82 | 50.93 |
| Stochastic model with 10 score states | 5.77 | 3.38 | 6.67 | 2.22 | 30.38 |
| Stochastic model with 20 score states | 6.22 | 3.63 | 7.30 | 2.21 | 33.13 |
| Coxian | 7.13 | 3.55 | 11.06 | 5.43 | 45.23 |

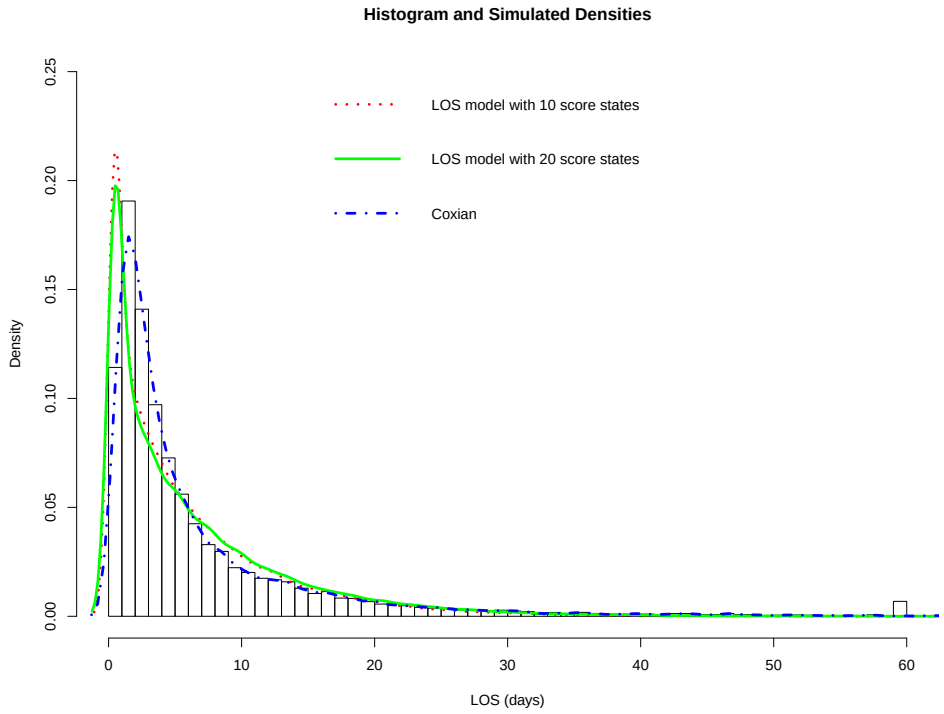



Figure 5.7: Histogram of empirical LOS and density functions of fitted LOS models

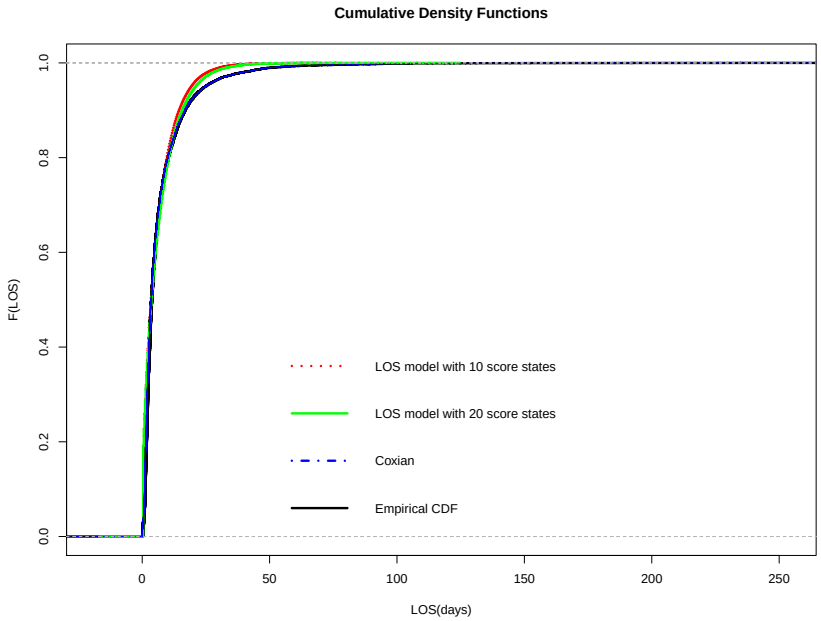



Figure 5.8: Cumulative density functions of empirical LOS and fitted LOS models

The Coxian distribution outperforms the LOS model in matching the moments. The stochastic LOS model slightly underestimates the mean LOS (12.5%), closely matches the median and overly underestimates the standard deviation and skewness. However, it matches the shape of the underlying distribution. Additionally, the Coxian distribution is better in modeling the long tails, as its 99% percentile is closest to that of the empirical LOS. We should note that the 20-score state model performs better than the 10-score state model. One possible explanation is that we may lose some information through discretization when creating scores from probabilities. However, a more detailed analysis is needed to conclude that a score set with more states is better, since having more states will lead to less reliable transition probability estimates due to a decrease in the sample size.

There are some possible explanations of the moderate fitting performance of the stochastic model, which also addresses the possible improvement directions: (1) The Markovian assumption of the score process and geometrically distributed transfer delay assumption are restrictive. (2) The model assumes the physical conditions of the patients can be modeled with a single scalar based on probability of readmission. However, in practice physicians con-

sider many other things such as whether downstream unit provides adequate care regarding resources to treat patients when giving discharge decisions. One future direction is to extend the state description to incorporate conditions. (3) The predictive power of the readmission model is not excellent (AUC-0.77); hence it would be bold to state the transfer readiness score models the patients' health conditions. We believe that the results are encouraging and a better fitting descriptive model can be built and utilized.

## 5.6 CONCLUSIONS

We present a statistical and probabilistic analysis of ICU LOS. In particular, we build linear and logistic regression models to identify factors yielding longer LOS's. Next, we present an explicit stochastic model of LOS that incorporates patient physiology modeled by transfer readiness score, as well as the transfer delay dynamics. We characterize the LOS process for a DTMC and a CTMC score processes. We show that the resulting LOS process is a phase-type distribution. Finally, we test the goodness of fit of the model with the LOS data. The stochastic model slightly underestimates the LOS, whereas it captures the shape of the empirical LOS distribution. The Coxian distribution provides the best performance in matching the moments and the long tail of the LOS distribution. We demonstrate that a simple stochastic model based on physiology can moderately predict LOS and improvements on the score potentially improves the goodness of fit.

# 6.0 CONCLUSIONS AND FUTURE DIRECTIONS

Modeling patient flow through an ICU is challenging because, unlike other patient flow models, a stochastic and dynamic model of physiology is crucial. This is very challenging, and requires enormous volumes of highly detailed data. This dissertation focuses data-driven approaches to optimally manage operations in the ICUs by: (1) creating a dynamic transfer readiness score and an explicit stochastic LOS model based on patient physiology and transfer delay dynamics; (2) developing an optimization model to make anticipative bed requests using the created dynamic and stochastic score.

In Chapter 3, we construct prediction models to estimate readmission and death probabilities upon transfer to a downstream unit. Our models outperform any published model in predicting readmissions. We show that the inclusion of dynamic data at least partially explains improved performance. For both the ICU readmission and late mortality models, we identified some potentially modifiable factors. Accurate predictions of performance measures are essential for optimal management of units. In that regard, we construct a new Transfer Score to model the dynamic nature of patient physiology by employing proposed prediction models.

In Chapter 4, we develop an optimization model to make anticipative bed requests using the created dynamic and stochastic score. We illustrate that an anticipative transfer request policy significantly increases the efficiency by reducing the transfer delays (up to 50% reduction) and hence reducing the congestion in the ICU. We further investigate the sensitivity of policy change upon cost parameter estimation errors by using robust models, and demonstrate that proactive strategies are more beneficial than reactive current policy in most scenarios.

In Chapter 5, we construct an explicit stochastic LOS model based on transfer score and delay dynamics. We characterize the LOS process under certain conditions. We compare the performance of the model with the Coxian distribution. We demonstrate that a model based on physiology can moderately explain a complex outcome, and can be utilized for planning purposes.

Our study have some certain limitations and potentially future extensions. Our data set reflects some bias towards more complete data on sicker patients. The readmission and mortality prediction models need to be tested with more data. Throughout the study, we assume that patient's physiology can be modeled with a single scalar based on probability of readmission. However, in practice physicians consider many other things such as whether downstream unit provides adequate care regarding devices, drugs, or training of nurses. One future direction is to extend the state description to incorporate aforementioned conditions. A better state description will also improve the performance of the anticipative transfer request policies. However, new approximation algorithms are also needed to generate policies in a reasonable time for such extended state space. Constructing different approximation methods with optimality bound guarantees is another interesting research direction to pursue. Our simulation model considers an isolated ICU-downstream system. A more complex simulation considering other units, i.e., hospital network as a whole, can be developed to assess the benefits of the anticipative policies. Testing the goodness of fit of the LOS model with different score processes and transfer request policies is an immediate future research direction. Finally, determining the minimal state definition so that the transfer readiness score process is Markovian is an exciting problem, since the Markov property is critical in optimization models.

# APPENDIX

## PROBABILITY DISTRIBUTION FIT DIAGNOSTIC PLOTS

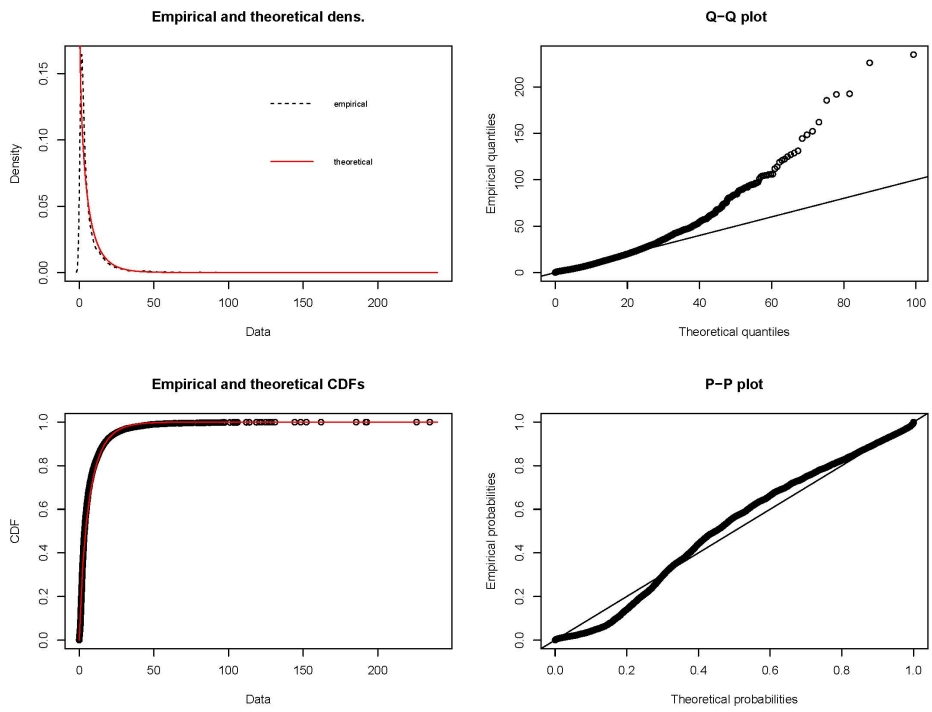



Figure A1: Weibull distribution fit diagnostic plots

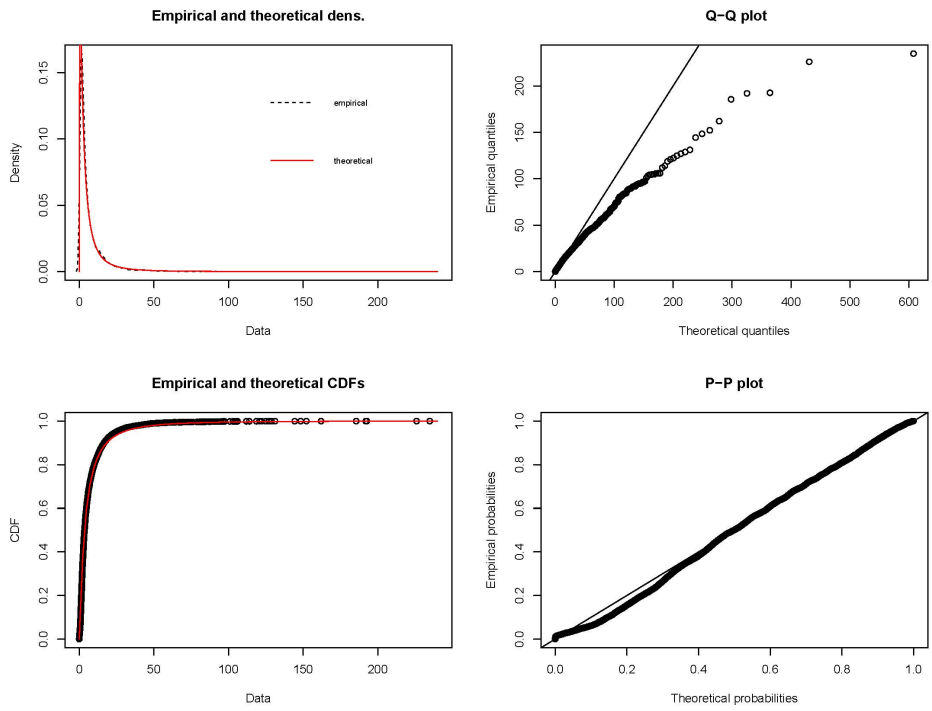



Figure A2: Lognormal distribution fit diagnostic plots

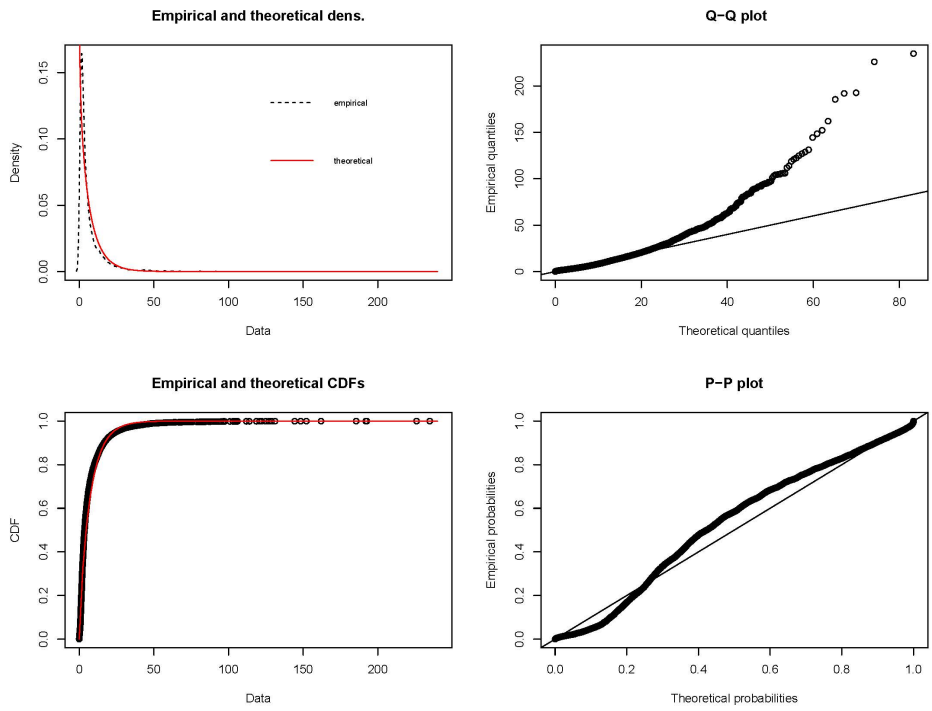



Figure A3: Gamma distribution fit diagnostic plots

## BIBLIOGRAPHY


[1] D. Adelman and A. J. Mersereau. Relaxations of weakly coupled stochastic dynamic programs. *Operations Research*, 56(3):712–727, 2008.

[2] D. Adler, E. Dupuis-Lozeron, J. C. Richard, J. P. Janssens, and L. Brochard. Does inspiratory muscle dysfunction predict readmission after intensive care unit discharge? *American Journal of Respiratory and Critical Care Medicine*, 190(3):347–350, 2014.

[3] M. Antonelli, R. Moreno, J. L. Vincent, C. L. Sprung, A. Mendoca, M. Passariello, L. Riccioni, and J. Osborn. Application of SOFA score to trauma patients. *Intensive Care Medicine*, 25(4):389–394, 1999.

[4] M. Armony, C. W. Chan, and B. Zhou. Critical care in hospitals: When to introduce a step down unit? Working paper available from http://www.columbia.edu/~cc3179/icusdu_2015.pdf, 2015.

[5] S. Asmussen, O. Nerman, and M. Olsson. Fitting phase-type distributions via the em algorithm. *Scandinavian Journal of Statistics*, pages 419–441, 1996.

[6] A. Azari, V. P. Janeja, and A. Mohseni. Predicting hospital length of stay (phlos): A multi-tiered data mining approach. In *2012 IEEE 12th International Conference on Data Mining Workshops*, pages 17–24, 2012.

[7] O. Badawi and M. J. Breslow. Readmissions and death after ICU discharge: Development and validation of two predictive models. *PLOS ONE*, 7(11):1–15, 2012.

[8] F. Barilla, N. Barzaghib, F. H. Cheemac, A. Capoa, J. Jiangc, E. Ardemagnia, M. Argenzianoc, and C. Grossia. An original model to predict intensive care unit length-of stay after cardiac surgery in a competing risk framework. *International Journal of Cardiology*, 168:219–225, 2013.

[9] R. E. Barlow and F. Proschan. *Mathematical Theory of Reliability*. John Wiley & Sons, New York, 1965.

[10] M. L. Barrett, M. W. Smith, A. Elixhauser, L. S. Honigman, and J. M. Pines. Utilization of intensive care services, 2011. Technical report, Agency for Healthcare Re-

search and Quality, 2014. Available from http://www.hcup-us.ahrq.gov/reports/statbriefs/sb185-Hospital-Intensive-Care-Units-2011.pdf.

[11] D. P. Bertsekas and J. N. Tsitliklis. *Neuro-Dynamic Programming*. Athena Scientific, 1st edition, 1996.

[12] R. C. Bones, R. A. Balk, and F. B. Cerra. Definitions for sepsis and organ failure and guidelines for the use of innovative therapies in sepsis. The ACCP/SCCM consensus conference committee. American college of chest physicians/society of critical care medicine. *Chest*, 101:1644–1655, 1992.

[13] T. Bountourelis, L. Luangkesorn, A. J. Schaefer, L. M. Maillart, S. Nabors, and G. Clermont. Development and validation of a large scale ICU simulation model with blocking. In S. Jain, R. R. Creasey, J. Himmelspach, K. P. White, and M. Fu, editors, *Proceedings of the 2011 Winter Simulation Conference*, pages 1143–1153, Piscataway, NJ, 2011. Institute of Electrical and Electronics Engineers.

[14] M. L. Brandeau, F. Sainfort, and W. P. Pierskalla. *Operations Research and Health Care: A Handbook of Methods and Applications*. Kluwer Academic Publishers, Massachusetts, MA, 2004.

[15] S. Brown, S. J. Ratcliffe, J. M. Kahn, and S. Halpern. ICU readmissions in United States hospitals: Incidence and trends. In *A104. INTENSIVE CARE UNIT ORGANIZATION, OUTCOMES, AND RESEARCH*, pages A2374–A2374. American Thoracic Society, 2011.

[16] S. E. Brown, S. J. Ratcliffe, J. M. Kahn, and S. D. Halpern. The epidemiology of intensive care unit readmissions in the United States. *American Journal of Respiratory and Critical Care Medicine*, 185(9):955–964, 2012.

[17] T. G. Buchman, K. L. Kubos, A. J. Seidler, and M. J. Siegforth. A comparison of statistical and connectionist models for the prediction of chronicity in a surgical intensive care unit. *Critical Care Medicine*, 22(5):750–762, 1994.

[18] A. J. Campbell, J. A. Cook, G. Adey, and B. H. Cuthbertson. Predicting death and readmission after intensive care discharge. *British Journal of Anaesthesia*, 100:656–662, 2008.

[19] M. G. Castillo. Modelling patient length of stay in public hospitals in Mexico. Available from https://eprints.soton.ac.uk/345554/, May 2012.

[20] L. A. Celi and R. G. Mark. "Big data" in the intensive care unit. Closing the data loop. *American Journal of Respiratory and Critical Care Medicine*, 187(11):1157–1166, 2013.

[21] D. B. Chalfin, S. Trzeciak, A. Likourezos, and B. M. Baumann. Impact of delayed transfer of critically ill patients from the emergency department to the intensive care unit. *Critical Care Medicine*, 35(6):1477–1483, 2007.

[22] C. W. Chan, V. F. Farias, N. Bambos, and G. J. Escobar. Optimizing intensive care unit discharge decisions with patient readmissions. *Operations Research*, 60(6):1323–1341, 2012.

[23] C. W. Chan, V. F. Farias, and G. J. Escobar. The impact of delays on service times in the intensive care unit. To appear in *Management Science* available from http://www.columbia.edu/~cc3179/ED2ICUdelay_2013.pdf, 2016.

[24] C. W. Chan, G. Yom-Tov, and G. J. Escobar. When to use speedup: An examination of service systems with returns. *Operations Research*, 62(2):462–482, 2014.

[25] L. Chelluri, M. R. Pinsky, M. P. Donahoe, and A. Grenvik. Long-term outcome of critically ill elderly patients requiring intensive care. *The Journal of the American Medical Association*, 269(24):3119–3123, 1993.

[26] A. M. Cheung, C. M. Tansey G. Tomlinson, N. Diaz-Granados, A. Matte, A. Barr, S. Mehta, C. D. Mazer, C. B. Guest, T. E. Stewart, F. Al-Saidi, A. B. Cooper, D. Cook, A. S. Slutsky, and M. S. Herridge. Two-year outcomes, health care use, and costs of survivors of acute respiratory distress syndrome. *American Journal of Respiratory and Critical Care Medicine*, 174:538–544, 2006.

[27] B. A. Christensen. Improving ICU patient flow through discrete-event simulation. Master's thesis, MIT Sloan School of Management, 2012.

[28] C. A. Chrusch, P. O. Kendiss, P. M. McMillan, D. E. Roberts, and P. R. Grayl. High occupancy increases the risk of early death or readmission after transfer from intensive care. *Critical Care Medicine*, 37(10):2753–2758, 2009.

[29] G. Clermont, D. C. Angus, S. M. DiRusso, M. Griffin, and Walter T W. T. Linde-Zwirble. Predicting hospital mortality for patients in the intensive care unit: A comparison of artificial neural networks with logistic regression models. *Critical Care Medicine*, 29(2):291–296, 2001.

[30] G. Clermont, V. Kaplan, R. Moreno, J. L. Vincent, W. T. Linde-Zwirble, B. V. Hout, and D. C. Angus. Dynamic microsimulation to model multiple outcomes in cohorts of critically ill patients. *Intensive Care Medicine*, 30:2237–2244, 2004.

[31] J. K. Cochran and A. Bharti. Stochastic bed balancing of an obstetrics hospital. *Health Care Management Science*, 9:31–45, 2006.

[32] C. M. Coopersmith, H. Wunsch, M. P. Pink, K. M. Olsen W. T. Linde-Zwirble, M. S. Sommers, K. J. Anand, D. C. Angus, and C. S. Deutschman. A comparison of critical

care research funding and the financial burden of critical illness in the United States. *Critical Care Medicine*, 40(4):1072–1079, 2012.

[33] A. X. Costa, S. A. Ridley, A. K. Shahani, P. R. Harper, V. De Senna, and M. S. Nielsen. Mathematical modeling and simulation for planning critical care capacity. *Anesthesia*, 58(4):320–327, 2003.

[34] C. R. Couto, T. M. G. Pedrosa, T. C. de Paula, and E. R. P. Pedroso. Risk factors for nosocomial infection in a neonatal intensive care unit. *Infection Control & Hospital Epidemiology*, 27(6):571–575, 2006.

[35] D. R. Cox. A use of complex probabilities in the theory of stochastic processes. *Mathematical Proceedings of the Cambridge Philosophical Society*, 51:313–319, 1955.

[36] K. Davis, K. Stermikis, D. Squires, and C. Scheon. Mirror, mirror on the wall, 2014 update: How the U.S. health care system compares internationally. Technical report, The Commonwealth Fund, 2014. Available from http://www.commonwealthfund.org/publications/fund-reports/2014/jun/mirror-mirror.

[37] A. M. de Bruin and R. Bekker. Time-dependent analysis for refused admissions in clinical wards. *Annals of Operations Research*, 178:45–65, 2010.

[38] A. M. de Bruin, R. Bekker, L. van Zanten, and G. M. Koole. Dimensioning hospital wards using the Erlang loss model. *Annals of Operations Research*, 178:23–43, 2010.

[39] D. P. de Farias and B. Van Roy. The linear programming approach to approximate dynamic programming. *Operations Research*, 51(6):850–865, 2003.

[40] D. P. de Farias and B. Van Roy. On constraint sampling for the linear programming approach to approximate dynamic programming. *Mathematics of Operations Research*, 29(3):462–478, 2004.

[41] B. T. Denton, O. Alagöz, A. Holder, and E. K. Lee. Medical decision making: Open research challenges. *IIE Transactions on Healthcare Systems Engineering*, 1(3):161–167, 2011.

[42] G. Dobson, H-H Lee, and E. Pinker. A model of ICU bumping. *Operations Research*, 58(6):1564–1576, 2010.

[43] G. S. Doig, K. J. Inman, W. J. Sibbald, C. M. Martin, and J. M. Robertson. Modeling mortality in the intensive care unit: Comparing the performance of a back-propagation, associative-learning neural network with multivariate logistic regression. In *Proceedings of the Annual Symposium on Computer Application in Medical Care*, page 361. American Medical Informatics Association, 1993.

[44] L. G. Donowitz, R. P. Wenzel, and J. W. Hoyt. High risk of hospital-acquired infection in the ICU patient. *Critical Care Medicine*, 10(6):355–357, 1982.

[45] M. C. dos Santos, M. M. Boniatt, C. S. Lincho, J. A. S. Pellegrini, J. Vidart, F. E. M. Rodrigues, and S. R. R. Vieira. Inflammatory and perfusion markers as risk factors and predictors of critically ill patient readmission. *Revista Brasileira de terapia intensiva*, 26(2):130–136, 2014.

[46] R. O. Duda, P. E. Hart, and D. G. Stork. Pattern classification and scene analysis part 1: Pattern classification. *Wiley Chichester*, 2000.

[47] R. Dybowski, V. Gant, P. Weller, and R. Chang. Prediction of outcome in critically ill patients using artificial neural network synthesised by genetic algorithm. *The Lancet*, 347(9009):1146–1150, 1996.

[48] M. Elliott, L. Worrall-Carter, and K. Page. Intensive care readmission: A contemporary review of the literature. *Intensive and Critical Care Nursing*, 30(1):121–137, 2014.

[49] M. Fackrell. Modelling healthcare systems with phase-type distributions. *Health Care Management Science*, 12(1):11, 2009.

[50] M. J. Faddy. Examples of fitting structured phase-type distributions. *Applied Stochastic Models in Business and Industry*, 10(4):247–255, 1994.

[51] M. J. Faddy and S. I. McClean. Analysing data on lengths of stay of hospital patients using phase-type distributions. *Applied Stochastic Models in Business and Industry*, 15(4):311–317, 1999.

[52] F. L. Ferreira, D. P. Bota, A. Bross, C. Melot, and J. L. Vincent. Serial evaluation of the SOFA score to predict outcome in critically ill patients. *The Journal of the American Medical Association*, 286:1754–1758, 2001.

[53] G. D. Finlay, M. J. Rothman, and R. A. Smith. Measuring the modified early warning score and the rothman index: Advantages of utilizing the electronic medical record in an early warning system. *Journal of Hospital Medicine*, 9:116–119, 2014.

[54] A. Freitas, T. Silva-Costa, F. Lopes, I. Garcia-Lema, A. Teixeira-Pinto, P. Brazdil, and A. Costa-Pereira. Factors influencing hospital high length of stay outliers. *BMC Health Services Research*, 12(1):265, 2012.

[55] B. E. Fries. Bibliography of operations research in health-care systems. *Operations Research*, 24(5):801–814, 1976.

[56] S. A. Frost, E. Alexandrou, T. Bogdanovski, Y. Salamonson, P. M. Davidson, M. J. Parr, and K. M. Hillman. Severity of illness and risk of readmission to intensive care: A meta-analysis. *Resuscitation*, 80(5):505–510, 2009.

[57] S. A. Frost, V. Tam, E. Alexandrou, L. Hunt, Y. Salamonson, P. M. Davidson, M. J. A. Parr, and K. M. Hillman. Readmission to intensive care: Development of a nomogram for individualising risk. *Critical Care and Resuscitation*, 12:656–662, 2010.

[58] D. Gallant and K. Lanning. Streamlining patient care processes through flexible room and equipment design. *Critical Care Nursing Quarterly*, 24(3):24–76, 2001.

[59] L. Garg, S. I. McClean, M. Barton, B. J. Meenan, and K. Fullerton. Intelligent patient management and resource planning for complex, heterogeneous, and stochastic healthcare systems. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 42(6):1332–1345, 2012.

[60] M. Ghassemi, L. A. Celi, and D. J. Stone. State of the art review: The data revolution in critical care. *Critical Care*, 19(1):118, 2015.

[61] C. Goldfrad and K. Rowan. Consequences of discharges from intensive care at night. *Lancet*, 355:1138–1142, 2000.

[62] F. Gorunescu, S. I. McClean, and P. H. Millard. A queueing model for bed-occupancy management and planning of hospitals. *Journal of the Operational Research Society*, 53:19–24, 2002.

[63] M. Gray. Value: Operations research and the new health care paradigm. *Operations Research for Health Care*, 1(1):20–21, 2012.

[64] L. V. Green. How many hospital beds? *Inquiry*, 39(4):400–412, 2002.

[65] L. V. Green. Queueing analysis in healthcare. In R. W. Hall, editor, *Patient Flow: Reducing Delay in Healthcare Delivery*, pages 281–307. Springer, 2006.

[66] J. D. Griffiths, M. Jones, M. S. Read, and J. Williams. A simulation model of ICU occupancy in a critical care unit. *Journal of Simulation*, 4(1):52–59, 2010.

[67] J. D. Griffiths, N. Price-Lloyd, M. Smithies, and J. Williams. A queueing model of activities in an intensive care unit. *IMA Journal of Management Mathematics*, 17:277–288, 2006.

[68] P. R. Hachesu, M. Ahmadi, S. Alizadeh, and F. Sadoughi. Use of data mining techniques to determine and predict length of stay of cardiac patients. *Healthcare Informatics Research*, 19(2):121–129, 2013.

[69] N. A. Halpern, S. M. Pastores, and R. J. Greenstein. Critical care medicine in the United States, 1985–2000: An analysis of bed numbers, use, and costs. *Critical Care Medicine*, 32:1254–1259, 2004.

[70] P. R. Harper and A. K. Shahani. Modeling for the planning and management of bed capacities in hospitals. *Journal of the Operational Research Society*, 53:11–18, 2002.

[71] T. Hastie, R. Tibshirani, and J. H. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, 2nd edition, 2009.

[72] O. V. Hein, J. Birnbaum, K. Wernecke, M. England, W. Konertz, and C. Spies. Prolonged intensive care unit stay in cardiac surgery: Risk factors and long-term-survival. *The Annals of Thoracic Surgery*, 81(3):880–885, 2006.

[73] M. S. Herridge, A. M. Cheung, C. M. Tansey, A. Matte-Martyn, N. Diaz-Granados, F. Al-Saidi, A. B. Cooper, C. B. Guest, C. D. Mazer, S. Mehta, T. E. Stewart, A. Barr, D. Cook, A. S. Slutsky, and Canadian Critical Care Trials Group. One-year outcomes in survivors of the acute respiratory distress syndrome. *New England Journal of Medicine*, 348:683–693, 2003.

[74] F. S. Hosein, D. J. Roberts, T. C. Turin, D. Zygun, W. A. Ghali, and H. T. Stelfox. A meta-analysis to derive literature-based benchmarks for readmission and hospital mortality after patient discharge from intensive care. *Critical Care*, 18(6):715, 2014.

[75] D. W. Hosmer, S. Lemeshow, and R. X. Sturdivant. *Applied Logistic Regression*. Wiley Series in Probability and Statistics. Wiley, 2013.

[76] W. Hu, C. W. Chan, and J. R. Zubizaretta. An examination of early transfers to the ICU based on a physiologic risk score. Working paper available from http://www.columbia.edu/~cc3179/EDIP_2015.pdf, 2016.

[77] Intensive Care Coordination Unit. Discharging a person from intensive care, 2017. Available from https://www.aci.health.nsw.gov.au/networks/intensive-care/patients-and-families/treatment-and-monitoring/discharging-a-person-from-intensive-care.

[78] G. N. Iyengar. Robust dynamic programming. *Mathematics of Operations Research*, 30(2):257–280, 2005.

[79] S. H. Jacobson, S. N. Hall, and J. R. Swisher. Discrete-event simulation of health care systems. In R. W. Hall, editor, *Patient Flow: Reducing Delay in Healthcare Delivery*, pages 211–252. Springer, 2006.

[80] Y. S. Jo, Y. J. Lee, J. S. Park, H. I. Yoon, J. H. Lee, C. T. Lee, and Y. J. Cho. Readmission to medical intensive care units: Risk factors and prediction. *Yonsei Medical Journal*, 56(2):543–552, 2015.

[81] A. Kaben, F. Correa, K. Reinhart, U. Settmacher, J. Gummert, R. Kalff, and Y. Sakr. Readmission to a surgical intensive care unit: Incidence, outcome and risk factors. *Critical Care*, 12:R123, 2008.

[82] S. Karlin and H. M. Taylor. *A Second Course in Stochastic Processes*. Academic Press, San Diego, 1981.

[83] G. Kasotakis, U. Schmidt, D. Perry, M. Grosse-Sundrup, J. Benjamin, C. Ryan, S. Tully, R. Hirschberg, K. Waak, G. Velmahos, E. Bittner, R. Zafonte, J. P. Cobb,

and M. Eikermann. The surgical intensive care unit optimal mobility score predicts mortality and length of stay. *Clinical Medicine*, 40:1122–1128, 2012.

[84] D. S. KC and C. Terwiesch. An econometric analysis of patient flows in the cardiac intensive care unit. *Manufacturing and Service Operations Management*, 14(1):50–65, 2012.

[85] S. P. Keehan, J. A. Poisal, G. A. Cuckler, A. M. Sisko, S. D. Smith, A. J. Madison, D. A. Stone, C. J. Wolfe, and J. M. Lizonitz. National health expenditure projections, 2015–25: Economy, prices, and aging expected to shape spending and enrollment. *Health Affairs*, 35(8):1522–1531, 2016.

[86] J. A. Kellum, F. E. Sileanu, R. Murugan, N. Lucko A. D. Shaw, and G. Clermont. Classifying AKI by urine output versus serum creatinine level. *Journal of the American Society of Nephrology*, 26(9):2231–2238, 2015.

[87] A. Khademi, D. Saure, A. J. Schaefer, S. Braithwaite, and M. S. Roberts. The price of nonabandonment: HIV in resource-limited settings. *Manufacturing and Service Operations Management*, 17(4):554–570, 2015.

[88] S. C. Kim, I. Horowitz, K. K. Young, and T. A. Buckley. Analysis of capacity management of the intensive care unit in a hospital. *European Journal of Operational Research*, 115:36–46, 1999.

[89] S. H. Kim, C. W. Chan, M. Olivares, and G. J. Escobar. ICU admission control: An empirical study of capacity allocation and its implication on patient outcomes. *Management Science*, 61(1):19–38, 2015.

[90] S. H. Kim, E. Pinker, J. Rimar, and E. Bradley. Refining workload measure in hospital units: From census to acuity-adjusted census in intensive care units. Working paper available from http://www-bcf.usc.edu/~songheek/index_files/RI_2016.pdf, 2017.

[91] W. A. Knaus, E. A. Draper, D. P. Wagner, and J. E. Zimmerman. APACHE II: A severity of disease classification system. *Critical Care Medicine*, 13(10):818–829, 1985.

[92] W. A. Knaus, E. A. Draper, D. P. Wagner, J. E. Zimmerman, M. Bergner, P. G. Bastos, C. A. Sirio, D. J. Murphy, T. Lotring, and A. Damiano. The APACHE III prognostic system. Risk prediction of hospital mortality for critically ill hospitalized adults. *Chest*, 100(6):1619–1636, 1991.

[93] A. A. Kramer, T. L. Higgins, and J. E. Zimmerman. The association between ICU readmission rate and patient outcomes. *Critical Care Medicine*, 41:24–33, 2013.

[94] A. A. Kramer, J. E. Zimmerman, and T. L. Higgins. Intensive care unit readmissions in U.S. hospitals: Patient characteristics, risk factors, and outcomes. *Critical Care Medicine*, 40:3–10, 2012.

[95] J. R. Le Gall, S. Lemeshow, and F. Saulnier. A new simplified acute physiology score (saps ii) based on a European/North American multicenter study. *Journal of the American Medical Association*, 270(24):2957–2963, 1993.

[96] L. R. Le Gall, P. Loirat, A. Alperovitch, P. Glaser, C. Granthil, D. Mathieu, P. Mercier, R. Thomas, and D. Villers. A simplified acute physiology score for icu patients. *Critical Care Medicine*, 12(11):975–977, 1984.

[97] C. P. Lee, G. M. Chertow, and S. A. Zenios. Optimal initiation and management of dialysis therapy. *Operations Research*, 56(6):1428–1449, 2008.

[98] R. S. Levin, E. T. Harley, J. C. Fackler, C. U. Lehmann, J. W. Custer, D. France, and S. L. Zeger. Real-time forecasting of pediatric intensive care unit length of stay using computerized provider orders. *Critical Care Medicine*, 40:3058–3064, 2012.

[99] K. V. Liang, F. E. Sileanu, G. Clermont, R. Murugan, F. Pike, P. M. Palevsky, and J. A. Kellum. Modality of RRT and recovery of kidney function after AKI in patients surviving to hospital discharge. *Clinical Journal of the American Society of Nephrology*, 11(1):30–38, 2016.

[100] G. Lim and E. K. Lee. *Optimization in Medicine and Biology*. Auerbach Publications, Boca Raton, FL, 2008.

[101] J. C. Lowery. Simulation of a hospital's surgical suite and critical care area. In J. J. Swain, D. Goldsman, R. C. Crain, and J. R. Wilson, editors, *Proceedings of the 1992 Winter Simulation Conference*, pages 1071–1078, Piscataway, NJ, 1992. Institute of Electrical and Electronics Engineers.

[102] J. C. Lowery. Multi-hospital validation of critical care simulation model. In G. W. Evans, M. Mollaghasemi, E. C. Russell, and W. E. Biles, editors, *Proceedings of the 1993 Winter Simulation Conference*, pages 1207–1215, Piscataway, NJ, 1993. Institute of Electrical and Electronics Engineers.

[103] A. Marazzi, F. Paccaud, C. Ruffieux, and C. Beguin. Fitting the distributions of length of stay by parametric models. *Medical Care*, 36(6):915–927, 1998.

[104] A. Marshall, C. Vasilakis, and E. El-Darzi. Length of stay-based patient flow models: Recent developments and future directions. *Health Care Management Science*, 8(3):213–220, 2005.

[105] A. H. Marshall and S. I. McClean. Conditional phase-type distributions for modelling patient length of stay in hospital. *International Transactions in Operational Research*, 10(6):565–576, 2003.

[106] A. H. Marshall, S. I. McClean, C. M. Shapcott, and P. H. Millard. Modelling patient duration of stay to facilitate resource management of geriatric hospitals. *Health Care Management Science*, 5(4):313–319, 2002.

[107] K. S. Mathews and E. F. Long. A conceptual framework for improving critical care patient flow and bed use. *Annals of the American Thoracic Society*, 12:886–894, 2015.

[108] M. Maxwell, M. Restrepo, S. G. Henderson, and H. Topaloglu. Approximate dynamic programming for ambulance redeployment. *INFORMS Journal on Computing*, 22(2):266–281, 2009.

[109] D. B. McLaughlin and J. M. Hays. *Healthcare Operations Management*. Health Administration Press, Chicago, IL, 2008.

[110] M. L. McManus, M. C. Long, A. Cooper, and E. Litvak. Queuing theory accurately models the need for critical care resources. *Anesthesiology*, 100:1271–1276, 2004.

[111] E. B. Milbrandt, A. Kersten, M. T. Rahim, T. T. Dremiszov, L. M. Cooper, D. C. Angus, and W. T. Linde-Zwirble. Growth of intensive care unit resource use and its estimated cost in Medicare. *Critical Care Medicine*, 36(9):2504–2510, 2008.

[112] P. H. Millard and S. I. McClean. *Modelling hospital resource use: A different approach to the planning and control of health care systems*. Royal Society of Medicine Press, 1994.

[113] L. Minne, A. Abu-Hanna, and E. de Jonge. Evaluation of SOFA-based models for predicting mortality in the ICU: A systematic review. *Critical Care*, 12:R161, 2008.

[114] B. A. Mobley, R. Leasure, and L. Davidson. Artificial neural network predictions of lengths of stay on a post-coronary care unit. *Heart & Lung: The Journal of Acute and Critical Care*, 24(3):251–256, 1995.

[115] M. F. Neuts. *Matrix-Geometric Solutions in Stochastic Models: An Algorithmic Approach*. Dover Publications, Inc., New York, 1981.

[116] A. Y. Ng and M. I. Jordan. On discriminative vs. generative classifiers: A comparison of logistic regression and naive bayes. In T. G. Dietterich, S. Becker, and Z. Ghahramani, editors, *Advances in Neural Information Processing Systems 14*, pages 841–848. MIT Press, 2002.

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

[118] A. Nimgaonkar, D. R. Karnad, S. Sudarshan, L. Ohno-Machado, and I. Kohane. Prediction of mortality in an Indian intensive care unit. *Intensive Care Medicine*, 30(2):248–253, 2004.

[119] I. Nouaouri, A. Samet, and H. Allaoui. Evidential data mining for length of stay (LOS) prediction problem. In *Automation Science and Engineering (CASE), 2015 IEEE International Conference on*, pages 1415–1420, 2015.

[120] A. W. Ong, L. A. Omert, D. Vido, B. M. Goodman, J. Protetch, A. Rodriguez, and E. Jeremitsky. Characteristics and outcomes of trauma patients with ICU lengths of stay 30 days and greater: A seven-year retrospective study. *Critical Care*, 13(5):R154, 2009.

[121] I. Ouanes, C. Schwebel, A. Francais, F. Philippart, A. Vesin, L. Soufir, C. Adrie, M. G. Orgeas, J. F. Timsit, and B. Misset. A model to predict short-term death or readmission after intensive care unit discharge. *Journal of Critical Care*, 27:422–429, 2012.

[122] Y. A. Ozcan. *Quantitative Methods in Health Care Management: Techniques and Applications*. Jossey-Bass, San Francisco, CA, 2005.

[123] P. M. Pardalos and H. E. Romeijn. *Handbook of Optimization in Medicine*. Springer-Verlag, New York, NY, 2008.

[124] R. Paterson, D. C. MacLeod, D. Thetford, A. Beattie, C. Graham, S. Lam, and D. Bell. Prediction of in-hospital mortality and length of stay using an early warning scoring system: Clinical audit. *Clinical Medicine*, 6:281–284, 2006.

[125] R. Payam, T. Lei, and L. Huan. Cross-validation. In L. Ling and Ö. M. Tamer, editors, *Encyclopedia of Database Systems*. John Wiley & Sons, Inc., Boston, MA, 2009.

[126] W. P. Pierskalla and D. J. Brailer. *Applications of Operations Research in Healthcare Delivery*. In: Handbooks in Operations Research and Management Science, 6: Operations Research and the Public Sector, S. M. Pollock, M. H. Rothkopf, A. Barnett (Eds.), Elsevier, Amsterdam, The Netherlands, 1994.

[127] G. L. Piper, L. J. Kaplan, A. A. Maung, F. Y. Lui, K. Barre, and K. A. Davis. Using the rothman index to predict early unplanned surgical intensive care unit readmissions. *The Journal of Trauma and Acute Care Surgery*, 77:78–82, 2014.

[128] W. B. Powell. *Approximate Dynamic Programming: Solving the Curses of Dimensionality*. Wiley-Interscience, Hoboken, New Jersey, 1st edition, 2007.

[129] M. L. Puterman. *Markov Decision Processes*. John Wiley & Sons, New York, 1994.

[130] A. Rais and A. Viana. Operations research in healthcare: A survey. *International Transactions in Operational research*, 18(1):1–31, 2011.

[131] M. Raksakietisak, P. Toomtong, P. Vorakitpokatorn, A. Sengleulur, and M. Sunjohndee. TISS 28 or SOFA: Good predicting factors for admission in a surgical intensive care unit longer than 24 hours. *Journal of the Medical Association of Thailand*, 89(5):648–56, 2006.

[132] V. J. Ribas, J. C. López, A. Ruiz-Sanmartín, J. C. Ruiz-Rodríguez, J. Rello, A. Wojdel, and A. Vellido. Severe sepsis mortality prediction with relevance vector machines.

In *Engineering in medicine and biology society, EMBC, 2011 annual international conference of the IEEE*, pages 100–103. IEEE, 2011.

[133] J. C. Ridge, S. K. Jones, M. S. Nielsen, and A. K. Shahani. Capacity planning for intensive care units. *European Journal of Operational Research*, 105:346–355, 1998.

[134] F. Rincon, S. A. Mayer, J. Rivolta, J. Stillman, B. A. Bernadette, S. V. Elkind, R. Marshall, and J. Y. chong. Impact of delayed transfer of critically ill stroke patients from the emergency department to the neuro-ICU. *Neurocritical Care*, 13(1):75–81, 2010.

[135] M. L. Rothman, S. C. Hedrick, K. A. Bulcroft, D. H. Hickam, and L. Z. Rubenstein. The validity of proxy-generated scores as measures of patient health status. *Medical Care*, 29(2):115–124, 1991.

[136] M. Rué, S. Quintana, M. Álvarez, and A. Artigas. Daily assessment of severity of illness and mortality prediction for individual patients. *Critical Care Medicine*, 29(1):45–50, 2001.

[137] J. A. Russell. She's out of the ICU now. That's a relief, isn't it?: The growing problem of ICU recidivism. *American Journal of Respiratory and Critical Care Medicine*, 185(9):906–908, 2012.

[138] S. Russell and P. Norvig. *Artifical Intelligence: A Modern Approach*. Prentice Hall, New Jersey, 2003.

[139] G. Saka, J. E. Kreke, A. J. Schaefer, C. C. Chang, M. S. Roberts, and D. C. Angus. Use of dynamic microsimulation to predict disease progression in patients with pneumonia-related sepsis. *Critical Care*, 11(3):R65, 2007.

[140] G. Schwarz. Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461–464, 1978.

[141] A. D. Sekulic, S. V. Trpkovic, A. P. Pavlovic, O. M. Marinkovic, and A. N. Ilic. Scoring systems in assessing survival of critically ill ICU patients. *Medical Science Monitor: International Medical Journal of Experimental and Clinical Research*, 21:2621, 2015.

[142] A. Sen, M. C. Keener, F. E. Sileanu, E. Foldes, G. Clermont, R. Murugan, and J. A. Kellum. Chloride content of fluids used for large-volume resuscitation is associated with reduced survival. *Critical Care Medicine*, 45(2):e146–e153, 2017.

[143] A. K. Shahani, S. A. Ridley, and M. S. Nielsen. Modeling patient flows as an aid to decision making for critical care capacities and organization. *Anesthesia*, 63(10):1074–1080, 2008.

[144] A. Shmueli, C. L. Sprung, and E. H. Kaplan. Optimizing admissions to an intensive care unit. *Health Care Management Science*, 6:131–136, 2003.

[145] L. J. Shuman, R. D. Speas, and J. P. Young, editors. *Operations Research in Health Care*. The Johns Hopkins Univeristy Press, Baltimore, MD, 1975.

[146] F. E. Sileanu, R. Murugan, N. Lucko, G. Clermont, S. L. Kane-Gill, S. M. Handler, and J. A. Kellum. AKI in low-risk versus high-risk patients in intensive care. *Clinical Journal of the American Society of Nephrology*, 10(2):187–196, 2015.

[147] L. Straney, A. Clements, J. Alexander, and A. Slateri. A two-compartment mixed-effects gamma regression model for quantifying between-unit variability in length of stay among children admitted to intensive care. *Health Services Research*, 47(6):2190–2203, 2012.

[148] I. W. M. Verburg, A. A. Atashi, S. Eslami, R. Holman, A. Abu-Hanna, E. de Jonge, N. Peek, and N. F. de Keizer. Which models can I use to predict adult ICU length of stay? A systematic review. *Critical Care Medicine*, 45(2):e222–e231, 2017.

[149] J. L. Vincent. Critical care - where have we been and where are we going? *Critical Care*, 17(1):S2, 2013.

[150] J. L. Vincent, R. Moreno, J. Takala, S. Willatts, A. De Mendona, H. Bruining, C. K. Reinhart, P. M. Suter, and L. G. Thijs. The SOFA (sepsis-related organ failure assessment) score to describe organ dysfunction/failure. *Intensive Care Medicine*, 22(7):707–710, 1996.

[151] J. L. Vincent, J. Rello, J. Marshall, E. Silva, A. Anzueto, C. D. Martin, R. Moreno, J. Lipman, C. Gomersall, Y. Sakr, and K. Reinhart. International study of the prevalence and outcomes of infection in intensive care units. *The Journal of the American Medical Association*, 302(21):2323–2329, 2009.

[152] J. L. Vincent and M. Singer. Critical care: Advances and future perspectives. *Lancet*, 376:1354–1361, 2010.

[153] J. Wagner, N. B. Gabler, S. J. Ratcliffe, S. S. Brown, B. L. Strom, and S. D. Halpern. Outcomes among patients discharged from busy intensive care units. *Annals of Internal Medicine*, 159(7):447–455, 2013.

[154] C. C. White and H. K. Eldieb. Markov decision processes with imprecise transition probabilities. *Operations Research*, 43:739–749, 1994.

[155] S. V. Williams. How many critical care beds are enough? *Critical Care Medicine*, 11(6):412–416, 1983.

[156] A. L. Woldhek, S. Rijkenberg, R. J. Bosman, and P. H. J. van der Voort. Readmission of ICU patients: A quality indicator? *Journal of Critical Care*, 38:328–334, 2017.

[157] E. G. Wong, A. M. Parker, D. G. Leung, E. P. Brigham, and A. I. Arbaje. Association of severity of illness and intensive care unit readmission: A systematic review. *Heart & Lung: The Journal of Acute and Critical Care*, 45(1):3–9.e2, 2016.

[158] D. Zhang and D. Adelman. An approximate dynamic programming approach to network revenue management with customer choice. *Transportation Science*, 43(3):381–394, 2009.

[159] J. E. Zimmerman, A. A. Kramer, D. S. McNair, F. M. Malila, and V. L. Shaffer. Intensive care unit length of stay: Benchmarking based on acute physiology and chronic health evaluation (APACHE) IV. *Critical Care Medicine*, 34:2517–2529, 2006.