4-20-2022 10:15 AM

Statistical Applications to the Management of Intensive Care and Step-down Units

Yawo Mamoua Kobara, *The University of Western Ontario*

Supervisor: David Andrew Stanford, *The University of Western Ontario*
Co-Supervisor: Camila de Souza, *The University of Western Ontario*
Co-Supervisor: Felipe Rodrigues, *King's University*
A thesis submitted in partial fulfillment of the requirements for the Doctor of Philosophy degree in Statistics and Actuarial Sciences

Abstract

This thesis proposes three contributing manuscripts related to patient flow management, server decision-making, and ventilation time in the intensive care and step-down units system.

First, a Markov decision process (MDP) model with a Monte Carlo simulation was performed to compare two patient flow policies: prioritizing premature step-down and prioritizing rejection of patients when the intensive care unit is congested. The optimal decisions were obtained under the two strategies. The simulation results based on these optimal decisions show that a premature step-down strategy contributes to higher congestion downstream. Counter-intuitively, premature step-down should be discouraged, and patient rejection or divergence actions should be further explored as a viable alternative for congested intensive care units (ICUs).

Secondly, an investigation of the length of stay (LOS) competition between the intensive care unit (ICU) and the step-down unit (SDU), two servers in tandem without a buffer between them was proposed using queuing games. Analysis of the competition was done under four different scenarios: (i) both servers cooperate; (ii) the servers do not cooperate and make decisions simultaneously; (iii) the servers do not cooperate but the first server, the ICU is the leader; (iv) the servers do not cooperate, the second server the SDU is the leader. Finally, a numerical analysis was performed. The results show that the length of stay decisions of each server depends critically on the payoff function's form and the exogenous demand. Secondly, with a linear payoff function, the SDU is only beneficial to the system if the unit cost is greater than its unit reward at the ICU. Perhaps most importantly, the critical care pathway performs better under coordination and or leadership at the ICU level.

Finally, first-day ventilated patients' ventilation time was analyzed using survival analysis. The probabilistic behaviour of the ventilation time duration was analyzed and the predictors of the ventilation time duration were determined based on available first-day covariates. Data were obtained from the Critical Care Information System (CCIS) about patients admitted to the ICUs in Ontario between July 2015 and December 2016. The log-logistic AFT model was found to be the best to relate the association between first-day covariates and the ventilation time.

Keywords: ICU/SDU, healthcare, patient flow, congestion, ventilation, Markov decision process, queuing games, survival analysis.

Lay Summary

In this thesis, I used statistics to address certain ICU-SDU server decision-making and ventilation time in the intensive care and step-down unit system.

First, when the critical care unit was overcrowded, a Markov decision process (MDP) model with Monte Carlo simulation was utilised to evaluate two patient flow strategies: prioritising premature step-down and prioritising patient rejection. Under the two techniques, the best decisions were made. The simulation findings based on these optimum judgments reveal that a premature step-down method leads to increased downstream congestion. Premature step-down should be avoided, and patient rejection or divergence measures should be investigated further as a possible solution for overcrowded intensive care units (ICUs).

Second, utilising queuing games, an analysis of the length of stay (LOS) rivalry between the intensive care unit (ICU) and the step-down unit (SDU) was proposed. Four scenarios were used to analyse the competition: I both servers collaborate; (ii) both servers cooperate but the first server, the ICU, is the leader; (iv) both servers cooperate but the second server, the SDU, is the leader. After then, there was a numerical analysis. The findings reveal that the payout function's shape and exogenous demand have a significant impact on each server's length-of-stay decisions. The SDU, on the other hand, has a linear payout function.

Finally, survival analysis was used to look at the ventilation time of first-day ventilated patients. Based on available first-day factors, the probabilistic behaviour of ventilation time duration was studied, and predictors of ventilation time duration were identified. Patients hospitalised to ICUs in Ontario between July 2015 and December 2016 were studied using data from the Critical Care Information System (CCIS). The best model for relating the connection between first-day variables and ventilation time was determined to be the log-logistic AFT model.

Co-Authorship Statement

I hereby declare that this thesis incorporates materials that are the results of joint research. As the main author, I was responsible for all the direct aspects of the research, including formulation of research questions, literature review, model formulation, analysis, coding, and preparing the first and final versions of the manuscripts.

Chapters 3 and 4 of the thesis are co-authored with Drs. David Andrew Stanford, Felipe Fontes Rodrigues, and Camila de Souza. In all cases, only my primary contributions towards these publications are included in this thesis. Dr. Felipe Fontes Rodrigues presented the research problems, provided the data, and contributed to the development of the methodology. He guided the analyses and provided suggestions and feedback for improving the text. Dr. David Andrew Stanford suggested the methodology, brought insights, and provided ideas to conduct the analysis. Dr. Camila de Souza aided in developing the methodology, and contributed to reviewing, improving, and giving feedback on the text. Chapter 3 has been submitted to the Journal of Operation Research. Chapter 4 has been submitted to the Journal Operation Research for Health Care for peer-review.

Chapter 5 of the thesis is co-authored with Drs. Felipe Fontes Rodrigues and Camila de Souza and Miss Megan Wismer. Dr. Felipe Fontes Rodrigues presented the research problem, the data, and contributed to the development of the methodology. Dr. Camila de Souza aided in developing the methodology, and contributed to reviewing, improving, and providing feedback on the analyses and the manuscript. Miss Megan Wismer collaborated with the literature reviews and feedback on the text. Chapter 5 has been submitted to the Journal of Operation Research for Health Care for peer-review.

To God be the glory.

"The Lord did not set his love upon me, nor chose me,
because I was more in number than any people;
for I was the fewest of all people"

Deuteronomy 7:7

À ma Grand-maman,
Je me souviens de ta foi sincère, de ton espoir
et de ton pur amour au quotidien!

À Maman et a Papa,
les meilleurs parents que l'on puisse demander,
merci d'être!

Acknowledgements

With profound gratitude, I extend a special thank you once more to all my supervisors. Your great help and support, your motivation and guidance, and your constructive feedback and corrections have constantly helped me improve myself and my work.

Dr. David Andrew Stanford, thank you for your patience and sincere interest in my development and for introducing me to Queuing Theory. Your mentorship, openness, and constructive interactions have had a significant influence on my understanding of the scope of academic research work. Thank you for the opportunities and support in attending conferences to present my work. I remembered you often saying "Yawo, we should make you an expert Markov decision planner." It is unfortunate your illness took the best part of you.

Dr. Felipe F. Rodrigues, thank you, not only for your exceptional help on this thesis but also for introducing me to Game and Contract Theory granting me an understanding of the social background to mathematical fields. Thank you for your encouragement. Your passion and perspective of the health care system are contagious.

Dr. Camila Pedroso Estevam de Souza, thank you, I appreciate your compromises, availability, and suggestions you made. Your co-supervision helped me complete my research.

Another thank you to the professors in the Department of Statistical and Actuarial Sciences who provided instruction, encouragement, guidance, and kindness in diverse ways. I appreciate the timely and effective advice, support, and encouragement from Dr. Reg Kulperger, Dr. Douglas Woolford, Dr. Marcos Escobar-Anel, Dr. Xioming Liu, and Dr. Kristina Sendova.

Finally, I am grateful to my parents, siblings, cousins, friends, and fellowship brethren for their never-ceasing encouragement, and moral and spiritual support. Thank you and God bless.

Contents

List of Figures

List of Tables

List of Appendices

List of Abbreviations

AI	Artificial Intelligence
ALC	Alternative Level of Care
AR	Admission Reward
CDF	Cumulative distribution function
CIHI	Canadian Institute for Health Information
CP	Cooperation game
CTAS	Canadian Triage and Acuity Scale
DES	Discrete Event Simulation
DTMC	Discrete Time Markov Chain
ED	Emergency Department
FCFS	First come first serve
GLM	Generalized linear model
HAP	High Acuity Patient
ICU	Intensive Care Unite
IMV	Invasive Mechanical Ventilation
IS	ICU Stackelberg game
KM	Kaplan-Meier
KPI	Key Performance Indicators
LAP	Low Acuity Patient
LHSC	London Health Sciences Centre
LOS	Length of Stay
MODS	Multi-organ Dysfunction Syndrome Score
MPD	Markov Decision Process
MSE	Mean Square Error
NEMS	Nine Equivalents of Nursing Manpower Score
PDF	Probability density function
PH	Proportional Hazard
RC	Rejection Cost
SDU	Step-down Unit
SR	Step-down Reward
SS	SDU Stackelberg game
ST	Simultaneous Decision game

Chapter 1

Introduction

1.1 Introduction and Motivation

Hospital units that face high demand and pressure include the intensive care unit (ICU) and the step-down unit (SDU). The ICU is an integral part of the modern healthcare system with many variant designations (e.g., Intensive Therapy Unit and Critical Care Unit). It is a special unit of hospitals that provides specialized treatment for critically ill patients. ICUs cater to patients with severe or life-threatening illnesses and injuries, which require constant care, close supervision from life support equipment and medication to ensure normal bodily functions. Admission in the ICU can either be a planned admission as part of recovery after surgery or an emergency measure if there are complications during surgery or an emergency after an accident. The time it takes to recover varies from person to person and depends on various medical and clinical factors. The ICU beds and nurses are the most expensive ones in a hospital regarding quality, training, and specialization. Therefore, there is a limited resource [65, 170, 142]. So an unnecessary use through overstays of this resource is detrimental to the hospital's effectiveness and efficiency especially with the high demand experienced in the ICU. Unfortunately, the ICU overstay phenomenon has not been vastly explored. The Canadian Institute for Health Information (CIHI) in 2016 found a noticeable increase in lack of capacity in the ICU due to increasing demand [74]. Commonly, the ICU may be over its capacity. At this moment, the service quality indicators, such as the wait times, length-of-stay, condition to discharge to the SDU, and the quality of care are of growing concern to the management and stakeholders patients and their families. During the COVID-19 pandemic, the ICU's congestion became more evident with more than 9000 Ontarians admitted to (ICUs) with COVID-19 related critical illness and the number of patients on ventilators was over 180% of pre-pandemic historical averages on the peak day ([104]).

Health care units in general and ICU in particular are long known for their congestion

problems [50, 7, 8]. Moreover, factors such as advancement in health care technologies and an aging population have increased the demand, the complexity of service requirements, and contributed to the accentuation of the problem [14, 171, 13, 132]. In the ICU, congestion and delay worsen patients' severity, increase morbidity and mortality, and so there is a constant need for improved patient flow management [100]. Decision-making in the healthcare system is of primary importance for patients' well-being and safety. Fast access to high quality health care is one of the recurrent indicators of quality of life globally [44]. Many services are performed by several consecutive queues in tandem, and due to the slow process at one end, the whole system becomes tense.

ICU beds are the most expensive service at a hospital and require well-trained and specialized nurses and physicians [65, 170, 142]. As patients' acuity lessens in the ICU, they no longer require an ICU bed and high nurse monitoring (one nurse per patient). One solution to reduce ICU congestion was to create the SDU as an intermediary level of care to alleviate the ICU burden and cost. But instead of a reduction in congestion, with time, the ICU/SDU system became as congested as before. In addition, since its introduction, there has been relatively little research on its benefits and, medically, a lot of debate about its usefulness [8].

The SDU also known as High Dependency Care Unit or Transitional Care Unit or Progressive Care Unit or Level 2 Care Unit, is an intermediate care unit between the ICU and the general ward. The SDU is designed to care for critically ill patients recovering in the ICU yet require higher and continuous monitoring than the patients in the ward. Gotsman and Schrire [61] introduced the concept of SDU in 1968. They proposed a patient-care area with specialized monitoring and nursing care for cardiac patients who no longer require full intensive care but are not ready for discharge to a regular ward [61]. The idea was that patients who can breathe unaided by ventilation equipment for example, would be transferred to an SDU to continue their recovery. Prin and Wunsch [129] gave an extensive review and utility of the SDU units in the hospital. The creation of the SDU was to improve the efficiency of patient flow and reduce cost. But it has been observed and reported without ample support by hospital managements that, in the ICU/SDU system, readmission, shortage of bed capacity, limited health care resources and the downstream congestion in the case of tandem queue services contribute to overstay in the ICU [100, 12]. Overstaying the system when the system is in high demand is critically damaging and has a crucially high repercussion on the arriving patients that demands the use of the ICU. The average ICU occupancy of teaching and large hospitals in Canada was estimated to be 90% in 2009 [74] and is increasing. Such a highly utilized unit could benefit from the smallest amelioration of its management. Due to the COVID-19, currently, the ICU occupancy is above 100% in Ontario [45]. Studies have shown that bottleneck congestion and crowded hospitals cause higher patient mortality rates and physician error in many hospital

units [14, 83, 33].

Overstaying the system when the system is in high demand is critically damaging and has a crucially high repercussion on the arriving patients demanding ICU use. Overstay in a hospital unit is defined as the stay in the unit that exceeds the discharged time. Overstay in the emergency department (ED) is referred to as "boarding" and is a well-researched situation [32, 108]. Many researchers have identified downstream congestion as the main source of overstay. This assertion has not been proved empirically by available data from the SDU. We postulate that other factors such as patient characteristics or external characteristics might influence their overstay. In the ICU, decisions are often made in the face of uncertain and incomplete information, pressure from managers, patients, their families, time, and capacity constraints. These are conditioned by a shortage of bed capacity, limited health care resources, an additional burden to the system due to staff's stress.

In the literature, guidelines addressing decisions on patient flow in a congested ICU/SDU system are varied and subjective [117, 116, 85, 151, 76]. In practice, hospitals have resorted to premature discharge, which is done by moving current inpatient from ICU to SDU to admit an arriving patient. Secondly, who should make patient flow decisions, the ICU or the SDU? Should the decision be centralized or decentralized? Such questions have yet to be answered in the literature. Consequently, in many hospitals where patients are to be served by different units, overstay and off-service have become a recurrent problem for the management. And the well-being of patients in many health care systems is at risk. Off services occur when the hospital admits patients to units that belong to different services because the required units are congested. Off service is part of the practices that hinder patients from receiving the proper care.

In the ICU, life and death decisions are made or received daily. These decisions are often made in the face of uncertain and incomplete information, pressure from managers, patients, their families, time, and capacity constraints. The problems associated with the ICU/SDU system are of two aspects; resource and patient management and service performance. ICU resources are scarce. First, the ICU capacity is limited; the ICU deals with a shortage of beds and nurses. The fragility of patients and the severity of the illness in the ICU create the need for a specially trained nurse for a 1:1 or at most a 1:2 nurse to patient ratio. Thus, the provision of more ICU beds would necessitate the employment of more specially trained clinicians and nurses.

Due to the high operational cost, increasing ICU demand, and the congestion of the other hospital units, the problem of admission and discharge (i.e., patients flow in general) guidelines surfaced. Most guidelines in the literature addressing decisions on patient flow in the ICUs have had significant limitations, and this literature does not yet provide a consistent view

of a mathematically generated model to use [117, 116, 151, 76, 85]. ICUs flow is a complex and intricate process that depends on multiple characteristics. For the most ICUs, the patients' flow decision is taken by the intensive care specialist. These decisions are based on the health condition (severity and status), the therapeutic proneness, the physiological state, and the patients' wishes [117]. But for a scarce, costly, sensitive, and highly demanded service like the ICU, the flow decisions must involve a trade-off balancing the importance of an individual against the overall benefit of the system to society. The incomplete data available and the consistent answer to the question of whether patients would obtain better outcomes if they had been admitted earlier into the ICU is of interest. An intuitive solution to this problem in a social system would be providing more resources in terms of beds and personnel. However, the possibility of offering more critical care beds and employing more nursing staff is challenging and must be weighed up against the relative benefit. VanBerkel and Blake [165] found that increasing capacity alone is not enough to stabilize patient flows, but faster services are also necessary. What about increasing capacity without the complementary provision of qualified staff needed? We know that congested environments trigger negative responses in the patient [47, 5]. Problems involving patient flow through discharge and admission management are increasing due to the increasing demands resulting from a growing and aging population [63] or unexpected circumstances such as the COVID-19 pandemic [147]. In addition to the question of the role of the SDU in a congested environment [129], the question of its geographic position and leadership between the two units also emerged. Should the SDU be under the leadership of the ICU management, that is, the two units having a central administration or not. The literature does not provide information on this. To open the discussion, we will provide a competition of a patient length-of-stay between the two units and study the difference between the centralized and decentralized settings to optimize critical indicators.

Several research papers have developed statistical and mathematical models that showed and reviewed the clinical, medical, financial, and economic importance and impact of the SDU [129, 94, 101, 50, 35]. However, the problems tackled by these articles are discussions about planning within the hospital. Secondly, these studies are limited to forecasting the likelihood of mortality, rejection, capacity for the ICU. The last bed problem in the ICU/SDU system due to the downstream congestion in the SDU and interdependence of ICU, SDU, and other hospital units are often overlooked. Thus, this thesis investigates and compares two patient flow decision strategies in a congested environment. This adds options to decision-makers to consider in case the demand rate to the ICU is increasing and there is a capacity shortage. For example, during the COVID pandemic, the ICU capacities are constrained, yet many more people are requesting the service of the ICU.

A defining characteristic of intensive care is life support and dedicated nursing staff. Per-

haps the most essential organ support machine in the ICU is the invasive mechanical ventilator. Invasive ventilation is used for patients without any strength and ability to breathe independently. The ventilator takes over the breathing and enables the body to receive oxygen and recover. The duration of these supportive therapies is clinically relevant to outcomes, especially during the Covid-19 pandemic. From the statistical perspective, these quantities are challenging to estimate due to episodes being time-dependent and potentially multiple and being influenced by the competing, terminal events of discharge alive and death.

In this thesis, we present three research projects. In the first, a simulation of the Markov decision model to investigate decisions with or without premature ICU step-downs. Decisions need to take place in a sequence and thus need to be planned ahead. The benefit or drawback of an action may not be immediately evident to a myopic decision maker due to the randomness of the future state of the system; nonetheless, may help achieve a higher future payoff (or a lower future cost). Therefore, MDPs are a principled methodology for this kind of problem [89]. The second project is a queuing game model between two servers (ICU and SDU) in tandem. The third project focus on applying survival analysis methods to characterize invasive ventilators' use time at admission in ICUs, considering various clinical covariates.

1.2 Patients Flow through the ICU/ SDU system

The hospital studied uses the patient/nurse ratio as a proxy for patient readiness to be moved to a lower level of care. As part of their routine, every patient is scored daily on a 56-point scale known as "Nine equivalents of nursing manpower use score" or "NEMS". The NEMS is closely related to patient health because as the patient's health improves, less nursing attention is needed, resulting in a lower NEMS. Empirically, a score below ten is considered to be a recovering patient (RP); scores between 11–25 are low acuity patients (LAP), and from those more than 25 as high acuity patients (HAP). HAPs arrive at the ICU individually from different sources. The emergency department (ED), a unit with varying patient severity, provides the highest proportion of ICU patients. From our data sample, about 38% of the patients come from the ED, 22% from the ward, 21% from the Operating room, and 20% from other places such as other hospitals or the SDU. 99.1% of the admission into the ICU are unplanned, and patients require immediate medical care. With the priority triage policy used in many hospitals, they have little, or no control over admitting HAP arriving through the emergency route. Daily arrivals are essentially equally distributed, with Thursdays having the maximal admission. Hourly admission trends are also examined. Most admissions of patients fall in the afternoon or evening, between 4 pm and 8 pm. Once an available ICU bed is assigned, "service time" commences, during which patients are stabilized and receive active management of their

critical illness. Service terminates if a patient dies, or survives and is deemed clinically ready for a lower level of care and staff request transfer from the ICU to the downstream unit, known as a "booking." Physicians do not preemptively request transfer before the patient is clinically ready per hospital policy. While waiting for transfer, patients experience ICU "boarding" when they physically remain in the ICU bed but no longer receive high-intensity services. If a patient clinically deteriorates and requires re-initiation of critical care-level interventions, service time commences again until the physician requests a new booking time in the future; we only consider these final booking times in our analysis. Patients exit the ICU upon physical transfer to the SDU or ward, where they complete their stay until hospital discharge or bounce back into the ICU if they require critical care services again before discharge.

1.3 Thesis Organization

This thesis presents three projects modelling the decision-making in the ICU/SDU system under congestion and the time patients spend on invasive mechanical ventilation. It consists of six chapters with content summarized as follows.

Chapter 1 introduces the research background and the system considered. In addition, motivations, objectives, and the outline of the thesis are described.

Chapter 2 provides an overview of the mathematical and statistical background of the projects presented.

In Chapter 3, Project 1 entitled 'Patient Flow in Congested Intensive Care Unit /Step-down Unit system: To Premature Step-down or not?" is developed. A Markov decision process (MDP) model is constructed to model the flow process in the ICU and SDU systems.

In Chapter 4, Project 2 entitled "Intensive Care Unit / Step-Down Unit Queuing Game with Service Time Decisions" constructs a queuing game between the ICU and SDU. The equilibrium characteristics of the system are determined with each of the units seeking its reward. Given that there are numerous questions about the configuration of the ICU/SDU, whether it should be under a single leadership or different leaderships, we consider leadership in decision making on the length-of-stay (LOS) between the ICU and SDU.

Chapter 5 presents Project 3 entitled "Invasive Mechanical Ventilation Duration Prediction using Survival Analysis" where a survival analysis of the ventilation time of first-day ventilated patients is performed to characterize the ventilation time, determine predictors of the ventilation time, and predict ventilation time of each patient.

Finally, in Chapter 6, I concluded the thesis and itemized its main contribution.

Chapter 2

Background

In Section 2.1, a brief overview of optimization methods is presented. Section 2.2 gives an overview of the Markov decision process modelling. In Section 2.3, game theory is discussed. Section 2.4 presents some basic queuing theory results, and Section 2.6 introduces Survival models.

2.1 Optimization

In order to support human decision-making performance, statistical, mathematical, and computational methods are needed. Operational research (OR) in general, plays an important role in finding optimal solutions to problems in many parts of our lives. Optimization methods have been the backbone of OR. All statistical criteria in one way or the other, are, in its essence a well-formulated optimization problem.

Optimization is also referred to as mathematical optimization or mathematical programming [102]. Merriam-Webster dictionary (https://www.merriam-webster.com/dictionary/optimization) defines optimization as a process or methodology of making something as fully perfect, functional, or effective as possible according to previously set objectives. Mathematical optimization seeks to select the element to obtain the maximum or the minimum of some function relative to some set. For complex allocation and decision problems, optimization methods have been the underlying analysis principle since it offers a level of conciseness and clear operational simplicity indispensable. Optimization problems can be classified to be either constrained or unconstrained optimization. For example, equality constraints problems can be converted into unconstrained problems using the method of Lagrange multipliers. Constrained optimization models can be classified into linear and nonlinear programming models. Linear programming is the simplest. The general formulation of a constraint optimization problem can be stated as follows [102]:

$$\begin{aligned} &\min_x f(x) \\ &\text{subject to} \\ &\qquad h_i(x) = 0, \quad \text{for each } i = 1, 2, \ldots, m \\ &\qquad g_j(x) \leq 0, \quad \text{for each } j = 1, 2, \ldots, p \\ &\qquad x \in \mathcal{S} \end{aligned} \tag{2.1}$$

where $x = (x_1, x_2, \ldots, x_n)$ is a n-dimensional vector of unknowns and f, h_i, $i = 1, 2, \ldots, m$ and g_j, $j = 1, 2, \ldots, p$, are real-valued functions of the decision variables $x_1, x_2, ..., x_n$. The set $\mathcal{S}$ is a subset of the n-dimensional space of restriction on the decision variables. The function $f(x)$ is the objective function that needs to be optimized. $h_j(x) = 0$ is the m-equality constraints and $g_j(x)$ is the p-inequalities constraints. If we maximize the objective function, the inequality constraints will change from less or equal to greater or equal. Generally, additional assumptions are introduced based on the real-life situation problems we are trying to solve. The unconstrained optimization has no other conditions attached to it. It is generally formulated as:

$$\min_x or \max_x f(x) \tag{2.2}$$

There are other ways to classify optimization problems. For example, optimization problems may be classified based on the structure of the problem (i.e. deterministic or stochastic), the number of the objective functions (single or multiple objective functions), type of decision variable (integer, continuous or mixed), the constraint structure (linear, non-linear, other), and based on the optimization structure (convex, non-convex or quasi-convex). Figure 2.1 shows the classification of optimization models.

2.1.1 Linear Optimization Problems

Linear optimization problems are problems in which the objective function and the constraints are all linear functions of the decision variables [102, 22]. In a matrix-vector notation, it is formulated as 2.3:

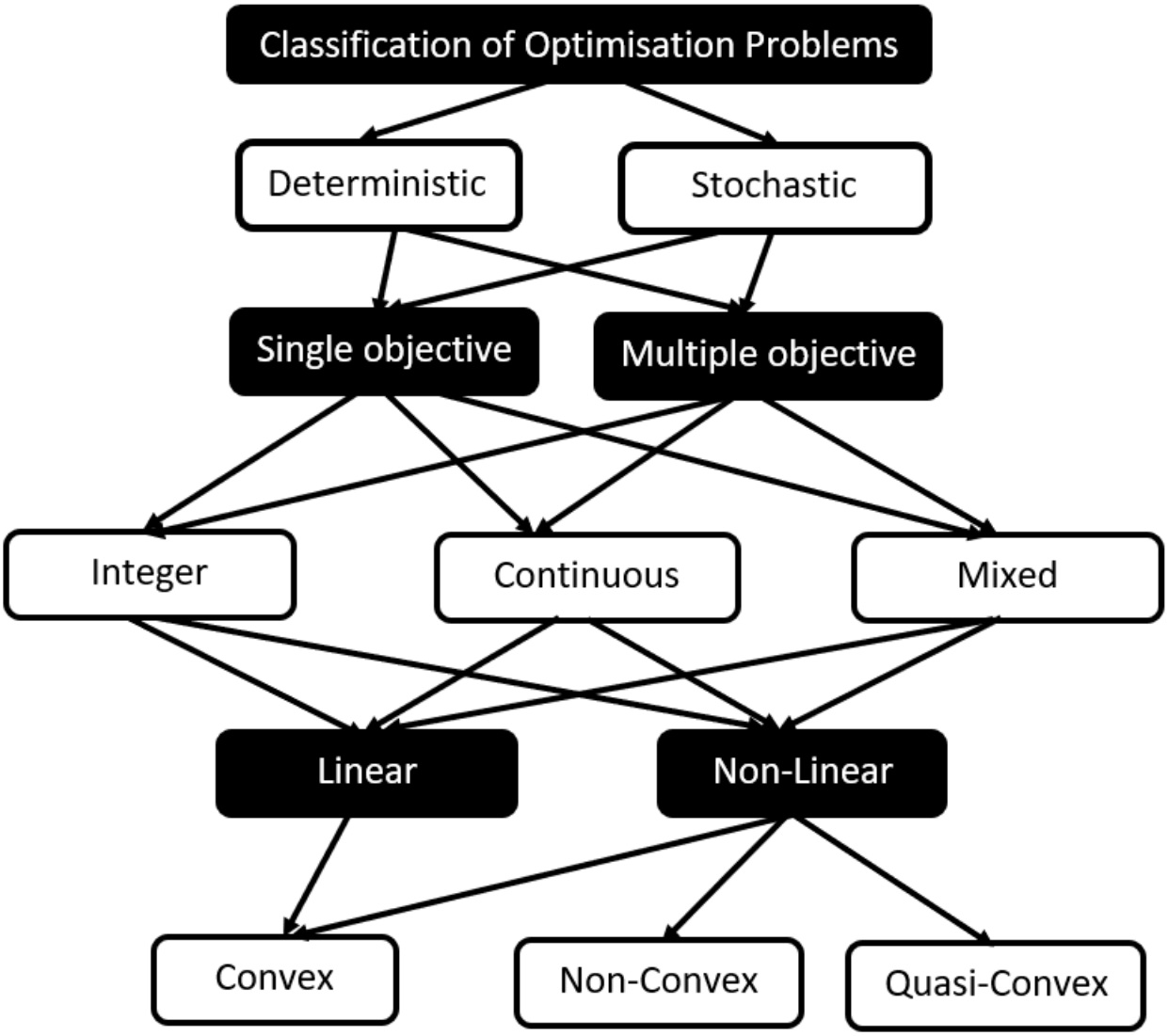


Figure 2.1: Classification of optimization problems *(formulated by the author)*.

$$\begin{aligned} &\min_{x} \quad c^T x \\ &\text{subject to} \\ &\qquad Ax = r \\ &\qquad Bx \leq 0 \text{ and } x \geq h \end{aligned} \tag{2.3}$$

where x is a n-dimensional column vector, c^T is a n-dimensional row vector, A is a $m \times n$ matrix, r is a m-dimensional column vector, B is a $p \times n$ matrix, and h is an n-dimensional column vector.

Linear optimization problems with continuous decision variables are termed linear programming (LP). Linear programming is the most used method in formulating the most naturally occurring problems. For example, Luenberger and Ye [102] and Boyd et al. [22] observed that many constraints and objective functions that arise in practice are linear. Moreover, the formulation is done with a modest effort since the mathematics is tidier, the model is precise and concise, the theory is richer, and the computation simpler compared to the non-linear ones [102].

The simplex method is usually used for manual computations. The simplex method in minimization is to start with a basic feasible solution of the constraint set of a problem and proceed to another, in such a way as to continually decrease the value of the objective function until a minimum is reached [102]. There are elegant, effective and efficient algorithms available in many languages to compute and obtain the solution for more extensive and complex problems [102, 22].

2.1.2 Non-linear Optimization problems

In linear optimization problems, we have seen that the objective function and all the constraints are linear functions of the decision variables. However, at other times, the objective functions and constraints may be non-linear. Those cases correspond to non-linear optimization problems, and intrinsically, these problems are harder to solve [23, 22, 102]. The form of the problem remains the same as in Equation 2.1, but the type of constraints and objective function are non-linear.

Among the types of non-linear functions, we can list those with polynomial objective functions, and exponential objective functions with linear constraints. An example of a polynomial objective function is the quadratic objective function. Solving non-linear optimization models

is challenging. The type of non-linearity determines the sufficiency and necessity for a solution. Convex objective functions tend to be the easiest to solve. Convex or concave objective function with convex or concave constraint set problem is solved with methods known as convex optimization. For quadratic objective function and linear constraints, quadratic optimization methods are used. A non-linear problem that satisfies the Karush–Kuhn–Tucker (KKT) conditions, has an optimal solution [23, 22, 138]. And convexity is a sufficient condition. There have been two ways to solve optimization problems: classical and computational. Classical methods use analytical means that solve differentiable functions. This method is used when the underlying conditions are fulfilled. Computational methods use computer algorithms that are designed for high-dimensional search. The following example taken from [23] illustrates how nonlinear programs can arise in practice.

Portfolio Selection Example

An investor has \$5000 and two potential investments. Let x_j for $j = 1$ and $j = 2$ denote his allocation to investment j in thousands of dollars. From historical data, investments 1 and 2 have an expected annual return of 20 and 16 percent, respectively. Also, the total risk involved with investments 1 and 2, as measured by the variance of total return, is given by $2x_1^2 + x_2^2 + (x_1 + x_2)^2$, so that risk increases with total investment and with the amount of each investment. The investor would like to maximize his expected return and at the same time minimize his risk.

$$\begin{aligned} \max_x f(x) =& 20x_1 + 16x_2 - \theta(2x_1^2 + x_2^2 + (x_1 + x_2)^2) \\ \text{s. t.} \quad & g_1(x) = x_1 + x_2 \leq 5, \\ & x_1 \geq 0, x_2 \geq 0. \end{aligned} \tag{2.4}$$

The constant $\theta > 0$ is the trade-off between risk and return.

2.2 Overview of Markov decision process (MDP)

A Markov decision process (MDP) is a mathematical framework for modelling discrete time-sequential decisions of an intelligent decision-maker with stochastic outcomes [130]. MDPs are useful for studying discrete dynamic programming optimization problems [72] with applications in finance, operation research, artificial intelligence (AI), and many other domains [17, 19, 127, 157]. Like all other sequential decision problems, MPD problems involve a dynamic system where inputs are selected sequentially after observing past outputs [130].

2.2.1 Components of a Markov decision process

An MDP model is generally built as a 4-tuple $(S, A_s, p(s'|s,a), r(s,a))$, where S is the set of possible states, A_s is the set of actions from which to choose in each state. As in a Markov chain, the state space represents the possible conditions in which the system may find itself. In each state $s \in S$ at time t an action $a \in A_s$ is taken, and the agent receives an immediate reward, $r(s,a)$, then the system moves randomly to state s' at time epoch $t+1$ according to the transition probability distribution $p(s'|s,a)$. The decision-maker aims to determine the collection of actions in each state that maximizes the expected discounted reward $V(s)$, given by

$$V(s) = E\left(r(s_0,a_0) + \gamma r(s_1,a_1) + \gamma^2 r(s_2,a_2) + \dots\right), \tag{2.5}$$

where $0 < \gamma < 0$ is the discount factor, $t = 0, 1, 2, \dots$ is the time epochs, and $r(s_i, a_i)$ represent deterministic or stochastic rewards.

The decision epoch is the periodic moment the decision-maker must make a decision. Let $\mathcal{T}$ be a finite or infinite sequence of the natural numbers of the form $(1, 2, 3, ..., T_{max})$ or $(1, 2, 3, ...)$ respectively, denoting the decision epochs, also called time steps, at which actions need to be taken. The MDP is called finite horizon if $T_{max} < \infty$ else it is infinite horizon. For discrete-time problems, decisions are made at decision epochs, while, decisions are made continuously at random points of time when certain events occur when dealing with continuous-time problems. Continuous MDP's are best dealt with in control theory methods based on dynamic system equations.

The collection of actions over a horizon is called a policy. A policy π maps each state $s \in S$ to an action $a \in A_s$. Given a policy π, the expected value function V^π is defined as

$$V^\pi(s) = E^{(\pi)}\left(r(s_0,a_0) + \gamma r(s_1,a_1) + \gamma^2 r(s_2,a_2) + \dots\right) \tag{2.6}$$

This is the expected discounted return of the payoffs when policy π is used.

The optimal policy is the collection of action, $a \in A_s$ from each state $s \in S$ over the horizon that maximizes the expected return. The optimal policy π^* is quickly recuperated using the optimal value function in computation. More generally, $V^{\pi^*}(s)$ recursively satisfies the Bellman equation [17] given by

$$V_k^{\pi^*}(s) = \max_a \left\{ r(s,a) + \gamma \sum_{s' \in S} p(s'|s,a) V_{k-1}^{\pi}(s) \right\} \tag{2.7}$$

and the optimal policy π^* is obtained as

$$\pi^* = arg\max_a \left\{ r(s,a) + \gamma \sum_{s' \in S} p(s'|s,a) V_{k-1}^{\pi}(s) \right\}. \tag{2.8}$$

2.2.2 Solving MDPs

Here we briefly describe methods, approximations, and algorithms used to solve MDPs. The fundamental approach to solving an MDP is the use of iterative algorithms. Many iterative methods are available to find the optimal or approximately optimal policies for the MDP [19] but the value iteration, the policy iteration, and the linear programming stand out.

Value iteration

Howard [72] was the originator of the value iteration algorithm. Value iteration uses dynamic programming methods iteratively to determine the value function of each state. The optimal value is obtained when the value of the iteration becomes stationary, satisfying the vector equation $V^* = LV^*$ [17, 130], where L is the Bellman operator. The algorithm is given as follows [20, 130]:

Value Iteration Algorithm

Step 1	Select $v^0 \in \mathcal{R}$, set $n = 0$, and specify $\epsilon > 0$.
Step 2	Compute $v^{n+1}(i) = Lv^n(i)$ for all $i \in$ S.
Step 3	If $\|v^{n+1} - v^n\| < \epsilon(1-\gamma)/2\gamma$, go to **Step 4**, otherwise, increase n by 1 and return to **Step 1**.
Step 4	Return with the actions attaining the maximum.

The run time of every step of the value iteration algorithm is complex. It requires $\bar{A}|S|^2$ multiplications and divisions, where $\bar{A}$ is an average number of actions per state, and the total maximum number of iterations needed by the algorithm is polynomial in $|S|$ and $\bar{A}$ [149]. It also converges slowly with an exponential rate to discount factor γ when the discount factor approaches 1.

Policy iteration

Another iterative algorithm to solve the MDP is the policy iteration algorithm. The policy iteration algorithm operates the policy directly. It begins by assessing an arbitrary policy and then uses the value function of that policy by an iterative update to find better policies according to

$$\pi^* = \arg\max_a r(s,a) + \gamma \sum_{s' \in S} p(s'|s,a) V_{k-1}^{\pi}(s). \tag{2.9}$$

The policy iteration algorithm is as follows [20, 130]:

Policy Iteration Algorithm

Step 1 Select arbitrary decision rule d_0, set $n = 0$.
Step 2 (Policy evaluation) Compute v^n by solving $(I - \gamma P_{d_n})v = r_{d_n}$.
Step 3 (Policy improvement) Choose the decision rule d_{n+1} such that $d_{n+1} \in argmax_{d \in D}\{r_d + \gamma P_d v^n\}$ setting $d_{n+1} = d_n$ if possible.
Step 4 If $d_{n+1} = d_n$, set $d^* = d_n$, stop.
Otherwise, increment n by 1 and return to **Step 2**.

The policy iteration algorithm requires $|S|^3$ multiplications and divisions, converging much faster than the value iteration [174]. The detailed discussion of the complexity of iterative methods of MDP can be found in Littman et al. [99], Papadimitriou and Tsitsiklis [121].

Linear Programming

The linear programming approximation method proceeds by transforming the MDP into an equivalent LP and approximates the value function by assuming a specific parametrized form then solving the chosen approximations to obtain the approximate values of the value function, v_{ALP} [106, 2, 42, 130]. The Approximate v_{ALP} is then used to determine the optimal policy in each state. A popularized LP approximation of MDP formulations is given in De Farias and Van Roy [42], Puterman [130].

From Powell [127] and Puterman [130], we know that, if

$$V(s) \geq \max_a (R(s,z) + \sum_{s \in S} P_{(s'|s,z_i)} V(s')),$$

then $V(s)$ is an upper bound on the value of being in each state. This means that the problem of finding the optimal values can be solved using the primal linear program

$$\min \sum_{j \in \mathcal{S}} d(s) V(j,z) \tag{2.10}$$

subject to

$$V(s,z) \geq r(s,z) + \gamma \sum_{j \in S} p(j|s,z) V(j,z) \qquad \forall s, j \in \mathcal{S}, z \in \mathcal{Z}$$

where $d(s)$ is any positive value, $V(s,z)$ is state s value, $r(s,z)$ is the instantaneous result in state s, $p(j|s,z)$ is the transition probability and γ is the discount rate.

With this primal in Equation 2.10, we know from Denardo [43] that solving the dual provides an approximation to the weight of the actions. The dual is obtained as Equation 2.11

$$\begin{aligned} &\max \sum_{s} \sum_{z} r(s,z)W(s,z) \\ &\text{subject to} \\ &\sum_{z} W(j,z) - \gamma \sum_{s\in\mathcal{S}} \sum_{z\in\mathcal{A}} p(j|s,z)W(s,z) \leq d(j) \\ &\forall s, j \in \mathcal{S}, z \in \mathcal{Z} \end{aligned} \tag{2.11}$$

where $W(s,z)$, $s \in S$, and $z \in \mathcal{A}$ are the policy flow in state s when action z is taken. The normalized $W(s,z)$ is interpreted as the optimal steady-state probability that action z is applied in state s.

The cost function:

$$\sum_{s_t} \sum_{z_t} R(s_t, z_t)W(s_t, z_t)$$

is the steady-state average reward per transition. Then as a result, the probability mass function of action z_t in state s_t is obtained as :

$$\pi(s_t, z_t) = \frac{W(s_t, z_t)}{\sum_{z_t} W(s_t, z_t)} \tag{2.12}$$

The action that provides the optimal policy is the action with the maximum probability in that state and is defined as

$$z_t^* = \arg\max_{z_t \in \mathcal{A}_{s_t}} \pi(s_t, z_t),$$

where $\mathcal{A}_{s_t}$ is the set of all the actions possible in state s_t.

2.3 Overview of Queuing Game

2.3.1 Game Theory

Game theory is the mathematical theory of games and has been popularized by John Von Neumann and Morgenstern [166]. It is concerned with the logic of decision-making in situations where self-interested entities interact and how those interactions should be structured to lead to a "better" abstract concept called utility. In economic literature, utility or payoff function is a function representing a consumer's preference. A game is any situation in which a decision-maker optimizes its utility by anticipating the reactions to his actions by one or more other decision-makers.

While in optimization problems, one decision-maker seeks a value $x \in \mathcal{X}$ ($\mathcal{X} \subseteq \mathcal{R}^m$ is a closed, unbounded domain of the decision variable x) that maximizes/ minimizes a typical

function $f(x)$, game theory on the other hand, is concerned with the situation with two or more decision-makers called players. The function each player has to maximize/ minimize is called the payoff function/ utility and depends on the decision variable of other players.

In the case of two players, if Player 1 has to decide on $x_1 \in \mathcal{X}_1$ and Player 2 on $x_2 \in \mathcal{X}_2$, the goal of Player i, $i = 1, 2$ is

$$\max \, or \min_x f_i(x_1, x_2). \tag{2.13}$$

Abstractly, a game has three components:

- A set of players.
- Sets of actions or decisions or strategies, $S_i, i = 1, 2, \ldots$ available to each player.
- Payoffs functions $f_i : S_1 \times S_2 \times \cdots \rightarrow \mathcal{R}$ defining each player's preference.

Various relevant concepts of the game process referring to different situations have been proposed in the literature. Based on the order of play, that is the order of decision-making, games can be distinguished into sequential-move and simultaneous-move games. In sequential-move games, decision-makers choose their actions one after the other, and in simultaneous-move games, they choose their actions simultaneously. In simultaneous-move games, we can distinguish between a cooperative or a non-cooperative game. In non-cooperative games, players do not communicate with each other and work independently to achieve their selfish goals. In a cooperative game, players can discuss their actions and agree.

Games are also classified based on the number of alternative strategies available to each player. Games may be finite or infinite. A finite game has all players with a finite number of strategies. It is infinite if at least one player has an infinite number.

A game is also classified based on the nature of the players' payoffs. In zero-sum games, the sum of the players' payoffs is zero. It is an interaction in which one person's gain is equivalent to another's loss, so the net change in wealth or benefit is zero. Otherwise, it is called a nonzero-sum game. The state of the information available to each player is another extensive way to classify games. We have complete and incomplete information games. Finally, a game is classified based on the involvement of time. If time is a factor considered in any player's decision-making, the game is dynamic; else, the game is static. In dynamic games or evolutionary games, the decision variable is a function of time $x(t)$. The time variable can be discrete or continuous over a range.

For static games, solution schemes depend on the formulation of the problem and are divided into cooperative and non-cooperative solutions. In general, there is no optimal solution for the game instead an equilibrium solution. Since an outcome that is optimal for one player can be detrimental for the other one. For cooperative games, the Pareto optima are sought.

Definition A tuple of strategies $s_1^*, s_2^*, \ldots, s_n^*$ is said to be Pareto optimal if there exists no other tuple $(s_1, s_2, \ldots, s_n) \in \mathcal{S}_1 \times \mathcal{S}_2 \times \cdots \times \mathcal{S}_n$ such that

$$f_1(s_1, s_2, \ldots, s_n) > f_1(s_1^*, s_2^*, \ldots, s_n^*) \text{ and } f_2(s_1, s_2, \ldots, s_n) \geq f_2(s_1^*, s_2^*, \ldots, s_n^*)$$

or

$$f_2(s_1, s_2, \ldots, s_n) > f_2(s_1^*, s_2^*, \ldots, s_n^*) \text{ and } f_1(s_1, s_2, \ldots, s_n) \geq f_1(s_1^*, s_2^*, \ldots, s_n^*)$$

where $f_i()$ is the payoff function of player i and s_i is player i's action or strategy.

Under Pareto optimality, it is impossible to strictly increase one of the players' payoff without strictly decreasing the payoff of the others. For $\eta_i \in [0, 1]$, such that $\sum_i \eta_i = 1$, the tuple $(s_1, s_2, \ldots, s_n)$ that maximize the optimization problem

$$\max_{s_1 \in \mathcal{S}_1, s_2 \in \mathcal{S}_2, \ldots, s_n \in \mathcal{S}_n} \sum_i (\eta_i f_i(s_1, s_2, \ldots, s_n)) \tag{2.14}$$

is Pareto optimal.

There are two major non-cooperative solutions: the Nash equilibrium and the Stackelberg equilibrium. The Nash equilibrium is the main solution concept for symmetric non-cooperative games. The Nash equilibrium [115] is an n-tuple of optimal strategies, one for each player, such that anyone who deviates from it unilaterally cannot possibly improve his payoff, as long as the other players stick to their equilibrium strategy.

Definition The tuple $(s_1^*, s_2^*, \ldots, s_n^*)$ is Nash equilibrium of the game if, for every strategy $s_i \in \mathcal{S}_i, i = 1, 2, \ldots, n$ denotes player i's strategy, we have

$$f_i(s_1^*, s_2^*, \ldots, s_n^*) \geq f_i(s_1^*, s_2^*, \ldots, s_{i-1}^*, s_i, s_{i+1}^*, \ldots, s_n^*), \tag{2.15}$$

where $f_i()$ is the payoff function of player i.

Strategies $(s_1, s_2, \ldots,)$ is a Nash equilibrium if and only if it is a fixed point of the best reply map. In general, a Nash equilibrium may not exist or be unique. However, there is no general method for finding the Nash equilibrium and mostly optimization methods such as LP, graphical methods, algorithms or heuristic approaches are used. When the payoff functions are differentiable, Nash equilibrium is obtained by the first-order conditions of the player's payoff concerning the player's strategy.

The other important solution scheme of the non-cooperative game is the Stackelberg equilibrium. The Nash equilibrium provides a solution to non-dominating or asymmetric games. The Stackelberg equilibrium specifies the behaviour when one of the players (the leader) can

impose his strategy. It is named after the economist Von Stackelberg for his work [153]. It assumes the leading player announces his strategy in advance, and the follower player chooses accordingly. For an illustration, suppose Player 1 is the leader and Player 2 is the follower who reacts to Player 1's decision. When Player 1 announces its strategy, Player 2 responds to maximize his payoff according to Player 1's decision. If $s_1 \in \mathcal{S}_1$ denotes Player 1's strategy, then Player 2 chooses its optimal strategy $s_2^* \in \mathcal{S}_2$ to maximize its payoff denoted by $f_2(s_1^*, s_2)$. Player 2 assumes the knowledge of Player 1's strategy. Player 2 chooses its optimal strategy s_2^* such that

$$f_2(s_1, s_2^*) \geq f_2(s_1, s_2), \quad \forall\ s_2 \in \mathcal{S}_2. \tag{2.16}$$

This best response function of the follower is known as the reaction function and is obtained as

$$\Gamma(s_1) \in \arg\max_{s_2 \in \mathcal{S}_2} f_2(s_1^*, s_2). \tag{2.17}$$

Using the follower's best response, the goal of the leader is now to maximize the composite function $f_1\left(s_1, \Gamma(s_1)\right)$.

Definition A couple of strategies $(s_1^*, s_2^*) \in \mathcal{S}_1 \times \mathcal{S}_2$ is called a Stackelberg equilibrium if $f_1(s_1, s_2) \leq f_1(s_1^*, s_2^*)$.

The computation is done backward since the leader needs the reaction function of the follower to compute its best response. A complete discussion and review of game theory is provided in Osborne and Rubinstein [120].

2.3.2 Queuing Theory

In this section, we will review some basics of queuing theory. The review will elaborate on notations, queue disciplines, and elementary queues key performance indicators.

Agner Krarup Erlang in 1909 [53] did a seminal work on queuing systems. During the Second World War and the advent of modern computing, the field has evolved with varied results and applications even in healthcare planning.

A queuing system is characterized by several elements: the arrival pattern, service pattern of servers, queue discipline, the system capacity, the number of service channels, and the number of service stages.

Since the description of the characteristics of a queue becomes very wordy, Kendall [82] popularised a standard shorthand notation used to describe queuing systems. In his abbreviation A/B/X/Y/Z, the first and the second character indicates the distribution of the arrival and service time respectively, the third, the number of parallel channel of servers, the fourth describes the system capacity, the fifth the queue discipline by its given acronym and the sixth

indicates the pool size of customers system draw from. For example, M/D/S/∞/FCFS/∞ could represent an ICU unit where inter-arrival time distribution is exponential, service time distribution is deterministic, the system has S beds (servers), no restriction on the number of people allowed in the system, patients are treated on a first come first serve basis, and an infinite population pool to draw from.

We used three types of performance indicators of interest: a measure of time customers spends in the system, a measure of the number of customers in the system and a measure of the performance time of servers. Little's formulas are simple yet powerful relationships between the average number of customers in the queue, L, the mean waiting time, and the arrival rate at any time, assuming a steady-state system ($\lambda = \mu$). Little showed that if $N(t)$ is the random number of customers in the system, the expected number of customers in the system is given by

$$L = \mathrm{E}(N(t)) = \sum_{n=0}^{\infty} n p_n = \lambda W, \tag{2.18}$$

where p_n is the probability that there are n customers in the system. The simplest queue model is the M/M/1 queue. For this queue, the inter-arrival time and the service times are exponential with rates λ and μ, respectively. At equilibrium, the first principle flow balance equations are given by

$$\begin{aligned} (\lambda + \mu) p_n &= \mu p_{n+1} + \lambda p_{n-1} \quad (n \geq 1) \\ \lambda p_0 &= \mu p_1. \end{aligned} \tag{2.19}$$

Using an iterative method or generating functions of operators, the probability distribution of the full steady-state for the M/M/1 queue system is obtained as

$$p_n = \left(1 - \frac{\lambda}{\mu}\right)\left(\frac{\lambda}{\mu}\right)^n. \tag{2.20}$$

Some basic measures of effectiveness are derived as follows. The expected number in the system

$$L = \frac{\lambda}{\mu - \lambda}. \tag{2.21}$$

The expected number in the queue is

$$L_q = \frac{\lambda^2}{\mu(\mu - \lambda)}. \tag{2.22}$$

It follows from Little's formula that the expected time in the system is

$$W = \frac{1}{\mu - \lambda} \tag{2.23}$$

and the expected time in queue is

$$W_q = \frac{\lambda}{\mu(\mu - \lambda)}. \tag{2.24}$$

For a complete introduction to the queuing theory, see Shortle et al. [146], Kleinrock [87].

2.3.3 Queuing Games

A queuing game is the application of game theory to strategically manage customers and servers' decision in a queue. It models situations where two or more players (customers or servers) control different components of the queue. Customers want to spend the least time in queue and must decide on the queue to join. Servers want to attract as many customers as possible. So, they try to provide the shortest time in queue or in service and the best service experience possible. In a queuing game, a model is constructed so that decisions about queue lengths, waiting time, queueing cost and other vital indicators can be studied and predicted.

The queuing game literature is replete with models that capture customer's behaviour in the system. In this thesis, we focus on the system's behaviour. Queuing games are mostly defined as a non-cooperative game with $N = \{1, \ldots, n\}$ finite set of players with a set of actions $A_i,\ i \in N$ for each player. A pure strategy for player i is an action $a \in A_i$. A mixed strategy assigns a probability distribution for selecting an action from A_i. A strategy profile is a set $s = \{s_1, \ldots s_n\}$ of strategies used by each player. The payoff of player i is, therefore, $f_i(s)$. Strategy s_i^* dominants if $f_i(s_1, \ldots, s_i^*, \ldots s_n) \geq f_i(s_1, \ldots, s_i, \ldots s_n)$ for all strategies s_i available for player i. A strategy profile is an equilibrium strategy if it is the set of the best response of each player. An equilibrium may not always exist. In queuing games, queues are mainly classified as being observable or unobservable. In observable queues, customers arriving at the station know the queue length and decide to join the queue or not. The queue can also be considered from the social welfare point of view. In this case, the objective is to maximize the benefit for the entire society, both the customers and the servers.

2.4 Survival Analysis

2.4.1 Introduction and Basic Concepts

Survival analysis is a collection of statistical methods for data analysis where the outcome variable of interest is the time to the occurrence of an event known as survival time. It is also referred to as time to the event analysis. Survival time is the length of time measured from the beginning of a process to the occurrence of a given event of interest. In health sciences and insurance for example, survival time or time to the event can be time to death, or time to a sickness or a handicap or time until an accident or time to the failure of an organ. Therefore, the measurement of the survival time, requires an exact time origin, a time measurement scale, and a clear definition of an event of interest.

The interesting characteristic of survival data that led to the development of new statistical methods is the presence of censored observations. Censored observations occur when some observations have not experienced the event of interest before the end of the data collection or study. The precise time to the event of these observations is unknown. Censoring occurs when the survival time information about some individuals is only partially known, and the exact survival time is unknown. There are three types of censoring: 1) right censoring, 2) left censoring, and 3) interval censoring.

Right censoring occurs when the time it took the event of interest to occur is greater or equal to the observed length of time. Let C denote the censoring time and X the observed event time. The observed survival time is X if the event occurs or C if it is censored, whichever comes first. Let $(T;\delta)$ denote the observed data. $T = min(X, C)$ and

$$\delta = I_{(X \leq C)} = \begin{cases} 0, & \text{if } X > C \quad \text{(Event unobserved, right censored)} \\ 1, & \text{if } X \leq C \quad \text{(Event observed).} \end{cases} \tag{2.25}$$

Left censoring occurs when the time it took the event of interest to occur is less or equal to the observed length of time. Interval censoring occurs when the event of interest is known to have happened within a two-time interval, but the actual survival time is not known.

In an experimental setting, Type I censoring occurs when in an experiment that is set to stop at a predetermined time, some observations have not experienced the event. Those events right-censored. Type II censoring occurs when in the experiment, a predetermined number of subject are observed to have experienced the event of interest and the remaining are right-censored. A random or non-informative censoring is when each observation has a censoring time that is statistically independent of their survival time. Right censoring is very common in survival time data, but left censoring is relatively rare. Therefore, the terms “censoring” or

"censored" will be used in this thesis to mean "right censoring".

2.4.2 Survival Time Distribution

This part of the presentation is based on the textbook Kalbfleisch and Prentice [78]. Survival times are completely characterized by any of three functions: the survival function, the probability density function, and the hazard function.

Let T be a random variable denoting the survival time. The event time may be a discrete or continuous non-negative random variable. The probability density function (PDF) of the continuous random variable is specified as $f(t)$ and the cumulative distribution function (CDF) is

$$F(t) = Pr(T \leq t) = \int_0^t f(s)ds, \quad t \geq 0. \tag{2.26}$$

The CDF is the probability of the event happening before or at most at time t. The probability of the event happening after time t is known as the survival function and is defined as

$$S(t) = Pr(T > t) = \int_t^\infty f(s)ds, \quad t \geq 0. \tag{2.27}$$

Of interest is the hazard function $h(t)$, which is the instantaneous failure rate at t given survival up to time t, i.e.,

$$h(t) = lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t}, \ t \geq 0. \tag{2.28}$$

The relationship between the hazard rate, the PDF, and the survival function is given as

$$h(t) = \frac{f(t)}{S(t)} = -\frac{d \log(S(t))}{dt}, \tag{2.29}$$

and the integral of the hazard rate is the cumulative (or integrated) hazard function $H(t)$, defined as

$$H(t) = \int_0^t h(s)ds. \tag{2.30}$$

The relationship between the survival function and the cumulative hazard function is

$$S(t) = \exp(-H(t)), \tag{2.31}$$

and we also have

$$f(t) = h(t)\exp\left(-H(t)\right) = h(t)S(t) \tag{2.32}$$

The survival function, the probability mass function and the hazard functions are analogously specified for the discrete variable. For a discrete random variable T taking well-ordered values $0 \leq t_1 < t_2 < \ldots$, let $P(T = t_i) = f(t_i), i = 1, 2, \ldots$, be the probability mass function then the survival function is defined as

$$S(t) = \sum_{j|t_j \geq t} f(t_j) = \sum f(t_j) I_{t_j \geq t}, \tag{2.33}$$

where the indicator function

$$I_{(t_j \geq t)} := \begin{cases} 0 & \text{if } t_j < t \\ 1 & \text{if } t_j \geq t. \end{cases} \tag{2.34}$$

The hazard at time t_i, $h(t_i)$ is the conditional probability of failure at time t_i given that the individual has survived up to time t_i,

$$h_i = h(t_i) = P(T = t_i | T \geq t_i) = \frac{f(t_i)}{S(t_i)} = \frac{S(t_i) - S(t_{i+1})}{S(t_i)} = 1 - \frac{S(t_{i+1})}{S(t_i)}, i = 1, 2, \ldots, \tag{2.35}$$

thus

$$1 - h(t_i) = \frac{S(t_{i+1})}{S(t_i)}, \tag{2.36}$$

and

$$\prod_{i|t_i < t} (1 - h(t_i)) = \frac{S(t_2) \times S(t_3) \times \ldots S(t_{i+1})}{S(t_1) \times S(t_2) \times \ldots S(t_i)} = S(t) \tag{2.37}$$

since $S(t_1) = 1$ and $S(t) = S(t_{i+1})$.

2.4.3 Non-Parametric Estimation of the Survival Models: Kaplan-Meier Analysis

Every statistical data analysis starts conveniently with a data summary through descriptive statistics. This section presents the numerical and graphical summaries of the survival time data. For the survival times, this analysis is done through estimates of the survival function and hazard function using the data. These methods are said to be non-parametric methods because

no parametric assumptions are made about the distribution of survival time.

There are generally two non-parametric methods: the life-table analysis and the Kaplan Meier (KM) analysis. The life-table analysis method was the first to be developed and provided more detailed statistics. However, the Kaplan-Meier analysis method is superior in many cases and, with the advent of computers, is now the method of choice [80, 131, 91, 137]. Whereas the life-table approach is based on grouped data, the KM focus on analyzing individual data. Here, we present only the Kaplan Meier analysis method.

If an uncensored survival sample of n distinct survival times is observed from a continuous homogeneous population, then the survival function can be estimated using the empirical survival function

$$S_n(t) = \frac{1}{n}\sum_{i=1}^{n} I_{(t_i>t)}, \tag{2.38}$$

where $I_{(t_i>t)}$ is the indicator function that takes the value 1 if $t_i > t$ and 0 otherwise. This simple and convenient summary is the proportion alive at time t. Graphically, it is a step function that decreases by n^{-1} at each observation. However, survival data often contain censoring and for this purpose, the KM estimator is a more consistent and convenient method for estimating the survival function. The KM estimator uses only the data on the time to the event without any covariate to estimate the survival curves.

Let n be the sample size of the observed survival time. Let $0 \leq t_{(1)} < \cdots < t_{(m)} < \infty$ be the distinct ordered observed times of events. Suppose that d_j observations experience the event of interest at time t_j and u_j observations are censored in the interval $[t_j, t_{(j+1)})$. Let $n_j = (d_j + u_j) + \cdots + (d_m + u_m)$ be the size of the risk set at time t_j, where the risk set denotes the collection of individuals alive and uncensored just before $t_{(j)}$. The Kaplan-Meier or product-limit estimate of the survival function, $S(t)$, is defined by

$$\hat{S}(t) = \prod_{j|t_{(j)}\leq t}\left(\frac{n_j - d_j}{n_j}\right) = \prod_{j|t_{(j)}\leq t}\left(1 - \frac{d_j}{n_j}\right) \tag{2.39}$$

The Kaplan-Meier estimate is a step function with discontinuities or jumps at the observed event times. If there is no censoring, the K-M estimate coincides with the empirical survival function.

Non-parametric Maximum Likelihood

The K-M estimator has been shown to maximize the discrete likelihood. Considering both the contribution to the likelihood of cases that die and those that are censored at time t, the likelihood function is given as

$$L = \prod_{i=1}^{m} \left[S_{(i-1)} - S_{(i)}\right]^{d_i} S_{(i)}^{u_i} \tag{2.40}$$

where $[S_{(i-1)} - S_{(i)}]$ is the probability of experiencing the event at time $t_{(i)}$ and $S(t_{(i)})$ is the probability of a censored observation between time interval $[t_{(i)}, t_{(i+1)})$. Note that in the likelihood function, it is assumed that all censored observations between the given interval have the same likelihood. Without loss of generality, take $t_{(0)} = 0$ and $S(t_{(0)}) = 1$, and from Equation 2.37,

$$S(t_{(j)}) = \frac{S(t_{(1)}) \times S(t_{(2)}) \times \cdots \times S(t_{(j)})}{S(t_{(0)}) \times S(t_{(1)}) \times \cdots \times S(t_{(j-1)})}.$$

Let $h(j) = \frac{S(t_{(j)})}{S_{(j-1)}}$, then $S(j) = h(1) \times h(2) \times \cdots \times h(j)$. Substituting $S(j)$ by $h(j)$ in Equation 2.40, we obtained the following binomial likelihood of the parameter $h(j)$.

$$\begin{aligned} L &= \prod_{j=1}^{m} \left[[1 - h(j)]^{d_j} \left[h(1) \times h(2) \times \cdots \times h(j-1)\right]^{d_j + u_j} \left[h(j)\right]^{u_i} \right] \\ &= \prod_{j=1}^{m} \left[1 - h(j)\right]^{d_j} \left[h(j)\right]^{n_j - d_j} \end{aligned} \tag{2.41}$$

Therefore, the maximum likelihood estimator of $h(j)$ is equal to

$$\hat{h}(j) = 1 - \frac{d_i}{n_j}. \tag{2.42}$$

The K-M estimator follows from multiplying these as

$$\hat{S}(t) = \prod_{j:t_{(j)} \le t} \hat{h}(j) = \prod_{j:t_{(j)} \le t} \left[1 - \frac{d_j}{n_j}\right] \tag{2.43}$$

Greenwood [62] were the first to provide confidence intervals for the survival probability. First, the sample variances of $\hat{h}(j)$ follow from the usual binomial formula

$$Var(\hat{h}(j)) = \frac{h_j(1 - h_j)}{n_j} \tag{2.44}$$

The sample variance of the K-M estimate, $\hat{S}(t)$, is obtained by applying the delta method twice.

Let $\log(\hat{S}(t)) = \sum_{i=1}^{j} \log(\hat{h}_j)$. Applying the delta method for the first time, we have

$$Var(\log(\hat{h}_j)) = \left[\log'(h_j)\right]^2 var(h_j) = \left[\frac{1}{h(j)}\right]^2 \frac{h_j(1 - h_j)}{n_j} = \frac{(1 - h_j)}{n_j h_j}. \tag{2.45}$$

Since the event and its censoring are independent, $cov(\hat{h}_i, \hat{h}_j) = 0$ and because $\log(\hat{S}(t))$ is a sum,

$$Var(\log(\hat{S}(t))) = Var(\sum \log(\hat{h}_j)) = \sum \frac{(1-h_j)}{n_j h_j} = \sum \frac{(d_j)}{n_j(n_j-d_j)}. \tag{2.46}$$

Applying the delta method again for the second time, we get the variance of the survivor function from the variance of its log:

$$Var(\hat{S}(t)) = \left[\hat{S}(t)\right]^2 Var(\log(\hat{S}(t))) = \left[\hat{S}(t)\right]^2 \sum \frac{(d_j)}{n_j(n_j-d_j)} \tag{2.47}$$

This equation is known as Greenwood's formula. Using Greenwood's formula, the K-M estimator is a consistent and convergent estimator of the survival function [26, 4, 57]. Thus the confidence intervals can be constructed based on the normal approximation of $S(t)$.

As pointed out by Kalbfleisch and Prentice [78], many other authors consider first the cumulative hazard function $\Lambda(t)$. Nelson-Aalen estimated the cumulative hazard function as

$$\hat{\Lambda}(t) = \sum_{t_j \le t} d_j/n_j = \sum_{t_j \le t} \hat{h}_j. \tag{2.48}$$

Breslow and Crowley [26] suggested that the survival function be estimated as

$$\hat{S}(t) = \exp\left(-\hat{\Lambda}(t)\right). \tag{2.49}$$

Breslow's estimator and the K-M estimator are asymptotically equivalent, and usually are quite close to each other, particularly when the number of deaths is small relative to the number exposed.

To compare two or more K-M survival functions, the survival curves can give us a graphical view. To check the statistical significance of the difference observed, a commonly used formal non-parametric statistical test is the Mantel-Haenszel log-rank test [107].

Let $t(1) < t(2) < \cdots < t(m)$ denote the ordered event times across all groups. Suppose that d_i events occur in the whole sample and d_{ij} events occur at $t(i)$ in group j, and n_i subjects at risk in the whole sample while n_{ij} subjects are at risk just before $t(i)$ in group $j \quad (i = 1, 2, \ldots, m)$. If the survival probabilities are the same in all groups, then the d_i events at time $t(i)$ are distributed among the k groups in proportion to the number at risk. Thus, conditional on d_i and n_{ij},

$$E(d_{ij}) = \frac{n_{ij} d_i}{n_i}. \tag{2.50}$$

Given n_j and d_j, the distribution of the counts conditional on both the row and column totals is hypergeometric

$$\frac{\binom{d_j}{d_{ij}}\binom{n_j - d_j}{n_{ij} - d_{ij}}}{\binom{n_j}{n_{ij}}}. \tag{2.51}$$

The mean is given above and the variance and the covariance are given as

$$Var(d_{ij} = \frac{d_i(n_i - d_i)n_{ij}(n_i - n_{ij})}{n_i^2(n_i - 1)}, \tag{2.52}$$

and

$$Cov(d_{iv}, d_{iw}) = -\frac{d_i(n_i - d_i)n_{iv}n_{iw}}{n_i^2(n_i - 1)}. \tag{2.53}$$

Let $\vec{d_i}$ denote the vector of the number of events for all groups at time $t(i)$. Let the mean of $\vec{d_i}$ be $E(\vec{d_i})$ and var-cov matrix $var(\vec{d_i})$. The overall sum is given as

$$D = \sum_{i=1}^{m}[\vec{d_i} - E(\vec{d_i})] \quad V = \sum_{i=1}^{m} Var(\vec{d_i}). \tag{2.54}$$

Under the null hypothesis

$$H_0 : S_1(t) = S_2(t) = ... = S_k(t), \tag{2.55}$$

the quadratic form

$$Q = D^{'}V^{-}D, \tag{2.56}$$

where $D^{'}$ is the transpose of D and V^{-} is the generalized inverse of V, has a χ^2 distribution with degree of freedom $k - 1$. For $k = 2$, we have

$$z = \sqrt{Q} = \frac{\sum(d_{i1} - E(d_{i1}))}{\sqrt{\sum var(d_{i1})}}. \tag{2.57}$$

An approximation for $k \geq 0$ which does not require matrix inversion is

$$\sum_i \sum_j \frac{(O_{ij} - E_{ij})^2}{E_{ij}}, \tag{2.58}$$

where O_{ij} is observed, and the E_{ij} is expected deaths at time $t_{(i)}$ in group j. There is an alternative test such as the Wilcoxon's test, but the M-H test is the most popular.

2.4.4 Common Parametric Distribution Functions for Survival Data

While non-parametric methods work naturally well with observed data from homogeneous populations in building the survival function, one can also consider the analysis of survival data making parametric assumptions for the distribution of survival time. In this section, we introduce widely used parametric probability density functions to characterize survival time. Some of the often used models include but are not limited to Exponential, Weibull, Gompertz-Makeham, Gamma, Generalized Gamma, log normal, Gumbel, Frechet, Generalized Extreme Value, Log-Logistic, Exponential power, Inverse-Gaussian, Pareto, and the Generalized-Gamma distributions. This presentation is based on the textbooks Klein and Moeschberger [86], Kalbfleisch and Prentice [78] and Sun [156].

The Exponential Distribution

The exponential distribution has a constant hazard function

$$h(t) = \lambda,\ t \geq 0, \tag{2.59}$$

the density function is given as

$$f(t) = \lambda \exp(-\lambda t),\ t \geq 0,\ \lambda > 0, \tag{2.60}$$

and the survival function is obtained as

$$S(t) = \exp(-\lambda t). \tag{2.61}$$

Thus, a survival time, T has an exponential distribution with parameter λ is denoted $T \sim \exp(\lambda)$. The mean and standard deviation are $1/\lambda$. The exponential distribution has a fairly simple mathematical form and interesting properties that make it mathematically tractable and easy to manipulate. The memoryless property states that

$$P(T > t + s|T > s) = P(T > t). \tag{2.62}$$

This property means that given that a subject survived s units of time, its chances of surviving an additional t is the same as if the subject just started its life. Although the exponential distribution has been historically very popular, its constant hazard rate appears too restrictive in both health and industrial applications.

The Weibull Distribution

The Weibull distribution is an important extension of the exponential distribution with three parameters. It is denoted $T \sim W(\lambda, \theta; p)$. The hazard function is defined as

$$h(t) = \lambda^p p(t-\theta)^{p-1}, \quad \lambda > 0, t > \theta, p > 0. \tag{2.63}$$

The probability density function, $f(t)$ with shape parameter p , scale parameter λ , and location parameter θ is given by

$$f(t) = p\lambda^p(t-\theta)^{p-1}\exp(-\lambda^p(t-\theta)^p), \quad \lambda > 0, t > \theta, p > 0, \tag{2.64}$$

and the survival function, $S(t)$ is defined as

$$S(t) = \exp(-\lambda^p(t-\theta)^p), \quad \lambda > 0, t > \theta, p > 0. \tag{2.65}$$

When the location parameter $\theta = 0$, then T^p can be expressed as an exponential distribution as $T^p \sim \exp(\lambda)$. Clearly, the log of the Weibull hazard function is a linear function of log time:

$$\log(h(t)) = (p-1)log(t-\theta) + p\log(\lambda) + log(p). \tag{2.66}$$

This is used as an empirical test for the Weibull distribution.

Gompertz-Makeham Distribution

The first parametric family of distribution to smooth mortality tables is the Gompertz distribution. It is a three-parameter distribution that assumes that the hazard rate increases in geometrical progression. Thus

$$h(t) = \theta\lambda\exp(\lambda t), \quad \lambda, \theta > 0, t \geq 0. \tag{2.67}$$

The survival function is given as

$$S(t) = \exp(-\theta(e^{\lambda t-1})), \quad \lambda, \theta > 0, t \geq 0, \tag{2.68}$$

and the corresponding density function is obtained as

$$f(t) = \lambda\theta\exp(\lambda t - \theta(exp(\lambda t) - 1)), \quad \lambda, \theta > 0, t \geq 0. \tag{2.69}$$

λ is a scale parameter and θ is known as the frailty parameter. An interesting characterization of the Gompertz distribution is that like the Weibull, the log of the hazard is linear in time,

therefore closely related to the Weibull distribution. The Gompertz distribution can be thought of as a log-Weibull distribution. The Gompertz distribution family was extended by [105] with the hazard rate redefined as

$$h(t) = \alpha\theta\lambda + \theta\lambda \exp(\lambda t), \alpha, \lambda, \theta > 0, \quad t \geq 0. \tag{2.70}$$

Thus the survival distribution is given as

$$S(t) = \exp(-\theta(\exp(\lambda t) - 1) - \theta\lambda\alpha t), \tag{2.71}$$

and the density is

$$f(t) = \left(\alpha\theta\lambda + \theta\lambda e^{\lambda t}\right) \exp\left(-\theta\left(e^{\lambda t - 1} - \theta\lambda\alpha t\right)\right). \tag{2.72}$$

Log-Normal Distribution

T is log-normally distributed if $X = \log(T)$ is normally distributed. It is a skewed distribution that empirically fits many types of time to failure data. The PDF is given as

$$f(t) = \frac{1}{t\sigma\sqrt{2\pi}} \exp - \frac{(\ln(t) - \mu)^2}{2\sigma^2}, \tag{2.73}$$

and the survival function obtained as

$$S(t) = 1 - \Phi\left(\frac{ln(t) - \mu}{\sigma}\right), \tag{2.74}$$

where Φ is the standard normal CDF.

The Gamma Distribution

The Gamma distribution is a 3-parameter probability density function, with shape, scale, and location parameters given by α, λ, γ respectively. It is denoted by $T \sim \Gamma(\alpha, \lambda, \gamma)$ with density function

$$f(t) = \frac{\lambda(\lambda(t - \gamma)^{\alpha - 1} e^{-\lambda t}}{\Gamma(\alpha)}, \tag{2.75}$$

and survival function

$$S(t) = 1 - I(\alpha, \lambda, \gamma), \tag{2.76}$$

where $I(\alpha, \lambda, \gamma)$, the incomplete gamma function is defined as

$$I(\alpha, \lambda, \gamma) = \int_0^t \frac{\lambda^{\alpha-1} e^{-t+\gamma} dt}{\Gamma(\alpha)}, \tag{2.77}$$

and

$$\Gamma(\alpha) = \int_0^\infty \lambda^{\alpha-1} e^{-t+\gamma} dt, \tag{2.78}$$

is the gamma function. There is no closed-form expression for the survival and hazard functions; however, there are numerical algorithms for the computation.

The Generalized Gamma Distribution

The generalized gamma distribution has density

$$f(t) = \frac{\lambda p(\lambda(t-\gamma)^{p\alpha-1} e^{-\lambda(t-\gamma)^p}}{\Gamma(\alpha)}, \quad \lambda \geq 0, \gamma \geq 0, \alpha \geq 0, p \geq 0, k \geq 0. \tag{2.79}$$

- When $p = 1$, it is a gamma distribution.
- When $k = 1$, a Weibull distribution.
- When $p = 1$ and $\alpha = 1$, an exponential distribution.
- When $k \to \infty$, a log-normal distribution.

The Gumbel Distribution

The Gumbel distribution is known as the Type I extreme value distribution and has two forms: the smallest extreme referred to as Minimum Extreme Value Type I and the largest extreme referred to as Maximum Extreme Type I. The density function of the Maximum Extreme Value Type I is given by

$$f(t) = \frac{1}{\beta} e^{-\frac{t-\mu}{\beta}} e^{-e^{-\frac{t-\mu}{\beta}}}, \quad \mu \geq 0, \beta \geq 0, \tag{2.80}$$

and that of the Minimum Extreme Type I is given by

$$f(t) = \frac{1}{\beta} e^{\frac{t-\mu}{\beta}} e^{-e^{\frac{t-\mu}{\beta}}}, \quad \mu \geq 0, \beta \geq 0, \tag{2.81}$$

where μ is the location parameter and β is the scale parameter. The Survival functions are defined as

$$S(t) = e^{-e^t}, \quad \mu \geq 0, \beta \geq 0, \tag{2.82}$$

and

$$S(t) = 1 - e^{-e^{-t}}, \quad \mu \geq 0, \beta \geq 0 \tag{2.83}$$

respectively.

The Fréchet Distribution

The Fréchet distribution is referred to as the Maximum Extreme Value Type II distribution. The Fréchet distribution has a shape parameter α, scale parameter β, and location parameter γ. The density is given by

$$f(t) = \frac{\alpha}{\beta}\left(\frac{\beta}{t-\gamma}\right)^{\alpha+1} \exp^{-\left(\frac{\beta}{t-\gamma}\right)^{\alpha}}, \quad t > \gamma, \alpha > 0, \beta > 0. \tag{2.84}$$

And the survival function is given by

$$S(t) = 1 - exp\left(-\left(\frac{\beta}{t-\gamma}\right)^{\alpha}\right). \tag{2.85}$$

Generalized Extreme Value (GEV) Distribution

The generalized extreme value (GEV) distribution is a family of continuous probability distributions that combines the Gumbel, Fréchet and Weibull distribution families. The density function for the generalized extreme value distribution with location parameter μ, scale parameter β, and shape parameter k is given as

$$f(t) = \frac{1}{\beta} r(t)^{k+1} \exp\left(-r(t)\right), \beta \geq 0, \tag{2.86}$$

where

$$r(t) = \begin{cases} \left(1 + \frac{k(x-\mu)}{\beta}\right)^{-1/k} & \text{if } k \neq 0 \\ e^{\frac{-(t-\mu)}{\sigma}} & \text{if } k = 0 \end{cases} \tag{2.87}$$

The survival function is then obtained as

$$S(t) = 1 - \exp\left(-r(t)\right). \tag{2.88}$$

Varying the shape parameter k by setting $k = 0$, $k > 0$ and $k < 0$ defines the sub-families as Gumbel, Fréchet and Weibull families respectively.

2.4.5 Regression Survival Models

Survival distribution methods work well on survival time for homogeneous populations. However, when we consider a more general problem where we have a vector x of covariates, survival distributions do not determine whether or not certain variables affect the survival times. There are explanatory variables that may characterize the survival time. Application of regression methods for analyzing survival data are therefore necessary and require suitable methods due to the existence of censored observations and the fact that survival times are positive variables, and are rarely normally distributed.

Cox Proportional Hazards Model: A Semi-parametric Model

The Cox proportional hazard model has been proposed by Cox [40]. The model is a regression model for survival analysis, which relates several predictors simultaneously to the survival time. The model is called semi-parametric because it does not assume any distribution for the survival time but assumes that the effects of the independent covariates upon the survival time are constant over time and are additive in one scale. The model assumes that the covariates have a proportional effect on the hazard rate. In general, the model is of the form

$$h(t, x) = h_0(t) f(x), \tag{2.89}$$

where $h_0(t)$ is the baseline hazard and $f(x)$ is the function of the relative risk associated with covariate values x that acts multiplicatively on the hazard function. Specifically, the relative risk model by Cox [40] specifies that

$$h(t, x) = h_0(t) \exp\left(\beta_0 + \sum_{i=1}^{p} \beta_i x_i\right) = h_0(t) \exp\left(\beta X\right), \tag{2.90}$$

where $h(t, x)$ is the expected hazard at time t, $h_0(t)$ is the baseline hazard at time t when all of the independent variables are zero and βs are the coefficients to be estimated. The corresponding survival functions known as the Lehman are is given as follows:

$$S(t, x) = [S_0(t)]^{\exp(\beta X)}, \tag{2.91}$$

where $S_0(t) = \exp\left[-\int_0^t h_0(s)ds\right]$.

A key characteristic of proportional hazards models is the constancy of all covariates at time 0. As a result, due to the proportionality of the hazard function, the ratio of the hazard rate of two different subjects with covariates X_1 and X_2 called the hazard ratio is a constant.

$$HR = \frac{h_0(t)\exp(\beta X_1)}{h_0(t)\exp(\beta X_2)} = \exp\left[(X_1 - X_2)\beta\right], \tag{2.92}$$

which is time-independent.

Estimation of the Cox Proportional Hazard Model

The baseline hazard function and the parameters, β of the model are estimated as proposed by Cox [40] using partial likelihood. The construction of the partial likelihood follows suit.

Let $t_1, t_2, \ldots, t_n$ denote the observed survival times, $t_{(1)} < t_{(2)} < \cdots < t_{(r)}$ be the observed distinct times of death, and $R_{(t_{(j)})}, j = 1, 2, \ldots, m$ be the risk set at time $t_{(j)}$, defined as the set of indices of the subjects that are alive just before time $t_{(j}$. The conditional probability that the ith individual experiences the event at $t_{(j)}$ given that one individual from the risk set on $R_{(t(j))}$ experiences the event at time $t_{(j)}$ is

$$\begin{aligned} P(\text{subject i dies}) &= \frac{P(\text{subject i dies at } t_{(j)})}{P(\text{one death at } t_{(j)})} \\ &= \frac{h_i(t_{(j)})}{\sum_{r\in R_{t(j)}} h_r(t_{(j)})} \\ &= \frac{h_0(t_{(j)})exp(\beta' X_i)}{\sum_{r\in R_{(t(j))}} h_0(t_{(j)})exp(\beta' X_r)} \\ &= \frac{\exp(\beta' X_i)}{\sum_{r\in R_{(t(j))}} \exp(\beta' X_r)}, \end{aligned} \tag{2.93}$$

which does not depend on the baseline hazard $h_0(t)$. Cox [40] proposed that the partial likelihood is given as the product of those probabilities resulting

$$L = \prod_{j=1}^{m} \frac{\exp(\beta' X_i(t_{(j)}))}{\sum_{r\in R_{(t(j))}} \exp(\beta' X_r(t_{(j)}))}, \tag{2.94}$$

where $X_i(t_{(j)})$ is the vector of covariate for subject i at risk at time $t_{(j)}$ and no observation is censored. For censored information, the likelihood is expressed by

$$L = \prod_{i=1}^{n} \left[\frac{\exp(\beta' X_i(t_{(i)}))}{\sum_{r\in R_{(t(i))}} \exp(\beta' X_r(t_{(i)}))} \right]^{\delta_i}, \tag{2.95}$$

where $\delta_i = \begin{cases} 1 & \text{if not censored} \\ 0 & \text{if censored} \end{cases}$ is the event indicator for the ith observation and no ties are assumed in the occurrence of events.

Parametric Cox Proportional Hazard Model

Parametric formulation of the Cox proportional hazard model assumes that the baseline hazard rate, $h_0(t)$ follows one of the parametric distribution functions described in section 2.4.4 and the hazard function depends on a covariates vector X of p independent variables. Different hazard functions produce a specific family of parametric PH models.

For example, if one assumes an exponentially distributed survival time, the baseline hazard rate, $h_0(t) = \lambda$ is constant then the hazard rate of the PH model obtained is given as

$$h(t|x) = \lambda exp(\beta X), \tag{2.96}$$

where β is a vector of coefficients to be estimated and X is the covariate vector. This model is known as the exponential PH model. An important extension of the exponential PH model is the piecewise exponential model. In this model, the survival time is divided into intervals $(t_j, t_{j+1}]$, $j = 1, 2, \ldots, k$, $t_1 = 0$ and the constant baseline hazard varies across intervals. (See [25, 70, 71]).

If the Gomperts distribution is assumed, the baseline hazard rate is given as $h_0(t) = \lambda \exp(\theta t)$, and the hazard rate of the PH model is given by

$$h(t/x) = \lambda \exp(\theta t) exp(\beta' X). \tag{2.97}$$

For the final example, if the Weibull distribution is assumed, then the baseline hazard rate is given by $\lambda p t^{p-1}$ with $\lambda, p > 0$. The Weibull PH model's hazard function with covariates vector X is given by

$$h(t|x) = \lambda p t^{p-1} \exp(\beta' X). \tag{2.98}$$

To determine the suitability of the parametric model, a non-parametric estimate of the survival time is performed. For example, if the hazard function is approximately constant over time, the exponential distribution might be used; if the hazard function increases or decreases monotonically with increasing time, a Weibull or Gompertz model should be considered.

The advantage of the parametric assumption is that the survival function is smooth and the model is flexible. But the downside is that the model is not parsimonious and does not lead to easy interpretations. And also, one requires the parametric model to be a good fit to the survival life data and must be tested. The maximum likelihood method is used to estimate the parameters instead of the partial likelihood in the Cox model.

Accelerated Failure Time Formulation

A proportional hazards model assumes a multiplicative effect of the covariate on the hazard rate. In contrast, the accelerated failure time model assumes an accelerative effect of the covariate directly on the survival time instead of on the hazard.

The AFT model formulation proposes a regression model for the log-transform survival time using the covariate as follows:

$$\log(T) = \beta_0 + \sum_{i=1}^{p} \beta_i x_i + \sigma\xi = \beta X + \sigma\xi, \tag{2.99}$$

where σ is a scale parameter and ξ, the error term is assumed to follow any of the survival parametric density functions discussed. If ξ follows an extreme value distribution, then the survival time T has a Weibull distribution. Depending on the parametric form assumed for the ξ, the AFT models are named according to the distribution of T obtained. Therefore, there can be exponential AFT models, Weibull AFT models, log-logistic AFT models, Log-Normal AFT models, or the gamma AFT models. The survival function of T_i is obtained following random variable transformation

$$\begin{aligned} S(t) &= P(T_i > t) \\ &= P(\log(T_i) > log(t)) \\ &= P\left(\beta_0 + \sum_{i=1}^{p} \beta_i x_i + \sigma\xi > log(t)\right) \\ &= P(\beta X + \sigma\xi > log(t)) \\ &= P\left(\xi_i > \left(\frac{log(t) - \beta X}{\sigma}\right)\right) \\ &= S_\xi\left(\frac{log(t) - \beta X}{\sigma}\right). \end{aligned} \tag{2.100}$$

Table2.1 summarizes the distributions of some commonly used ξ and their corresponding distributions of T_i. Generally, the survival function and its corresponding hazard function are of the form

$$S(t|X) = S_o(t/\nu(X)), \tag{2.101}$$

and

$$h(t|\nu(x)) = \frac{h_0[t|\nu(x)]}{\nu(X)}. \tag{2.102}$$

where S_0 is the baseline survival function, $h_0()$ is the baseline hazard rate, and ν is an acceler-

ation factor. The acceleration factor is given as

$$\nu(x) = \exp\left(\beta_0 + \sum_{i=1}^{p} \beta_i x_i + \sigma\xi\right) = \exp\left(\beta' X + \sigma\xi\right). \tag{2.103}$$

Table 2.1: Summary of parametric AFT models

Distribution of ξ	Distribution of T
Extreme value(1 parameters)	Exponential
Extreme value(2 parameters)	Weibull
Logistic	Loglogistic
Normal	Lognormal
LogGamma	Gamma

The most common estimation method for the AFT models is the maximum likelihood method. The likelihood function of the n observed survival times, $t_1, t_2, \ldots, t_n$ is given by

$$L(\beta, \sigma) = \prod_{i}^{n} \left[f(t_i)\right]^{\delta_i} \left[S(t_i)\right]^{1-\delta_i}, \tag{2.104}$$

where $f_i(t_i)$ and $S_i(t_i)$ are the density and survival functions for the ith individual at time t_i and

$$\delta_i = \begin{cases} 1 & \text{if not censored} \\ 0 & \text{if censored} \end{cases} \quad \text{is the event indicator for the ith observation.}$$

Substituting $S(t_i) = S_\xi\left(\frac{log(t)-\beta' X}{\sigma}\right)$ as obtained from Equation 2.100, the log-likelihood function is then obtained as

$$\log\left[L(\beta, \sigma)\right] = \sum_{i=1}^{n} \left[\delta_i \log\left(\sigma t_i + \delta_i \log\left(f_{\xi_i}(z_i) + (1-\delta_i)\log\left(S_{\xi_i}(z_i)\right)\right)\right)\right], \tag{2.105}$$

where $z_i = \frac{log(t)-\beta' X}{\sigma}$ and the maximum likelihood estimates of the unknown parameters, β and σ are obtained using the Newton-Raphson algorithm.

Chapter 3

Intensive Care Unit-Step-down Unit System Congestion: To Premature Step-down or not?

Abstract

A Step-Down Unit (SDU) provides an intermediate Level of Care for patients from an Intensive Care Unit (ICU) as their acuity lessens. However, SDU congestion and upstream patient arrivals force ICU administrators to incur a cost, either in the form of overstays or premature step-downs. When the ICU and the SDU are congested, patient flow decisions are challenging. We develop two patient flow policies using a Markov decision process model to select actions to optimize the system's net health service benefit. One allows for premature step-down actions, and the other allows for patient rejection actions when the system is congested. Our simulation results using the optimal actions obtained from the MDP models show that the policy with patient rejection has a net health service benefit that significantly exceeds that of the policy with the premature step-down option. Furthermore, based on the results, it is observed that premature step-down action contributes to congestion downstream. Counter-intuitively, premature step-down should therefore be discouraged, and patient rejection or diversion actions should be further explored as viable options for congested ICUs.

Keywords: Healthcare, ICU, SDU, patient flow, policy, congestion, MDP

3.1 Introduction

Intensive care units (ICUs) provide care to patients with high levels of acuity. Patient acuity has often been measured by standardized scores such as SOFA and the variants of APACHE [88, 90] as well as nursing manpower scores, such as the "Nine Equivalents of Nursing Manpower" (NEMS) score [125, 113]. Staffing is typically one nurse per ICU patient, and ICU beds are rarely idle. During recovery, the continued need for intensive care (and consequently, an ICU bed) diminishes and this is reflected in a lower NEMS score. To provide better continuity of care, so-called Step-down Units (SDUs) are intended for these recovering patients [94, 101, 59], with staffing typically one nurse per two SDU patients. In a congested setting, patients with lesser acuity who continue to occupy ICU beds represent at best a sub-optimal use of resources, and at worst may prevent an arriving high-acuity patient from getting the care she requires.

The NEMS score is a scoring derived by Miranda et al. [113] from the therapeutic intervention scoring system (TISS) to determine the required levels of intensive care needs, provide information on the severity and prognosis of patients' acuity and determine the needed number of nurses and their workload. The NEMS score is a value between 0-56 points, the sum of nine (9) patient's related factors (see Table: 3.2) which have an influence on nurses' workload during the administration of care [31, 112, 167].

ICU beds and staffing represent a high operational cost for any hospital [65, 170, 142]. Therefore, with its highly valued care and the increasing demand, it is necessary to improve the ICU flow to optimize the system's throughput [8]. For a patient to be discharged or transferred from the ICU, a physician's declaration of the patient's medical stability is required [117, 119]. Determining the patients' suitability to leave the ICU often takes time. Also, if there are no available beds downstream, patients may occupy ICU beds longer than medically necessary [154]. Furthermore, ICUs avoid rejecting patients as there is often a risk of death if a patient is turned down or left untreated. As such, patient arrivals may trigger the step-down of a suspected lower acuity patient to free up a bed for the recent arrival. Motivated by these phenomena, we define two types of ICU patient transfers or 'step-down': regular step-down, during which sufficiently low acuity patients (LAP) are moved from the ICU to the SDU and a premature step-down during which a HAP is moved from the ICU to the SDU before reaching her intended medical stability.

In the case of increasing demand for intensive care and a congested ICU, management needs to decide between rejecting a new patient in need of critical care or prematurely stepping down a current occupant. Possible future scenarios, including demand surges due to new diseases, may put the ICU in a precarious situation. For example, increasing numbers of coro-

navirus disease 2019 (COVID-19) cases may create a surge in demand for hospital admission and critical care. In such a situation, when the ICU is full, and resources are constrained who receives the service: the newly arriving patient or the existing patient [11]?

In this study, we compared two patients flow policies. The first combines the following actions: reject or admit an arriving HAP to the ICU, step-down or retain existing LAP in the ICU, premature step-down or retain HAP in the ICU, and prematurely discharge or retain a LAP from the SDU. In Policy 2, however, whenever the system is congested and there is an arriving patient, an existing HAP is prematurely stepped down in other to admit the arriving patient instead of rejection in Policy 1. In this policy, premature step-down of existing patients and admission of an incoming patient has priority over rejection of the latter under congestion. We only quantify the variation in the system's health service benefit in a congested environment via a metric that reflects the benefice or detriment of action, i.e an action either increases or decreases the system's health service benefit. We aim to assess the impact of the two decision policies on patient flow in a congested environment. The main difference between the two policies is that the first perform premature step-downs of existing high acuity patients (HAP) to avoid rejection of arriving patients when the ICU is full, whereas once the ICU is full, the second reject arriving patients. We sought to optimize the long-term health service benefit of these policies. In our methodology, we assigned relative weights to each atomic action, built and solve a Markov decision model to obtain optimal actions that made a policy based on such weights. The optimal actions of the two are then analyzed for sensitivity and used as inputs to simulate the hospital management flow and compare the two policies under an increasing rate of arrival.

The remainder of the chapter proceeds as follows. Section 3.2 provides an overview of related literature, with keen attention to the use of operation research applications to patient flows in the ICU. Section 3.3 briefly describes the empirical data used in this research. Section 3.4 describes the simple solution methodology of the infinite horizon Markov decision process model we proposed, its sensitivity analysis and the simulation model. In Section 3.5, we present the result of the decision rules for each policy, the sensitivity analysis, and the simulation results. In Section 3.6, we discussed the implication of the results obtained and the chapter closes in Section 3.7 with the conclusions and recommendations.

3.2 Overview of Related Literature

ICU Patient flow and capacity planning have received a lot of attention in the operation research literature, even more so during COVID-19 pandemic [147, 37, 114]. Several papers study patient flow and capacity planning based on resources such as beds and staffing and its scarcity.

Comprehensive reviews of this literature can be found in Bai et al. [12], Lin et al. [98]. They found various stochastic (queueing models, MDP), deterministic (mathematical programming) and empirical (statistical) methods used in the literature to model ICU patient flow.

Here, we focus our review on a few key papers in the literature on the decision-making in an ICU under congestion. ICU flow decision-making process is complex and challenging. Azcarate et al. [11] remark that flow decisions guidelines in the ICU are hindered by the absence of clear and objective metrics to determine which patients will continue to benefit from critical care. Levin [95] observed that only a small number of ICUs use written patient discharge guidelines and follow an empirical decision process and that consensus, rather than empirical evidence, dictates the importance of guidelines and policies.

In a congested scenario, the practice may be to triage current ICU patients [152]. Like Levin [95], most literature suggested that discharging patients is a way to relieve congestion pressure in the ICU. At the same time, the lack of beds in other downstream units of the hospital can also cause discharge delays or overstays. A majority of the unsuccessful discharges from ICU were due to a lack of beds in the downstream or disagreement over admitting services in the wards [150, 95, 8].

Downstream congestion has also been observed to cause blocking in the ICU. It keeps patients from moving [38]. Mathews and Long [110] found that in the USA, ICU patients who are ready for transfer to a downstream unit often stay in the ICU for longer than clinically necessary. Most of such patients remain in a critical care bed and thereby delaying admission for other incoming patients. In their studies, Mathews and Long [110] examined varying discharge policies in times of capacity strain in the Emergency Department (ED). Shi et al. [144] develop a stochastic network queuing model with dynamic discharge policies, which reduce admission delays and ED wait times for admission to the ward in times of peak utilization.

Markov Decision Process (MDP) models have been used actively in recent years in hospital resource and inventory management in general, and ICU resources and service modelling in particular. Broyles et al. [27], Dobson et al. [47], Patrick et al. [123], Patrick [122], Chan et al. [34], Li et al. [97], Nunes et al. [118] all used discrete-time MDP to model discharge and admission decisions in the ICU with many dissimilarities in the models. One major difference is the state definition of the particular MDP model. Nunes et al. [118] presented an MDP model for elective (non-emergency) patient admissions to promote more efficient utilization of hospital resources, thereby preventing idleness or excessive use of these resources. Their solution approach was the value iteration algorithm. Their model was able to generate an optimal admission control policy that maintained resource consumption close to the desired levels of utilization. They however report difficulties using it since it is a complex model due to its stochastic dynamic and high dimensionality and requires the development of cus-

tomized solution methods. Patrick [122] analyzed several scenarios that explore the trade-off between patient-related measures (lead times) and physician- or system-related measures (revenue, overtime, and idle time). They did so by using a MDP model with a three-dimensional state: the current number of patients who can be booked in advance, the number of previously booked appointments and the new demand, to analyze several scenarios that explore the trade-off between patient-related measures (lead times) and physician- or system-related measures (revenue, overtime and idle time). Through simulation, the paper demonstrates that, over a wide variety of potential scenarios and clinics, the MDP policy does as well or better than open access in terms of maximizing net health service benefit as well as providing more consistent throughput. Chan et al. [34] examined priority demand-driven ICU discharge policies on patient mortality and total readmission load. They define the state to be the number of different types of patients in the ICU to reflect the total occupancy of the ICU. They designed an approximation algorithm and found optimal policies in certain regimes. Li et al. [96] applied an MDP approach to study the admission decisions. Their state definition is similar to Chan et al. [34]. but they combined both the number of different types of patients in the ICU and the number of available beds. Using this model, a lower and upper bound of the parameter was therefore found to evaluate the admission policy and improve the policy. Edbrooke et al. [51] examined the cost-effectiveness of ICU admission by comparing patients who were accepted into ICU after ICU triage to those who were not accepted while attempting to adjust such comparison for confounding factors. They found that not only does the ICU appear to produce an improvement in survival, but the cost per life saved falls for patients with greater severity of illness, that is timely ICU admission reduces 28-day mortality by 30%. Other studies demonstrate that delaying ICU admission can prolong ICU length-of-stay [32] and increase the risk of death [30]. Thus, a vicious cycle is born. Chronic beds shortages contribute to admission delays and longer wait times. This further increases the length-of-stay (LOS) and creates exacerbating beds shortages.

More recently, Li et al. [97] developed an analytical framework of an MDP model with the system state characterized by the number of two patient types in the ICU to quantify the impact of the number of reserved beds and suggest when to prematurely discharge current patients. After extensive numerical experiments were performed to analyze the effect of each parameter on the total survival benefits, the model is established to strike a balance between rejection of incoming patients and premature discharge.

There are metrics developed by the clinical community as systematic criteria to evaluate the patient health severity status. Rodrigues et al. [134] working with a large dataset from an academic hospital use a discrete event simulation to show the benefits of SDU beds in improving both patient flow and costs in a highly congested hospital based on a metric called the Nine

Equivalents of Nursing Manpower (NEMS). Shmueli et al. [145] used Acute Physiology and Chronic Health Evaluation II (APACHE II) to evaluate patient severity. The NEMS and the APACHE are based on the clinical observation of the patients and are generally assigned daily based on data available within an ICU stay. Strand and Flaatten [155] provided a review of several versions of three different prognostic scoring systems to review the severity metrics in the hospital. Kim et al. [84] estimated the cost of denied ICU care for all the medical patients admitted to 21 hospitals through the EDs. They empirically found that ICU congestion could have a significant impact on ICU admission decisions and patient outcomes.

The step-down and discharge process in the ICU-SDU system is currently prolonged beyond the acute care days of patients. The unavailability of empty beds in the SDU results in patients using the ICU even when they have no need of it [8, 94]. The congestion of the SDU contributes to that of the ICU and therefore produces an increased length-of-stay. When patients who do not need the ICU service stay in the ICU longer than the need, this prevents the admission of others who need it the most and exposes them to higher risk of mortality. The question worth asking here is what the hospital can do in terms of step-down policy planning to reduce not only length-of-stay but also increase health service benefit of patients who request ICU. To answer this question, we must first look into how the step-down process is performed currently.

Studies on the impact of SDUs are primarily limited to observational and simulation-based models with different objectives [111, 49, 133, 129]. Most studies to date recommend the SDUs as a safe care option for patient who does not need ventilation [129]. But, these studies have not been conclusive on the benefit of the SDU in reducing mortality. Many hospitals have used SDUs specifically as an alternative to full intensive care and this practice is thought to be an alternative level of care. Continued research and data collection from the SDU is needed and required in this arena to contribute to the development of the patient discharge process to characterize and completely specify the medical and physiological step-down to SDU policy, and to allow for comparison of outcomes across different units.

In this research, we will look at congestion from the last bed problem perspective. When the ICU is full how to decide between rejecting a new patient in need of critical care and creating a vacancy by prematurely discharging a current occupant? Azcarate et al [11] offer a review of literature on this clinical management dilemma with factors to consider and the patient health consequences of each decision. However, they noted that mathematical models of ICU management practices overlook these health factors and its consequences on patients. To the existing literature, we propose the determination of actual decisions to be taken in congestion instead of the determination of a certain threshold considering risk factors. The decisions about all the aspects of patients' flow are considered instead of focusing on a subset. We suggest

focusing on the congestion area instead of solving a bulky state space with the state that is not relevant for congestion. Finally, the result of our simulation model counter-intuitively suggests that rejection of HAP when the system is full is better than to the conventional practice of prematurely stepping down high acuity.

3.3 Data Description

Empirical data from the London Health Sciences Centre (LHSC) have been used to estimate the distribution function of the transition used for the simulation. The LHSC is a multi-site health care facility. It has two main hospitals: University Hospital and Victoria hospital which includes the Children's Hospital at London Health Sciences Centre. The data used is a set of a four-year record containing more than 70000 logs with nearly 8000 patients from January 2015 to December 2018. The Patient information includes patient age, gender, admitting diagnosis, admitting source, discharge destination, and daily NEMS scores till discharge. The NEMS is closely related to patient health because as the patient's health improves, less nursing attention is needed, resulting in a lower NEMS. Empirically, a score below 10 is considered to be a "Very Low Acuity" patient; scores between 11–25 would be "Low Acuity" patient, and from 26–56 a "High Acuity" (See Table 3.1) [134]. From Figure 3.1, it is observed that more than 95 % of patients requesting the ICU's service has their NEMS score higher than 25. It is therefore reasonable to assume that all patients requesting the ICU are HAP.

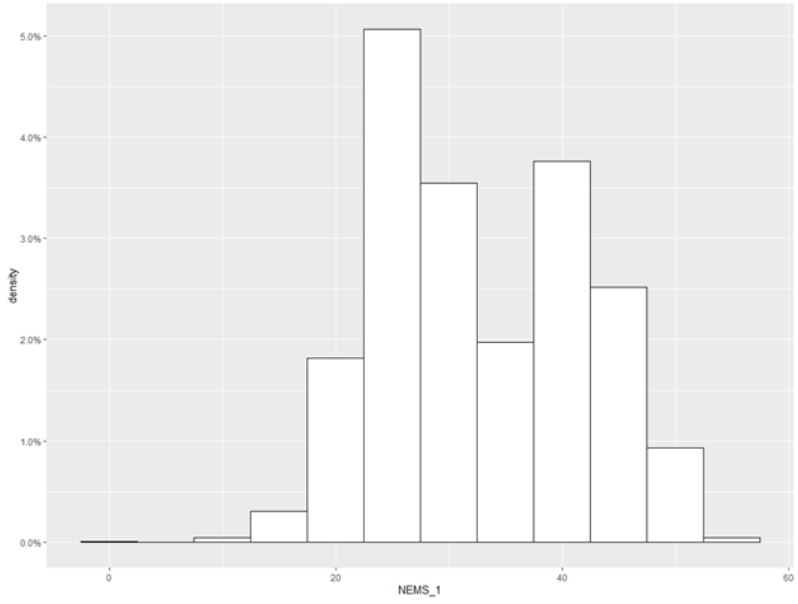


Figure 3.1: First day NEMS score of Victoria hospital patients

Table 3.1: Severity and Levels of Care Characteristics. (*Source:Rodrigueset al.* [134])

Level of Care	Beds characteristics	Patient per nurse ratio	Cost ($ /patient-day)	NEMS Score
1	Standard Ward beds: No organ support, no ventilation	3 or more to 1	$600	≤ 10
2	Step-down beds: Support single failed organ system, no ventilation	2 to 1	$2,000	11 to 25
3	Intensive care beds: and multiple organ support	1 to 1	$3,500	26 to 56

Table 3.2: NEMS Component (*Source:Mirandaet al.* [113])

Items	Points
1. Basic monitoring: hourly vital signs, regular record and calculation of fluid balance	9
2. Intravenous medication: bolus or continuously, not including vasoactive drugs	6
3.Mechanical ventilatory support: any form of mechanical/ assisted ventilation, with or without PEEP	12
4. Supplementary ventilatory care: breathing spontaneously through an endotracheal tube; supplementary oxygen any method, except if (3) applies	3
5. Single vasoactive medication: any vasoactive drug	7
6. Multiple vasoactive medications: more than one vasoactive drug, regardless of type and dose	12
7. Dialysis techniques: all	6
8. Specific interventions in the ICU: such as an endotracheal intubation, the introduction of a pacemaker, cardioversion, endoscopy, emergency operation in the past 24 h, gastric lavage; routine interventions such as X-rays, echocardiography, electrocardiography, dressings, the introduction of venous or arterial lines, are not included	5
9. Specific interventions outside the ICU: such as surgical intervention or diagnostic procedure; the intervention/procedure is related to the severity of illness of the patient and makes an extra demand upon manpower efforts in the ICU	6

Patients arrive at the ICU individually from different sources. The ED, a unit with varying patient severity provides the highest proportion of ICU patients. From the hospital studied, about 38% of the patients come from the ED, 22% from the ward, 21% from the Operating

room, and 20% from other places such as other hospitals or the SDU. The majority of the admission into the ICU are unplanned and requires immediate medical care. With the priority triage policy used in many hospitals, the hospital we study has very little, or no control over admitting HAP arriving through the emergency route. Daily arrivals are essentially equally distributed with Thursdays having the maximal admission. Hourly admission trend is also examined. Figure 3.4 is the distribution of the inter-arrival time of the patients. The average inter-arrival time is 6.47 hours and is approximately exponential. The system's capacity is 30. Figure 3.2 is the daily ICU occupancy in the year 2018. There is no evident trend available in the data. Nevertheless, occupancy is observed to be often higher than capacity. Congestion is a daily routine. In most cases, to offset overcapacity, patients are placed elsewhere at an alternative level of care (ALC). Figure 3.3 is the time plot of the various acuity level observed in the system. The evolution of the patients' acuity levels depends on the severity and the care they received.

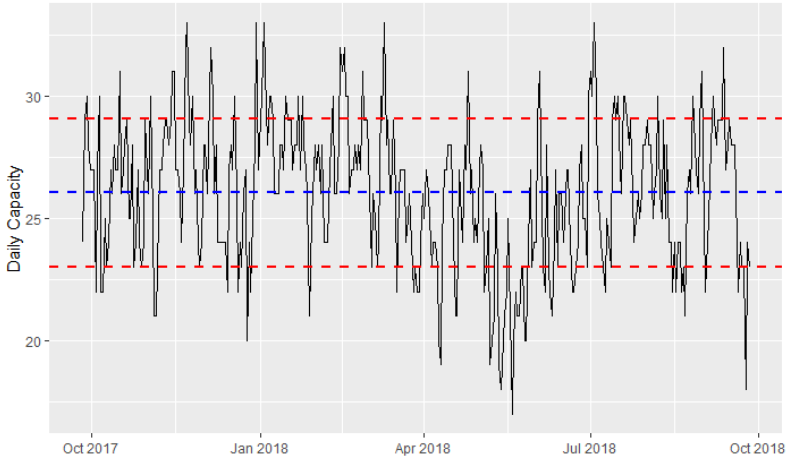


Figure 3.2: Time plot of daily Victoria hospital occupancy in 2018. (Blue dashed line represents the mean and the red dashed lines represent one standard deviation below and above the mean.)

The daily number of the various category of patients in the ICU is recorded. From that, we obtain the distribution of the number of patients that recover from one acuity level to the other in the system. Figure 3.5 is the density function of the number of patients that recovered from high acuity to low acuity at the ICU. These are patients with NEMS score between 11 and 25 that are destined to move from the ICU to the SDU. Figure 3.6 is the density function of the number of patients that recovered from low acuity to recovered in the system. These are patients with NEMS Score less than 10 that are moved from the SDU to the general ward. The patients that died or leave the ICU directly to the house are termed the discharged. Since the number of recovered patients is countable, and the number of occurrences is independent,

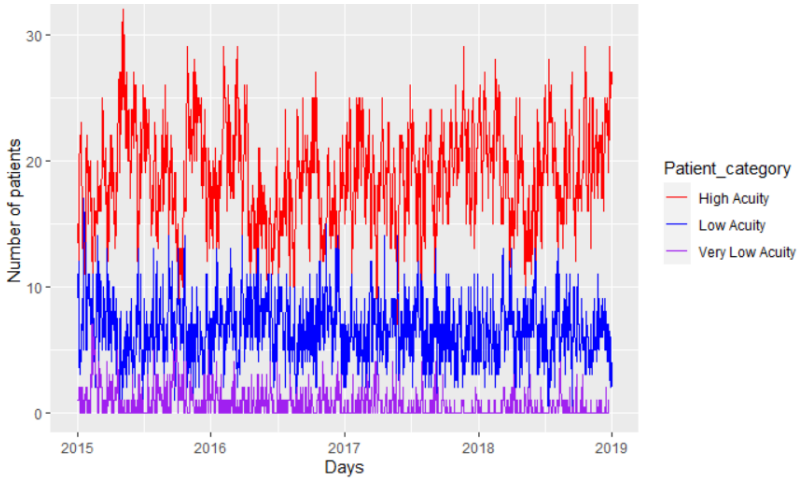


Figure 3.3: Time plot of daily number of acuity levels' occupancy at the Victoria hospital.

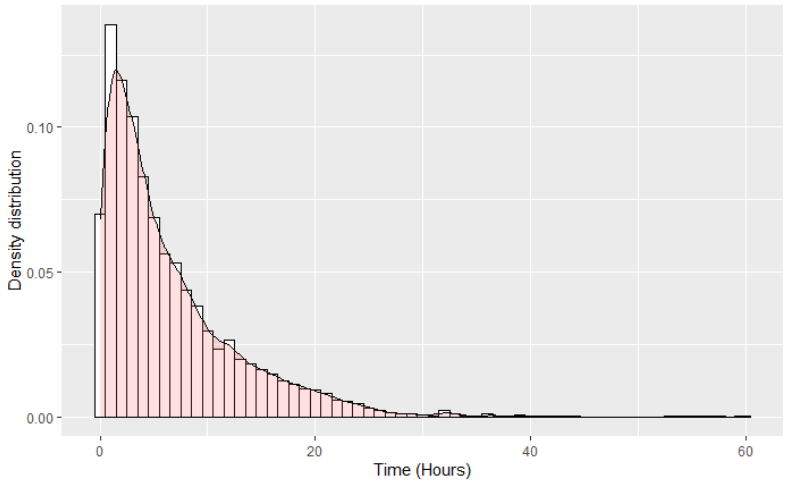


Figure 3.4: Density distribution of the system's inter-arrival time.

and the estimated average rates are approximately equal and independent of every occurrences, we can safely assume that the recovery processes follow a Poisson distribution with parameter estimates 2.45 and 3.22 for the high and low recoveries respectively. Table 3.3 provides the descriptive statistics of the recovery processes.

Table 3.3: Descriptive statistics of daily recovery process

	Mean	Var	Median	Min	Max	Skew	Kurtosis
High- Low	2.45	2.56	2	0	9	0.55	−0.058
Low-Recovered	3.22	3.67	3	0	11	0.60	0.19

On average, the daily transition matrix from one acuity level to another is tabulated in Table

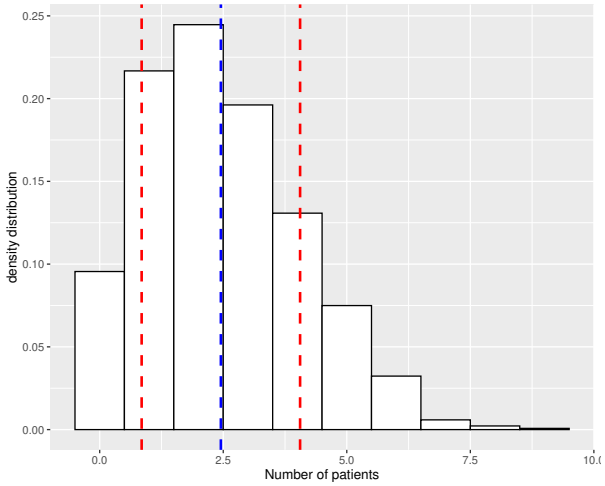


Figure 3.5: Daily distribution of the number of patients that move from high acuity to low acuity. (Blue line represents the mean and the red lines are the one standard deviations from the mean.)

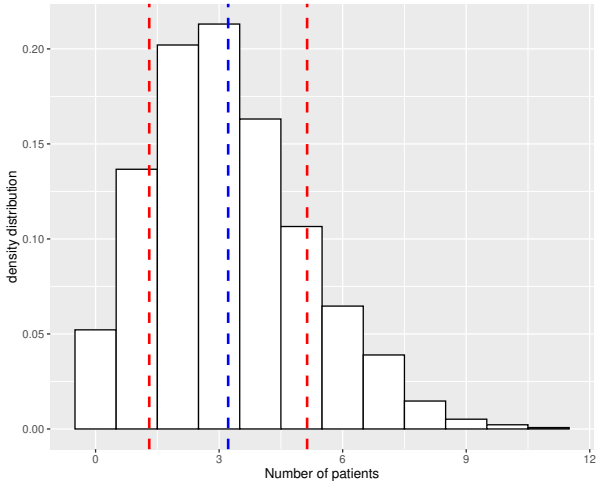


Figure 3.6: Daily distribution of the number of patients that move from low acuity to recovered. (Blue line represents the mean and the red lines are the one standard deviations from the mean.)

3.4 given as

Table 3.4: Average daily transition probability

	High	Low	Recovered	Discharged
High	0.7760	0.1360	0.0020	0.0690
Low	0.1505	0.4342	0.0040	0.346
Recovered	0.0700	0.2250	0.0899	0.6140
Discharged	0	0	0	1

3.4 Methodology

Patient flow through the SDU in most hospitals is assumed to come from two areas: either directly from the ICU, or as a direct entry if the ICU is full. Patients are admitted to the SDU directly from the ICU if they are deemed to be less acute for the ICU, but not so sick that they require ICU care; and alternatively, acute patients who cannot be admitted to the ICU due to congestion are admitted to the SDUs. A direct entry into the SDU is another example of an Alternative Level of Care (ALC). That is, patients are sometimes placed in the SDU before ICU care if the ICU is too congested to immediately admit the patient. Cady et al. [29] and Eachempati et al. [50] for example only admit post-ICU patients into their SDU, while others allow different admission patterns. Like Cady et al. [29] we assume in our setup that all SDU patients are post-ICU.

Patients arriving at the ICU are assumed to be HAP only and are admitted immediately to the ICU if there is a bed. In reality, some patients admitted to the ICU may have their NEMS score less than 25. But due to the presence of multiple factors described in Table 3.2, they are treated in the ICU. We assume only two types of patients in the ICU: high and low acuity. HAP have NEMS score greater or equal to 25 and the LAP have NEMS score less than 25 [134]. At a decision epoch, which is continuous, a patient flow decision has to be made. So when the ICU is full and a HAP arrives at the ICU, he or she is rejected or a less acute patient may be stepped down to make a space or a HAP may be prematurely stepped down. A critical patient who is admitted to the ICU will be treated until either she reaches a stable enough low acuity state to be stepped-down to the SDU or she dies and is discharged.

We propose a Markov decision process (MDP) to model the patient flow dynamics. Our objective is to maximize what we call "the net health service benefit" of the system's flow under a congested environment. The accumulation of the rewards and the costs associated with each of the actions under a policy gives the net health service benefit of a policy. Our goal is then to find the set of actions that maximizes the net health service benefit.

3.4.1 State Space and Action Set

At a decision epoch, $t \in [0, \infty)$ considered to be continuous, the decision-maker has a set of decisions to make: (i) Admit or reject an arriving patient to ICU if any (rejection can be an alternative level of care or an off service), (ii) step-down or retain a LAP to the SDU, (iii) prematurely discharge or retain recovering patient from the SDU to the ward and (iv) prematurely step down or retain a HAP from the ICU to the SDU (if the ICU is full and there is an arrival, premature may be considered). Since we are looking at decisions under congestion, the system state is defined in terms of the congestion zone or the last beds zone [11]. The

system's state is denoted by $s_t = (x_t^1, x_t^2, y_t, q_t)$ where $x_t^1 \in 0, 1, \ldots, B_I$ is the number of HAP occupying the B_I congestion zone of the ICU, $x_t^2 \in 0, 1, \ldots, B_I$ is the number of LAP occupying the B_I congestion zone of the ICU, $y_t \in 0, 1, \ldots, B_S$ is the number of LAP occupying the B_S congestion zone of the SDU, and $q_t \in 0, 1$ is the number of arriving patients.The state-space at time epoch t is defined as:

$$\mathcal{S} := \left\{ s_t = (x_t^1, x_t^2, y_t, q_t), 0 \leq x_t^1 + x_t^2 \leq B_I, 0 \leq y_t \leq B_S, 0 \leq q_t \right\}. \tag{3.1}$$

For each state $s_t \in \mathcal{S}$, $z_t = (z_{1t}, z_{2t}, z_{3t}, z_{4t})$, $\mathcal{A}(s_t)$ denote a feasible action that can be taken. And the action state is given as:

$$\mathcal{A}(s_t) := \{z_t = (z_{1t}, z_{2t}, z_{3t}, z_{4t}), z_{1t} \in \{0, 1\}, z_{2t} \in \{0, 1\}, z_{3t} \in \{0, 1\}, z_{4t} \in \{0, 1\}\} \tag{3.2}$$

$z_{1t} = 1$ denotes admission otherwise rejection; $z_{2t} =$ denotes step-down to the SDU, otherwise, retain; $z_{3t} = 1$ denotes discharge to the ward or otherwise retain; and, $z_{4t} = 1$ denotes premature step down otherwise retain. Every action must satisfy the capacity constraint expressed as

$$\begin{aligned} x_t^1 + x_t^2 + z_{1t} - z_{2t} - z_{4t} &\leq B_I \quad (ICU) \\ y_t + z_{2t} + z_{4t} - z_{3t} &\leq B_S \quad (SDU) \\ \forall t &\in [0, \infty). \end{aligned} \tag{3.3}$$

The actions are taken such that the system cannot accept more than its capacity. The rejection of arriving patients when the capacity is exceeded guarantees that.

If there are enough beds in the ICU, then the decision may seem simple, admit all arriving patients. If there is enough capacity in the SDU, then step down all LAP in the ICU. We know that if the ICU is not full, there is no need to discharge prematurely [97]. The main concern that motivates this research is the problem of congestion, or "the last bed" in both the ICU and the SDU. Rejecting patients is undesirable and may be impractical. If congestion exists in the ICU, patients that would have been admitted may wait in other hospital's units (e.g. ED, surgical wards, general wards) but maybe recorded as ICU patients. Keeping a patient in the ICU while that patient is supposed to be discharged to the SDU due to SDU congestion is also undesirable, since this patient consumes resources he or she is no longer in need of and preventing others from using those resources. It may also not be in the hospital's best interest to have too many idle beds. Since we are considering only congestion, instead of considering the whole system capacity of the units to build our state space, we consider only the congestion zone. For simplicity, we assume a toy example of the last two beds of the ICU and the last bed

of the SDU correspond to the congestion zone so that the problem will become the last bed problem in the medical literature [48, 135, 11].

Table 3.5 describes the state space. The first policy allows admit or reject, step-down or retain a LAP in the ICU, and premature discharge or not from the SDU. The actions are coded as tabulated in Table 3.6. A feasible combination of these actions is described in Table 3.7. The second policy allows in addition to the actions of Policy 1, premature step-down of HAP. The actions are coded as shown in Table 3.8. A feasible combination of these values gives us a complete description of the action space for the first case as shown in table 3.9.

Table 3.5: Description of state space of the system with arrival, high acuity patients (HAP), low acuity patients (LAP) in the intensive care (ICU) or step-down unit (SDU).

State number	HAP ICU (x_1)	LAP ICU (x_2)	LAP SDU (y)	Arrival (q)
1	0	0	0	0
2	0	1	0	0
3	0	2	0	0
4	1	0	0	0
5	1	1	0	0
6	2	0	0	0
7	0	0	1	0
8	0	1	1	0
9	0	2	1	0
10	1	0	1	0
11	1	1	1	0
12	2	0	1	0
13	0	0	0	1
14	0	1	0	1
15	0	2	0	1
16	1	0	0	1
17	1	1	0	1
18	2	0	0	1
19	0	0	1	1
20	0	1	1	1
21	0	2	1	1
22	1	0	1	1
23	1	1	1	1
24	2	0	1	1

Table 3.6: Actions in the Policy 1

Action	Value	Description
a1	0,1	Admission of high acuity to ICU
a2	0,1	Step-down of low acuity to SDU
a3	0,1	Discharge of low acuity out of SDU

Table 3.7: Feasible actions under the Policy 1

Action number	(z_1)	(z_2)	(z_3)
1	0	0	0
2	0	0	1
3	0	1	0
4	0	1	1
5	1	0	0
6	1	0	1
7	1	1	0
8	1	1	1

Table 3.8: Actions in the Policy 2

Action	Space	Description
a1	0,1	Admission of high acuity patients (HAP) to ICU
a2	0,1	Step-down of low acuity patients (LAP) to SDU
a3	0,1	Discharge of LAP out of ICU
a4	0,1	Premature Step-down of HAP

Table 3.9: Feasible actions under the Policy 2

Action number	(z_1)	(z_2)	(z_3)	(z_4)
1	0	0	0	0
2	0	0	0	1
3	0	0	1	0
4	0	0	1	1
5	0	1	0	0
6	0	1	0	1
7	0	1	1	0
8	0	1	1	1
9	1	0	0	0
10	1	0	0	1
11	1	0	1	0
12	1	0	1	1
13	1	1	0	0
14	1	1	0	1
15	1	1	1	0
16	1	1	1	1

3.4.2 Health Service Benefit Rewards and Costs

The reward and cost structure of this study is defined as the health service benefit of an individual in the hospital. Since ICU managers desire to serve the HAP in the ICU and the LAP in the SDU, the reward comes from two sources: admitting patients in the ICU is given a higher

reward and stepping down a LAP from the ICU to the SDU to accommodate HAP. A reward, r_h is associated with every admission and a reward, r_l is associated with every natural step-down. Natural discharges are not rewarded in order not to double reward. The reward, r_l, is viewed as the profit of not using the ICU with higher cost and using the SDU with lesser cost but with the same result. It is considered the difference between the cost of rejecting an arriving patient to the ICU and the cost of the current patient using the ICU for a full recovery. The undesirable events are then seen as the elements that contribute to the cost. Rejecting patients, premature step-down, keeping a LAP in the ICU and premature discharge of the patient that has not fully recovered all contribute to the cost of health service benefit to individual patients in the hospital. A cost, c_h is associated with every HAP rejected; a cost, c_l is associated with every overstay of a LAP in the ICU; a cost and a c_p, is the cost of a premature step-down out of the ICU. The cost of a premature step down represents the cost of incomplete service at the ICU. The reward associated with the natural step down is considered the difference between the cost of rejecting an arriving patient to the ICU and the cost of the current patient using the ICU for a full recovery. The premature step-down cost is the cost of incomplete service at the ICU.

Health service benefit rewards are gains and health service benefit costs are losses. The accumulation of the rewards and the costs defined above produces the net health service benefit. This research aims to obtain the decision structure to maximize the discounted net health service benefit over a quarter. The structure of the net health service benefit can be written as:

$$r_{(s_t,z_t)} = r_h(s_t, z_t) + r_l(s_t, z_t) - c_h(s_t, z_t) - c_l(s_t, z_t) - c_p(s_t, z_t) \tag{3.4}$$

We assume that the cost of ICU refusal is equal in absolute terms to the reward of ICU use. Given that the SDU is poorly equipped and less monitored, we assumed that its absolute service cost is less than that of the ICU. Rejecting a patient has a higher negative effect compared to prematurely stepping down a patient. Likewise, premature step-down has a higher negative effect compared to premature discharge from the SDU. For our toy example, the baseline values for the computation are set as follows: the reward for admitting a patient is set to 100, the reward for stepping down a patient is 25, the cost for rejecting a patient is 100, the cost for overstay in the ICU is 50, and the cost for premature step-down of a HAP is set to 80. These values are chosen as relative weights of the consequences of each action. A measure of the effect of these actions is difficult. Since such research will be unethical. The idea is to start with a naive relative weight for each action. Thus, we performed sensitivity analysis of these values for its robustness.

3.4.3 Value Function and Transition Probability

The value function estimates how good it is for the decision-maker to perform a given map of actions to the state. That is a measure of each policy. A policy π is a distribution of a set of the feasible action in each state (a mapping from each state, $s \in \mathcal{S}$, and action, $z \in \mathcal{A}$, to the probability $p_{(s,z)}$ of taking action z when in state s). In other words, it is a mixed policy. The value function in state s under a policy π at time t, denoted $V_\pi(s_t)$, is the expected long run of the discounted rewards when starting in s and following the policy π thereafter. It is defined by

$$V_\pi(s_t) = \mathrm{E}_\pi\left(\sum_{i=0}^{T} \lambda^i r(s_{t+1}, z_t/s_t)\right) \tag{3.5}$$

where $r(s_{t+1}, z_t/s_t)$ is the reward to-go function at time $t+1$ given action z is taken in state s at time t and T is the time horizon. When $T = \infty$, we have an infinite time horizon. We considered the infinite horizon because our decision epochs happen continuously and arrivals occur randomly. The optimal value function of the MDP model specifies the maximal expected reward over the infinite horizon for each state and satisfies the Bellman's optimality equation [130] for all $s_t \in \mathbf{S}$ defined by

$$\mathrm{v}^*_\pi(s_t) = \max_\pi \left\{ r(s_t, z_t) + \lambda \sum_{s_{t+1}\in\mathcal{S}} P(s_{t+1}/s_t, z_t)\mathrm{V}_\pi(s_{t+1}) \right\} \tag{3.6}$$

Our objective is to determine the optimal policy π^* of the MDP. This policy specifies the distribution of the actions that optimize the value function for each state and is given by

$$\pi^*(s_t) = \arg\max_\pi \left\{ r(s_t, z_t) + \lambda \sum_{s_{t+1}\in\mathcal{S}} P(s_{t+1}/s_t, z_t)\mathrm{V}_\pi(s_{t+1}) \right\}$$

$P(s_{t+1}/s_t, z_t)$ is the probability of transiting from state s to another when action z is taken at time t. The transition probability described the interactive combination of the progression of patients' health status from one acuity level to another, death, the random arrival, and the action taken and is given by

$$P(s_{t+1}|s_t, z_t) = Pr(x^1_{t+1} = x^1_t - r_t + z_{1t} - z_{4t} - d_t, x^2_{t+1} = x^2_t + r_t - z_{2t}, y_{t+1} = y_t + z_{2t} - r_{2t} - z_{3t}) \tag{3.7}$$

where $P(x^1_{t+1} = x^1_t - r_t + z_{1t} - z_{4t} - d_t, x^2_{t+1} = x^2_t + r_t - z_{2t}, y_{t+1} = y_t - z_{3t})$ is the probability that at time $t+1$, the number of HAP patients in the ICU is $x^1_{t+1} = x^1_t - r_t + z_{1t} - z_{4t} - d_t$, the number of LAP in the ICU is $x^2_{t+1} = x^2_t + r_{1t} - z_{2t}$, and the number of LAP in the SDU, where

r_t is the number of people who recover from the x_t, and r_{2t} is the number of natural recoveries who left the SDU. At time epoch t, the system state is $\left(x_t^1, x_t^2, y_t, q_t\right)$. Between time epoch t and $t+1$, four processes influence the transition of the state. Arrivals q_t, recovery of r_t^1 HAP to LAP in the ICU, complete recovery of r_t^2 LAP that move out of the system at time $t+1$, and dead patients, d_t. We assumed that only HAP die. The system state at time $t+1$ becomes $\left(x_{t+1}^1 = x_t^1 - r_t + z_{1t} - z_{4t} - d_t, x_{t+1}^2 = x_t^2 + r_t - z_{2t}, y_{t+1} = y_t + z_{2t} - r_{2t} - z_{3t}\right)$. This is depicted in the diagram below.

$$
\begin{array}{lll}
Time & t & t+1 \\
 & & \\
 & x_t^1 & x_{t+1}^1 = x_t^1 - r_t + z_{1t} - z_{4t} - d_t \\
State(s) & x_t^2 & x_{t+1}^2 = x_t^2 + r_t - z_{2t} \\
 & y_t & y_{t+1} = y_t + z_{2t} - r_{2t} - z_{3t}
\end{array}
$$

The transition probability is obtained from four random events: the arrival, the recovery in the ICU, the recovery in the SDU, and the dead. q_t arrivals occur at time t with probability $Pr(q_t)$. This random arrival is independent of patients' recovery rate and the hospitals' management. r_{1t} recovered from high acuity to low acuity with probability $Pr(r_{1t})$, r_{2t} recovered from low acuity to recovered with probability $Pr(r_{2t})$, and d_t death occur with probability $Pr(d_t)$. Since these four processes are independent, we can safely approximate the transition probability as:

$$P(s_{t+1}|s_t, z_t) = Pr(r_{1t})Pr(r_{2t})Pr(q_t)Pr(d_t) \tag{3.8}$$

and the distribution of each of these events is estimated from data as shown in Table 3.4 in Section 3.3. Since a Poisson distribution is the limit of a binomial distribution with parameter $p = \lambda/n$, where λ is the Poisson rate, and n, the number of trials approaches infinity. Since we are considering continuous-time epochs, we can assume that at most one patient may arrive, at most one patient may recovery from high acuity to low, at most one patient may recover from low to recovered, and at most one patient will arrive. Each process can then be a Bernoulli process with parameter p. Because the assumption on n is subjective, if we assumed that the probability of an event occurring or not is equi-probable. Hence, the transition probability from state s_t to state s_{t+1} is $P(s_{t+1}|s_t, z_t) = \left(\frac{1}{2}\right)^{r_{1t}+r_{2t}+q_t+d_t}\left(\frac{1}{2}\right)^{4-(r_{1t}+r_{2t}+q_t+d_t)}$ with $r_{1t}, r_{2t}, q_t,$ and $d_t \in \{0, 1\}$. In the computation, we use the empirical probabilities in Table 3.4.

z_{1t} is admitted, z_{2t} is stepped down to the SDU, z_{3t} is prematurely discharged, and z_{4t} is prematurely stepped down from the ICU at time t. Our objective is to determine the weight of every action in every state and use the action with the maximum weight to proxy optimal action in the various state.

Several methods have evolved for solving MDPs and dynamic processes in general. Solution methods for infinite-horizon problems use policy iteration, value iteration and linear programming while finite-horizon problems are mostly solved backwards induction algorithms. In this paper, we use the linear programming method developed by Schweitzer and Seidmann [141] with more recent expansions by Adelman [2], De Farias and Van Roy [42], Puterman [130] to solve the MDP due to its simplicity and easy reproducibility. Bertsekas and Tsitsiklis [19], Powell [127], Manne [106], Adelman [2], De Farias and Van Roy [42], Puterman [130] are extensive literature on linear programming methods for solving MDPs. From Powell [127] and Puterman [130], we know that, if $v(s) \geq \max_a(R(s,z) + \sum_{s \in S} P_{(s'|s,z_i)} v(s'))$ where $P_{(s'|s,z_i)}$ is the transition probability, then $v(s)$ is an upper bound on the value of being in each state. This means that the optimal value function can be obtained, and the optimal actions determined by backward induction by solving the following linear program

$$\begin{aligned} &\min_{v} \sum_{z_t \in \mathcal{A}} d(s_t) v(s_t, z_t) \\ s.t. \quad &v(s_t, z_t) \geq R(s_t, z_t) + \lambda \sum_{j \in \mathcal{S}} p(j|s_t, z_t) v_\pi(s_t, z_t), \ \forall s_t \in \mathcal{S}, z_t \in \mathcal{A}. \end{aligned} \tag{3.9}$$

Where $d(s_t)$ is any positive value. Alternatively, the solution of the dual of Equation 3.9 shown in Equation 3.10 provides the distribution of the actions in each state [43].

$$\begin{aligned} &\max \sum_{s_t} \sum_{z_t} R(s_t, z_t) W(s_t, z_t) \\ s.t. &\sum_{z_t} W(j, z_t) - \sum_{s_t} \sum_{z_t} p(j|s_t, z_t) W(s_t, z_t) \leq d(j), \ \forall j \in \mathcal{S}. \end{aligned} \tag{3.10}$$

Where the normalized $W(s_t, z_t)$, $(s_t \in S\,, z_t \in \mathcal{Z})$ are interpreted as the steady-state probabilities that action z_t is applied when the system visit state s_t at the typical transition. There are in total $\#\,\mathcal{S}$ constraints, where $\#\,\mathcal{S}$ represents the total number of states in the states space. The cost function $\sum_{s_t} \sum_{z_t} R(s_t, z_t) W(s_t, z_t)$ represents the steady-state average reward per transition. From Ross [136] and Wang et al. [168] by strong duality, we know that the optimal objective value of the dual LP equals the optimal objective value of the primal LP. Therefore, given a solution to the dual, the optimal action can be approximated directly by a much simpler transformation as

$$z_t^* = \arg\max_{z_t \in \mathcal{A}_{s_t}} \left\{ \frac{W(s_t, z_t)}{\sum_{z_t} W(s_t, z_t)} \right\} \tag{3.11}$$

where $\mathcal{A}_{s_t}$ is the set of all the actions possible in state s_t.

3.5 Results

3.5.1 Optimization Results

As described above, we solved equation 3.10 and computed the values of the probabilities of each of the actions in every state. The action with the maximum probability in the state is the optimal action. We did so in R using lpSolveAPI package, a 64 bit laptop with 32 GB memory and an Intel(r) Core i7-4600U CPU at 2.69 GHz.

The optimization and the computation were done as described in Sections 3.5 and 4.1. Table 3.10 exhibits the optimal actions in each state for Policy 1. In states without arrival, the possible actions are reduced to step-down or retain LAP in the ICU to the SDU and/or discharge or retain a LAP from the SDU. Discharges occur only when there is at least one LAP in both the ICU and the SDU and this is when there is an arrival or not. Every discharge out of the SDU has been triggered by a step-down from the ICU, and the presence of two LAP in the system. Either the two LAP are in the ICU or there is one in the ICU and one in the SDU. Whenever there is a space in the SDU and there is a LAP in the ICU, a step-down is triggered. Once there is an arrival if there is a LAP in the ICU, a step-down is triggered. Whenever there is an arrival and the ICU has an empty bed, we admit. In the state where we have the two ICU beds taken by low acuity patients, the SDU is empty and there is an arrival, the recommended action is to admit and step down. In general, accept arrivals when you can, step down when you can, and discharge when needed.

Table 3.11 describes the optimal actions selected in the various states under Policy 2. In states where there is no arrival, the possible actions are reduced to step-down LAP from the ICU to the SDU and/or discharge a LAP from the SDU and/or premature discharge of a high acuity patient. It can be observed that whenever there is no arrival, a premature stepping down is not necessary. We note that the only state in which a premature stepping down is allowed is state 18, i.e. when all ICU beds are occupied, one SDU bed is available and there is an arrival, one of the ICU patients is prematurely stepped down and the arriving HAP admitted.

Whenever there is at least two LAP in the system and a LAP occupying the SDU bed, a discharge is triggered by a stepping down of a LAP in the ICU. Admission is denied when the ICU is full and the SDU is also full. If all the patients in the ICU are HAP, we may have a HAP in the SDU as well. That is the patient occupying the last bed in the SDU may also be a high acuity patient. In general, it is better to allow the recovering patient in the SDU to recover than to discharge him, with a cost, and admit an arriving patient in the SDU with another cost.

State	High acuity ICU (x_1)	Low acuity ICU (x_2)	Low acuity SDU (y)	Action taken without arrival	Action taken with arrival
1 /13	0	0	0	admit 0, step-down 0 and discharge 0	admit 1, step-down 0 and discharge 0
2 /14	0	1	0	admit 0, step-down 1 and discharge 0	admit 1, step-down 1 and discharge 0
3 /15	0	2	0	admit 0, step-down 1 and discharge 0	admit 1, step-down 1 and discharge 0
4 /16	1	0	0	admit 0, step-down 0 and discharge 0	admit 1, step-down 0 and discharge 0
5 /17	1	1	0	admit 0, step-down 1 and discharge 0	admit 0, step-down 1 and discharge 0
6 /18	2	0	0	admit 0, step-down 0 and discharge 0	admit 0, step-down 0 and discharge 0
7 /19	0	0	1	admit 0, step-down 0 and discharge 0	admit 1, step-down 0 and discharge 0
8 20	0	1	1	admit 0, step-down 1 and discharge 1	admit 1, step-down 1 and discharge 1
9 / 21	0	2	1	admit 0, step-down 1 and discharge 1	admit 1, step-down 1 and discharge 1
10 / 22	1	0	1	admit 0, step-down 0 and discharge 0	admit 1, step-down 0 and discharge 0
11 /23	1	1	1	admit 0, step-down 1 and discharge 1	admit 1, step-down 1 and discharge 1
12 /24	2	0	1	admit 0, step-down 0 and discharge 0	admit 0, step-down 0 and discharge 0

Table 3.10: Policy 1 Comparative Optimal Decision in States with Arrival and without Arrival

State	High acuity ICU (x_1)	Low acuity ICU (x_2)	Low acuity SDU (y)	Action taken without arrival	Action taken with arrival
1/13	0	0	0	admit 0, step-down 0 discharge 0 and pre step-down 0	**admit 1**, step-down 0 discharge 0 and pre step-down 0
2/14	0	1	0	admit 0, step-down 1 discharge 0 and pre step-down 0	admit 1, step-down 1 discharge 0 and pre step-down 0
3/15	0	2	0	admit 0, step-down 1 discharge 0 and pre step-down 0	admit 1, step-down 1 discharge 0 and pre step-down 0
4/16	1	0	0	admit 0, step-down 0 discharge 0 and pre step-down 0	admit 1, step-down 0 discharge 0 and pre step-down 0
5/17	1	1	0	admit 0, step-down 1 discharge 0 and pre step-down 0	admit 1, step-down 1 discharge 0 and pre step-down 0
6/18	2	0	0	admit 0, step-down 0 discharge 0 and pre step-down 0	admit 1, step-down 0 discharge 0 and pre step-down 1
7/19	0	0	1	admit 0, step-down 0 discharge 1 and pre step-down 0	admit 1, step-down 0 discharge 0 and pre step-down 0
8/20	0	1	1	admit 0, step-down 1 discharge 1 and pre step-down 0	admit 1, step-down 1 discharge 1 and pre step-down 0
9/21	0	2	1	admit 0, step-down 1 discharge 1 and pre step-down 0	admit 1, step-down 1 discharge 1 and pre step-down 0
10/22	1	0	1	admit 0, step-down 0 discharge 0 and pre step-down 0	admit 1, step-down 0 discharge 0 and pre step-down 0
11/23	1	1	1	admit 0, step-down 1 discharge 1 and pre step-down 0	admit 1, step-down 1 discharge 1 and pre step-down 0
12/24	2	0	1	admit 0, step-down 0 discharge 0 and pre step-down 0	admit 0, step-down 0, discharge 0 and pre step-down 0

Table 3.11: Policy 2 Comparing Optimal Decision in State with Arrival to States without Arrival

3.5.2 Sensitivity analysis of the costs and rewards

A sensitivity analysis was undertaken to check how changes in the baseline costs and rewards change the optimal action in the different states. The base parameters of the cost/reward used in the model are as follows. The reward for admitting a patient is 100, the reward for stepping down a patient is 25, the cost for rejecting a patient is 100, the cost for overstay in the ICU is 50, the cost for prematurely discharging a patient is 25 and the cost of premature step-down of a patient is 80. In general, the actions are robust to rewards and costs associated with each action within the neighbourhood. A large variation of the cost and reward parameters is necessary for a fundamental change in the decision. For example in state (0,2,0), i.e. 0 high acuity in the ICU, 2 low acuity in the ICU and zero low acuity in the SDU, using Policy 1, rewards must be increased from 100 to 800 before we observe a change in the action from action (0,1,0) i.e. admit 0, step-down 1 and discharge 0 to action (0,0,0) i.e. admit 0, step-down 0 and discharge 0 (see Table 3.12 first row).

In Policy 1, assuming that stepping down reward and the cost of overstay is always less than the rejection cost. With that assumption, the model is always robust to stepping down reward and the cost of overstay. And only the admission reward or rejection cost have the same change as summarized in Table 3.12. Note that the Tables summarise only states with a variation. When we are in state (1,1,0), i.e. one high acuity, one low acuity in the ICU last bed and no patient in the SDU last bed, the baseline action is to do nothing, but when the admission reward increased from 100 to 1000, the replacement action becomes stepping down the LAP from the ICU to the SDU. When the admission reward or the rejection cost increases, the system tends to perform fewer step-down actions. In Policy 2, increasing discharge cost while keeping all other rewards and cost constant significantly affects decisions in only two states mainly. In states (0,2,1,1) and (1,1,1,1), action (0,1,1,1) is replaced by action (0,0,0,0). In other words, when there are at least two LAP in the system, and the discharge cost is high, the optimal action recommended is to do nothing. Table 3.13 summarizes the variations observed when we increase the reward of admission, the cost for rejection and the step-down reward. Increasing the cost of overstay while keeping all other rewards and cost constant affects two states. In state (1,0,0,0), the action changed from action (0,0,0,0) to action (0,0,0,1). In state (1,0,0,0), the decision changed from action (0,0,1,0) to action (0,0,0,0). In state (1,0,0,1), the decision changed from action (1,0,0,0) to action (0,0,0,0). When the ICU is full and the SDU is full, no matter how we decrease the cost, premature step-down is not a better option to choose (See Table 3.9 for actions). In state (2,0,0,1), when the cost increased, it is not a better option to premature step-down. In state (1,0,0,0), when the cost is low, we can afford to premature step-down and admit once there is a space in the SDU. In state (1,0,0,0) even though the ICU is not full, when the cost of premature step-down is low, it is recommended to step down.

Table 3.12: Sensitivity summary of Policy 1.

States	Baseline actions	Threshold	Replacement actions
(0,2,0)	(0,1,0)	800	(0,0,0)
(0,0,1)	(0,1,0)	1000	(0,0,0)
(1,1,0)	(0,0,0)	1000	(0,0,1)
(0,2,1)	(0,1,1)	1000	(0,0,1)
(1,1,1)	(0,1,1)	1000	(0,0,1)
(0,1,1)	(1,1,0)	1000	(1,0,0)

Table 3.13: Sensitivity summary Policy 2.(AR: Admission Reward, RC: Rejection Cost, SR: Step-down Reward)

Parameters	States	Baseline action	Replacement action	Threshold
AR	(1,0,0,0)	(0,0,0,0)	(1,0,0,0)	250
	(2,0,0,0)	(0,0,0,0)	(1,0,0,0)	500
	(1,0,1,0)	(0,0,0,0)	(0,0,0,1)	1000
	(1,0,0,1)	(0,1,0,0)	(1,0,1,0)	1500
	(0,0,1,1)	(0,1,0,0)	(0,1,0,1)	500
SR	(0,0,1,0)	(0,0,0,1)	(0,0,0,0)	1000
	(2,0,1,0)	(0,0,0,0)	(1,0,0,1)	1000
RC	(2,0,0,0)	(0,0,0,0)	(1,0,1,0)	1000
	(0,2,1,0)	(0,0,1,1)	(0,0,0,0)	1000
	(1,1,1,0)	(0,0,1,1)	(0,0,0,0)	1000
	(1,0,0,1)	(0,1,0,0)	(1,0,1,0)	2000
	(0,1,1,1)	(0,1,1,1)	(0,1,0,0)	2000
	(2,0,1,1)	(0,0,0,0)	(0,0,0,1)	2000

3.5.3 Simulation

The model is built using Simul8 6.0 (See Figure 3.8). The software was chosen for its availability, flexible coding, simplicity, and its interactive display of sequential events. In the simulation frame, there are 30 ICU beds and 12 SDU beds. Decision epochs are continuous. The optimal decisions are triggered only when the system is in the congestion zone. Arrivals follow a Poisson distribution with a rate of 6.3 patients/day (Approximation from the hospital data, see Figure 3.4). As estimated in Section 3.3, the recovery process from High acuity to Low acuity follows a Poisson distribution with a rate of 2.45 patients/day, recovery from low acuity to recovered is a Poisson process with a rate of 3.22 patients/day, the discharge process from the system to elsewhere also follows a Poisson distribution with a rate of 4.47 patients/day, and the death process is also Poisson with a rate of 1.24 patients/day. Decision-making is a continuous process triggered by any of the previous processes. When there is space in the ICU, an arriving patient is automatically admitted. The internal decisions in the ICU and SDU are coded into

the system as tabulated in Table 3.10 or Table 3.11 depending on the policy considered. In the ICU, the LAP are then made distinct from the HAP. If there is space in the SDU, the LAP are moved to the SDU.

Two methods were used to validate the model: Experts face validation and the comparison of parameter estimates and the results of the simulation without the implementation of the optimal decisions. The average throughput is within 5.78% of the empirical estimate. Total LOS is lower but with a standard deviation of 3.4, resulting in no statistically significant difference compared to the empirical data (p-value > 0.05). The mean LOS was found to be within 9.27% of the empirical. Figure 3.7 is the plot of the superimposed distributions of the empirical and simulated LOS. Given that the empirical data contains some special patients (about 0.07% of the data) who have used the ICU for more than two months, we considered those observations as outliers and removed them from our analysis.

Indicator	Low 95 %	Simulation Average	Up 95%	Emp data Estimate	Difference
Throughput (Patients/year)	1332	1366	1399	1287	-5.78%
Average LOS (days)	4.04	4.11	4.18	4.53	-9.27%
Standard Deviation (days)	3.354	3.4	3.447	4.13	-17.67%

Table 3.14: Comparative Patients Performance Measures

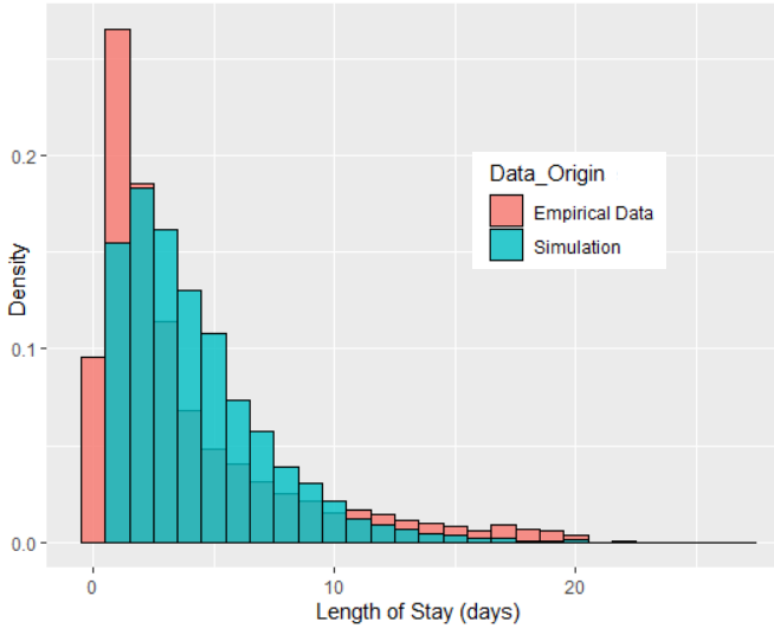


Figure 3.7: length-of-stay Distribution

Each policy was run for ten months duration with 300 replications of 50000 trials. We reported only records of the last four months of the simulation to have stable results. The

number of replications was recommended by the replication calculator, a function embedded in Simul8. We set the precision to a 95% confidence interval. A different random seed was used for each of the runs. We examine the costs incurred, the number of patients rejected and the number of patients prematurely stepped down under the two scenarios to find and compare the performance of the decision policies.

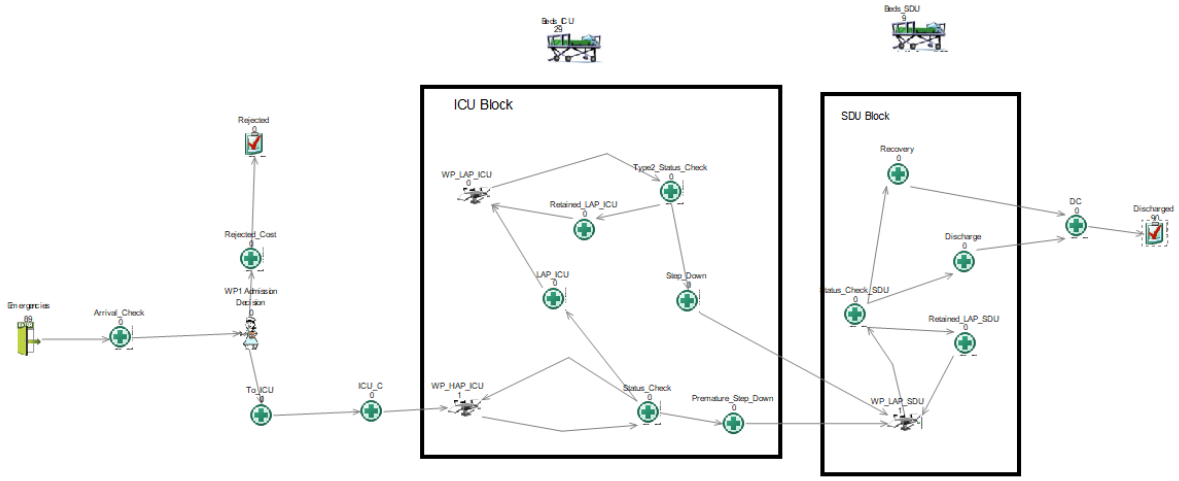


Figure 3.8: Screenshot of Simulation in Simul8

3.5.4 Simulation Results

The estimated performance indicators of the system when the arrival rate is 6 patients per day (from empirical data) are tabulated in Tables 3.14 and 3.16. From Table (3.14), the results suggest that on average, no patient is rejected with this average arrival rate. Policy 2 prematurely stepped down on average about 47% (253) of the patients admitted causing an average overstay of 10% (55) patients. Conversely, Policy 1 overstayed only 2.57 % (13) patients. ICU beds utilization is moderate but the SDU beds utilization is high as observed in real life. Policy 2 has on average a lesser utilization of the ICU (35%) compared to Policy 1 (49%). Policy 1 gives a less congested SDU utilization of about 58% on average compared to the 71.8% of Policy 2. In terms of health service benefit and cost, Policy 1 has an average cost per admitted of 13.57, an average total reward of 125.24, thus a net benefit of 111.67 per patient admitted. Oppositely, Policy 2 has an average cost per admitted of 55.40, an average total reward of 113.28, therefore a net average benefit of 57.896 per patient admitted.

A similar analysis was performed investigating the increasing arrival rate of the patients into

Performance measures	Policy 1			Policy 2			Diff
	Low 95%	Av	Up 95%	Low 95%	Av Total	Up 95%	
Admitted	462	515	568	534	537	540	-22
Rejected	0	0	0	0	0	0	0
Pre-stepped down	0	0	0	251.46	253.22	254.97	-253.22
ICU overstay	4.79	13.25	21.71	52.43	55.33	58.22	-42.08
ICU utilization (%)	41.81	47.81	53.82	34.68	34.95	35.22	12.86
SDU utilization (%)	53.08	57.73	62.37	71.41	71.81	72.22	-14.08

Table 3.15: Patients flow performance measures over four months.

Performance measures	Policy 1			Policy 2			Diff
	Low 95%	Av	Up 95%	Low 95%	Av Total	Up 95%	
Rejected Cost	0	0	0	0	0	0	0
Overstay Cost	0.52	0.90	1.09	4.91	5.16	5.39	-50.25
Discharge Cost	12.01	12.29	12.66	12.51	12.51	12.54	-0.22
Pre-Step-Down Cost	0	0	0	37.66	37.73	37.78	-37.73
Total Cost	13.18	13.57	13.91	55.07	55.40	55.70	-41.83
Admission Reward	100	100	100	100	100	100	0
step-down Reward	24.84	25.11	25.41	13.18	13.20	13.24	11.91
Total reward	124.98	125.24	125.56	113.28	113.28	113.28	11.97
Net benefit	111.06	111.67	112.40	57.57	57.90	58.20	53.77

Table 3.16: Average service performance per admitted patient.

the ICU. The results are summarized graphically in the following figures. The blue vertical line represents the point the arrival rate is equal to the service rate ($\lambda = \mu$). Policy 1 is plotted in green while Policy 2 is plotted in black.

Fig. (3.9) shows the average number of ICU requests made in the last four months of the simulation under various arrival rates with its 95% confidence interval. The confidence interval is tiny and imperceptible. As expected, the ICU demand grows linearly with an increasing rate of arrival. Fig. (3.10) the percentage of ICU requests admitted into the ICU during the last four months of the simulation under various arrival rates with its 95% confidence interval. The confidence interval is also indiscernible. When the arrival rate is less than the service rate, all patients are admitted, once the arrival rate becomes higher, we observe a linear decrease in the percentage admitted into the ICU. Fig. (3.11) the percentage of ICU requests rejected into the ICU during the last four months of the simulation under various arrival rates with its 95% confidence interval. When the arrival rate is less than the service rate, no patient is rejected, once the arrival rate becomes higher, we observe a liner increase in the percentage rejected patients. Fig. (3.12) is the percentage of admitted ICU patients who stepped down in Policy 2 during the last four months of the simulation under various arrival rates with its 95% confidence

interval. Fig. (3.13) is the the percentage of admitted ICU patients who prematurely stepped down in Policy 2 during the last four months of the simulation under various arrival rates with its 95% confidence interval. When the arrival rate is less than the service rate, a little above half of the admitted patients are stepped down and the remaining prematurely stepped down. When the arrival rate is higher than the service rate, premature step-down leaves room for normal step-downs. Note that in the figures, strategy means the optimal policy.

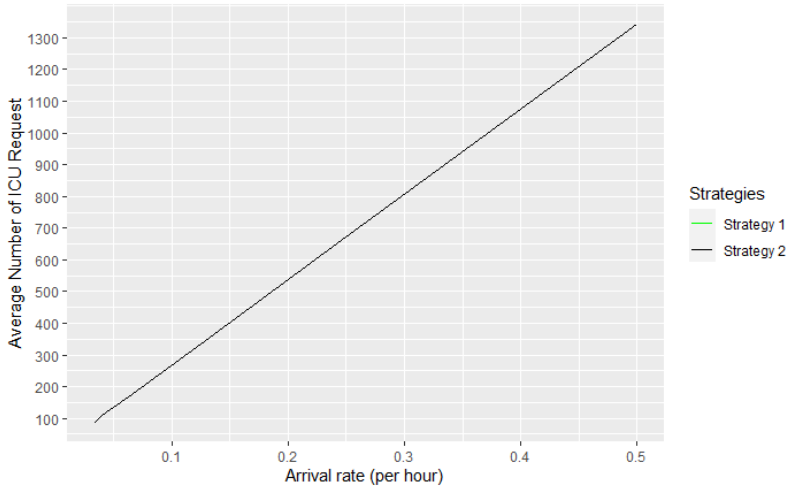


Figure 3.9: Average ICU requests

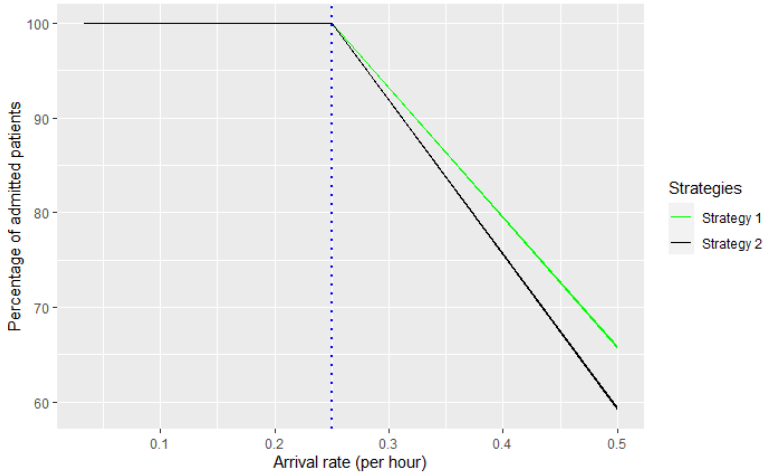


Figure 3.10: Percentage ICU Admission versus increasing arrival rate with its 95% CI. The blue vertical line represents, ($\lambda = \mu$), the point the arrival rate is equal to the service rate follows. The policy without premature step-down is plotted in green while the policy with premature step-down is plotted in black.

Fig. (3.14) and Fig. (3.15) are the average ICU utilization and SDU utilization with their 95% confidence interval. In the ICU, Policy 2 has a linearly increasing utility that is unaffected

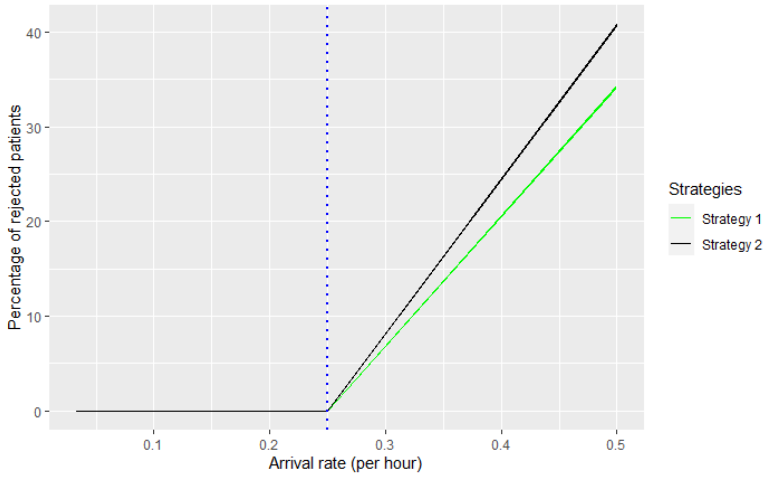


Figure 3.11: Percentage ICU rejection versus increasing arrival rate with its 95% CI. The blue vertical line represents, ($\lambda = \mu$), the point the arrival rate is equal to the service rate follows. Policy 1 is plotted in green while Policy 2 is plotted in black.

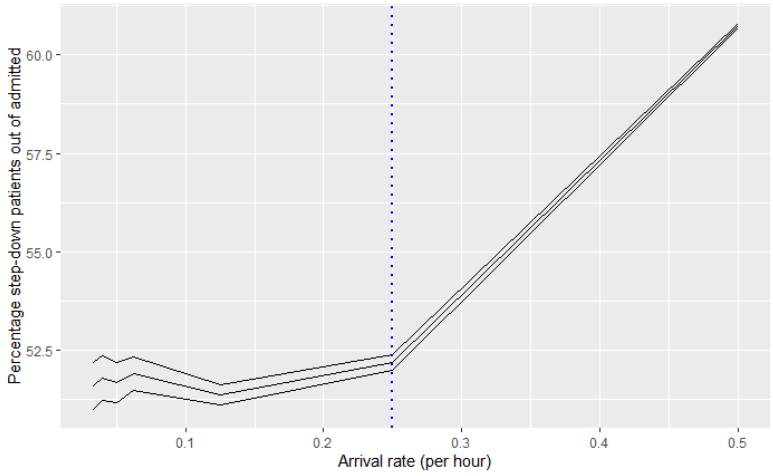


Figure 3.12: Percentage ICU step-downs using Policy 2 with its 95 % CI. The blue vertical line represents, ($\lambda = \mu$), the point the arrival rate is equal to the service rate follows.

by the steady condition. Policy 1's utility whoever, has a reduced steepness when arrival rates are greater than service rate. In general, Policy 1's ICU utility is higher than that of Policy 2. In the SDU, both policies have reduced utility steepness when arrival rates are greater than service rate. In general, Policy 2's SDU utility is higher than that of Policy 1. ICU utility and SDU utility in Policy 1 have equivalent trend while SDU's utility in Policy 2 is exorbitant.

Fig. (3.16) shows the average benefit per patient admitted. Both policies experience a decreasing trend with a higher steepness when the arrival rate is high. In general, Policy 1 has a higher benefit per admitted patient than Policy 2.

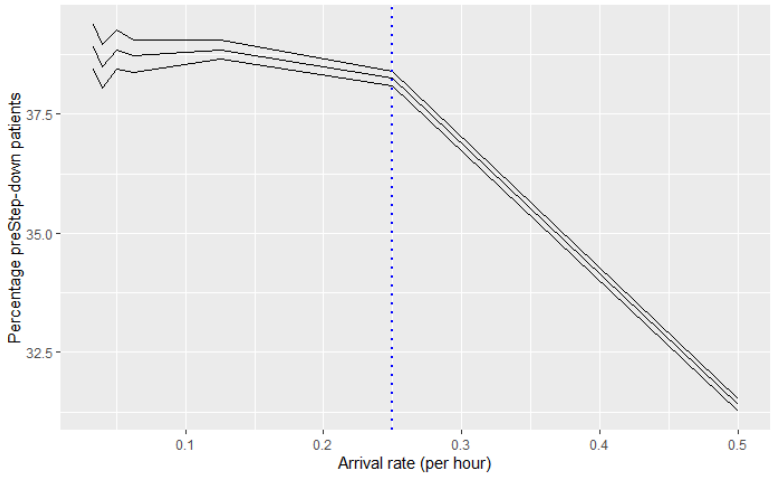


Figure 3.13: Percentage ICU premature step-downs using Policy 2 with its 95% CI. The blue vertical line represents, ($\lambda = \mu$), the point the arrival rate is equal to the service rate follows.

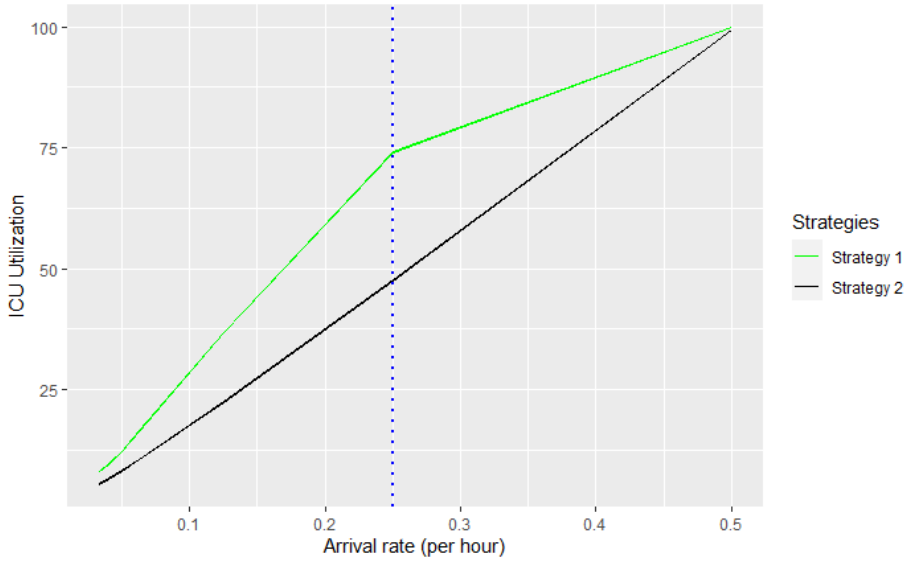


Figure 3.14: ICU utility versus increasing arrival rate with its 95% CI. The blue vertical line represents, ($\lambda = \mu$), the point the arrival rate is equal to the service rate follows. Policy 1 is plotted in green while Policy 2 is plotted in black.

3.6 Discussion

In the literature, the use of the SDU has been shown to considerably increase the ICU throughput [110, 94, 109, 134]. But the effect of premature step-down has not been compared to that of rejection in a high-demand system. The two decision rules are used to investigate the last bed problem of patient flow management in the congested environment. The sequential optimal solution stipulates admission whenever it is possible, i.e. if there is a mean of stepping down

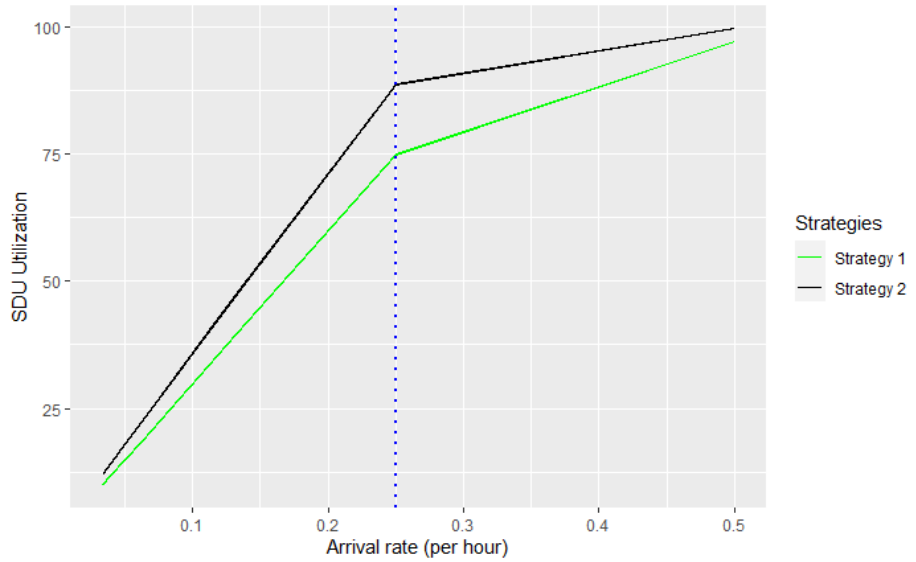


Figure 3.15: SDU utility versus increasing arrival rate with its 95% CI. The blue vertical line represents, ($\lambda = \mu$), the point the arrival rate is equal to the service rate follows. Policy 1 is plotted in green while Policy 2 is plotted in black.

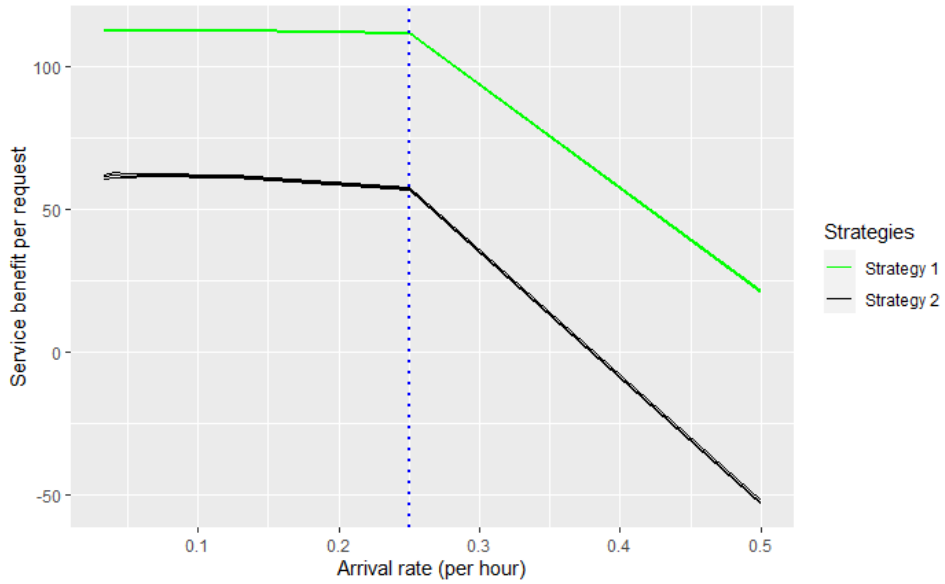


Figure 3.16: Average benefit per patient admitted and its 95% CI.

a patient to the SDU (either normal stepping down or premature stepping down, it should be done), reject whenever the ICU is full, the demand is high and the cost of premature discharge from the ICU is increased. Whenever both the ICU and the SDU are full, there is first a discharge action before a step-down action. If the cost of premature discharge is increased, then discharging patients become so costly that doing it will be rarely creating a congested ICU that prevents admission and therefore other actions may not even be possible. The interdependence

of the ICU and the SDU is clearly shown in this relationship. Because we assumed that the costs are lesser than the reward, the results are robust. As expected once the cost of premature step-down is greater than the reward of admission, the optimal action is to "do nothing", that is, it is preferable to reject arriving patients than prematurely discharging a present high acuity patient.

Because ICU lengths of stay may be long and Policy 2 has the default action to premature step-down, it initially tends to perform excessive premature step-downs. This not only is detrimental to the health service benefit of those patients but also seems to increase the SDU length-of-stay in the long run. Therefore, in the long run, the SDU is full and further normal step-downs are impossible. The flow rate into the SDU/ICU system becomes highly reduced. Fig. (3.10) and Table. (3.15) show that when the rate of arrival is less than the service rate, both policies admit on average a comparable number of patients, with Policy 1 having a higher variation due to the rejections. Premature step-down causes congestion downstream. Increasing the number of patients the SDU receives lengthens the patients stay at the ICU creating a downward congestion, preventing upstream patients to be moved out. When the SDU is congested, it creates overstay at the ICU even though the ICU may seen less busy with empty beds. Empty beds are costly to the ICU since the ICU beds' capacity is hard to change whether used or not [64].

Premature step-downs increased the SDU artificially creating higher overstays at both the ICU and the SDU and therefore reducing the normal step-downs. This creates congestion at the ICU and in the whole system creating more rejection (Fig. (3.11)). Using Policy 2, increases the number of premature step-downs as the arrival rate increases, and the SDU utilization increases faster than the ICU utilization. An increase in the SDU occupancy, increases the SDU overstays, prevents ICU normal and premature step-down and causes an increased ICU rejection. This partially explains the fewer admission by Policy 2 compared to Policy 1 when arrival rates increases. An increase in rejections leads to a drop in net survival. With an arrival rate of more than half a patient per hour whatever the policy or the system, there is no further improvement in the admittance of newly arriving patients, as the ICU is already full. In such states, the system's capacity is near 100% utilization. In general, even with the premature discharge, the system will reach a point under heavy traffic where the number of rejected patients, the ICU utilization and the SDU utilization in both policies will converge as the rate of arrival increases. In those busy states, though the policy without premature step-down rejects more patients, the policy with premature step-down rejects nearly as many and performs even more premature step-downs.

Even if a patient may be denied access to ICU due to its full capacity, nonetheless hospital management may be pressed to identify an individual to premature step-down to prevent

rejection. The perceived high risk of rejecting ICU patients and ethical considerations when rejecting a patient, constrain the practice of premature step-downs. Instead of thinking about the rejection as onsite rejection of patients, rerouting of ambulance can prevent the negative impact on those patients who may be rejected and for whom a lack of bed may prove fatal. Even if Policy 2 seems right and ethical overall, under high demand, it proves more detrimental than Policy 1. Likewise, we need to consider other arrangements when alternate levels of care are not available to patients and all beds are occupied.

3.7 Conclusions

This research is concerned with the modelling of an ICU supplemented by a Step-down Unit (SDU) to assist with efficient patient flow as patients recover, in a congested environment. The Nine Equivalents of Nursing Manpower Score (NEMS) data-set for the ICU which we considered served as the measure of patient recovery over time. We approximated the optimal actions in the resulting Markov decision process to enable patient flow under two cases. Both cases allowed for the admission of new HAP, the stepping down of existing HAP to the SDU, and the discharge of LAP from the ICU. The latter case also allowed for the premature stepping down into the SDU of HAP who have not yet completed the care they would ordinarily receive in the ICU.

Through numerical comparison, we discovered that the optimal policy for the latter case ended up doing more premature step-downs to admit a few more arrivals, relative to the case where premature stepping down was not allowed. In this way, premature step-downs were shown to impact future arrivals due to the greater level of occupancy downstream, which impedes the movement of recovering patients to the SDU. As a result of this study, we found that NEMS works well as a proxy for classifying daily patient health states and as such, translates well into a transition matrix to be used in an MDP model. As defined in this research, cohort health service benefit seems to be a good measure of the overall hospital's pressure in its acute units(ICU and SDU). We also found that the rewards and costs accumulated into the health service benefit are not very sensitive; in other words, considerable changes in the values of the rewards and costs are needed to change our general findings. In most cases, we observed that premature step-downs stress the acute care pathway and lead to further congestion downstream. In steady-state systems with lower utilization rates, we recommend Policy 1, with the use of an alternative level of care when there is an empty bed in the SDU in case the ICU becomes full. Surprisingly and counter-intuitively, in prolonged busy states (high utilization with high-demand scenarios), our findings recommend Policy 1. This policy does not allow for premature step-downs, while achieving similar levels of overall health service benefit perfor-

mance. The added benefit of Policy 1 is that it does so without the additional stress downstream which would further impact future arrivals.

Our model has its limitations. We look at the system level, not the individual. Due to its intensive level of care, the individual benefits from overstaying at the ICU, while the system under-performs. If an individual overstays in the ICU, his health service benefit does not decrease, however, the system suffers a dis-utility. Especially, if there is an arrival finding the ICU full. Secondly, NEMS is mostly used in Canada, so other jurisdictions may need to rely on other daily metrics to determine a patient's acuity such as APACHE (all versions) and SOFA. Furthermore, the MDP model captures only the congestion zone of the system's capacity. In a model with full capacity, the state space increases rapidly, making it less tractable and harder to solve both computationally and analytically. Moreover, the ICU/SDU ratio used in our MDP model is fixed at 2:1. This was so to help formulate and solve the MDP. Finally, the health service benefit as defined in the research may be an over-simplification of real-life phenomena in the ICU.

Chapter 4

Intensive Care Unit-Step-Down Unit Service Time Decisions Queuing Game

Abstract

In this paper, a length-of-stay competition between two servers in tandem without buffer between them is investigated using queuing games. This system typifies the relationship between the intensive care unit (ICU) and the step-down unit (SDU) of a hospital. We model and analyze the equilibrium length-of-stay decision under four different games (one cooperative and three non-cooperative games) as follows: (i) both servers cooperate; (ii) the servers do not cooperate and make decisions simultaneously; (iii) the servers do not cooperate and the first server, the ICU, is the leader (ICU Stackelberg); (iv) the servers do not cooperate and the second server, the SDU, is the leader (SDU Stackelberg). The payoff of the ICU is expressed as the difference between the service benefit and the waiting in queue penalty, while that of the SDU is the difference between the service benefit and the overstay penalty. The results show that length-of-stay decisions of each server depends critically on the payoff function's form and the the exogenous demand. Secondly, with a linear payoff function, the SDU is only beneficial to the system if the unit cost is greater than the unit reward at the ICU. Our results revealed also that payoffs depend on the substitutability in both ICU Stackelberg and SDU Stackelberg games. When most of the length-of-stay is spent at the ICU unit, our results suggest that the critical care pathway performs better under coordination and or leadership at the ICU level.

Keywords: Game theory, buffer-less tandem queue, cooperation, Stackelberg, simultaneous decision, length-of-stay allocation

4.1 Introduction

The intensive care unit (ICU) represents the severest and most costly level of care within a given hospital [169, 129]. To reduce cost and provide efficient care, the step-down unit (SDU) is used as an intermediate level of care between the ICU and the general ward to care for recovering patients and reducing patients' time spent in the ICU [66]. The ICU is often a congested unit due to its high demand. Therefore, the SDU is initially tasked to provide a transition for recovering patients as an alternative to increase the ICU's capacity or perform premature, demand-driven discharge of patients from ICUs to general care units [101]. However, the use of the SDUs has evolved, and there are considerable subjective views on their benefits and role in the hospital [8]. While there are opposing views on the role of the SDU to the hospital system, we will focus on its primal purpose, which is to provide a transitional place for recovering ICU patients. Of note is the fact that: Firstly, it is not clearly defined how long a patient would stay in the ICU before being moved to the SDU and secondly, the dependence or independence of the SDU in making its own decisions has not yet been studied. Therefore, it is necessary to study time partitioning (length-of-stay, or LOS) decisions between the ICU and the SDU to provide guidelines and optimal care decisions at each server. In addition, understanding the effect of power structure on the patients' LOS helps set up procedures that can reduce cost, congestion and improve coordination in the ICU/SDU system.

Though the ICU and SDU may be viewed as independent units, the SDU is created mostly to absorb ICU outpatients. In this case, a discharge decision from the ICU is dependent on the availability of space in the SDU. Decision-making will be different at the unit level, whether the two servers cooperate (as in a centralized planner) or compete. In the latter, each server is viewed as a decision-maker (or player) in a utility maximization game and an optimal setting is the one that provides the highest utility. On the other hand, in games where the objective is to maximize the throughput of a queue, we are in a queuing game [69, 58]. There are alternative ways of studying a queuing game, and the most used is viewing the customers as the players of the game, where the players observe (or not) the queue and decide whether to join or not such queue. In contrast, in this work, we view the servers as players competing against each other for the patients' LOS when servers in a series in the same feed-forward network compete. The objective of the competition is not the customers. The servers' objective or utility can be considered in terms of the share of the burden of care (patient LOS) of each of the servers.

In this work, we determine and characterize the LOS decisions between the ICU and the SDU using payoff functions that measure the burden of care in a congested system by addressing the question of competing servers in tandem without a buffer in between. To our best knowledge, this is an approach that has not been used frequently in the general queuing

game literature, much less in healthcare. We tackle this problem considering a centralized decision-making cooperative game as well as decentralized decision-making via simultaneous decision-making games and Stackelberg games.

The remainder of this paper is organized as follows: Section 4.2 provides the background literature in applications of queuing games in operations management and healthcare operations management. Then, in Section 4.3, we describe our system and the proposed model formulation. In Section 4.4, we present the results and discussion. Next, we present and discuss numerical results. Finally, the conclusion and recommendations are presented in Section 4.5.

4.2 Relevant Literature

There is a broad range of strategic decision-making research concerning customers and firms. For example, competition and cooperation among firms have been studied actively in game theory literature. Customers' strategic decisions to choose between firms have also received much attention in the literature. The works by Hassin [67] and Hassin and Haviv [68] provide an exhaustive literature review on the topic. This section, briefly presents some related papers that looked at queuing game progress in healthcare followed by papers with profit maximization objective in a network of servers' system.

In his seminal work, Arrow [9] laid the foundation of what is later known as the "contract theory". He argues that a physician's medical decision-making cannot be simply modelled as a profit-maximization problem as it is in the case of a firm's decision-making problem. Thinking in the same direction, Siciliani and Hurst [148] and Brekke et al. [24] studied the impact of hospital competition on waiting times. Siciliani and Hurst [148] modelled a market with only two dominant producers where a general practitioner refers the patients to the hospital with the lowest waiting time. The hospitals choose the supply of care and waiting time to compete for patients. The paper shows that substitutability among hospitals reduces the supply of care in equilibrium and results in longer waits. Brekke et al. [24] modelled hospitals competing in a spatially differentiated market. They considered two types of patients who differ in their treatment benefits: high segment and low segment. The hospitals simultaneously announce waiting times, but they cannot differentiate the two patient segments in terms of waiting times. Sadat et al. [139] consider a duopoly quality of care competition between two hospitals to capture a fraction of the total market demand. Patients decide on the hospital that provides the highest utility, as a function of price and the patient's perceived quality. They show that while patients may enjoy a positive utility based on demand and perceived quality of care, hospitals share the market demand based on their perceived quality of care and capacity. Chen et al.

[36] focus on price competition in a market of three service providers; one free public service provider and two private service providers who charge a price. The objective of the service providers is to choose the price that optimizes their profits by providing quality service, higher capacity to reduce wait time. They show that such a price competition between the private service providers in a market with a public service provider reaches a pure Nash equilibrium. They investigate the impact of competition and collaboration between the two private service providers on social welfare.

Attempt to maximize the network of servers through cooperation have also been studied in the literature. Anily and Haviv [6] using a transferable utility cooperation game, considered improving n servers' efficiency by pooling service capacities to serve individual streams of customers. They observed that for any subset of servers there exists cost-sharing allocations under which no partial subset can take advantage by leaving and forming a separate coalition. Karsten et al. [81] modelled M/M/C queue system by complete pooling of their resources and customer streams into a joint service system by providing sufficient conditions for the games under consideration to possess a core allocation using cooperative game theory concepts. Timmer and Scheinhardt [158] study cooperation between n-node tandem Jackson network servers in series to minimize the overall waiting time in the system. For two and three M/M/1 servers, they discover that cost-sharing through cooperation is possible in many ways and depends on the service and arrival rates. Zeng et al. [173] develop a marginal analysis algorithm and a greedy heuristic algorithm before conducting numerical studies to solve what they call the server transfer problem. They attempted to obtain a win-win capacity transfer solution of independent M/M/C queue systems through cooperation. Karsten et al. [81] model the resource pooling of collaborated servers to pool resources into a joint service system as Erlang loss systems. Their objective was to minimize additive fixed cost rate per server and penalty costs for lost customers. They identify a cost allocation that gives no subset of players an incentive to split off and form a separate pooling group. Bendel and Haviv [18] considered a tandem network of queues that cooperate by pooling resources with a transferable utility leading to a single combined server that satisfies the aggregated service demands with a greater service rate. They derived the core allocation and found out that the cost of a coalition is the steady-state mean total number of customers in the system formed by its members. Additional literature that has minimized the cost of independent servers using cooperative games are [159, 172, 160].

As we've shown, competition and cooperation between hospitals have a wide presence in the queuing game literature. But all have either being parallel servers competing for the jockeying customers in a queue or a network of queues cooperating. In this paper, we consider a game between two servers (two hospital units) in series and the servers compete for the service time of customers, not the customers themselves.

4.3 Proposed System and Model

This section describes and introduces the basic framework of our system and its associated model. We consider a system of two stations of servers (ICU and SDU) arranged in series without a queue between them. In the first station, the ICU can provide full service to the patients with a concave payoff function. To reduce cost, ICU will transfer patients to the second station, the SDU with a lower service cost to complete the service. We denote this system using Kendal's notation as the A/B/C-D/E system, where A is the arrival pattern to the first station, B is the service distribution of the servers at the first station, C is the number of servers at the first station, D is the service distribution at the second station and E is the number of servers at the second station. In Fig. (4.1), we consider an M/M/1-M/1 system, where the M stands for a Markovian or memory-less process. Customers arrive at the servers according to a Poisson process with a rate of λ. Length-of-stays at each station are exponentially independent and identically distributed with a mean length-of-stay $l_i = \frac{1}{\mu_i}, i = 1, 2$, where μ_i is the service rate at server i. The arrival rate is assumed to be known and the average length-of-stays (Patients' LOS) are the decision variables. The system is a first come first serve (FCFS) type of service and each newly arriving customer immediately goes into service if an idle server is available. If the customer finds the first station occupied, they wait in the queue. The waiting time in queue costs c at the first server. The cost of overstaying the first server because the second server is occupied is c. The servers are rewarded with benefits $r_i, i = 1, 2$, for service.

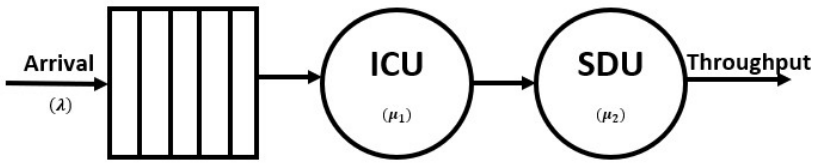


Figure 4.1: M/M/1-/M/1 System flow. Customers arrive at the servers according to a Poisson process with a rate of λ. length-of-stays at each of the stations are exponentially independent and identically distributed with mean length-of-stay $l_i = \frac{1}{\mu_i}, i = 1, 2$, where μ_i is the service rate at server i.

In the ICU/SDU system, new patients represent the customers. Patients do not decide on whether to join the queue or not, however, the ICU and the SDU make decisions on how long a patient stays in their respective services given that if it was only the ICU serving, the LOS will be l. Let the LOS at server i be $l_i, i = 1, 2$, then $l_{ICU} + l_{SDU} \leq l$. We formulate the decision models as a constrained optimization problem. The objective is to determine the optimal average length-of-stay for each of the two stations. We consider four possible power game scenarios for which the rules and structure of the game are given as follows:

1. ***Cooperation (CP)***. The two servers have a unified payoff they try to maximize. This case is similar to having one central planner overseeing the two units.

2. ***Simultaneous Decision (ST)***. The two servers have different payoff profits. Each server maximizes its own payoff. Decisions are made simultaneously. This case is similar to have two decentralized managers over the two servers.

   Stackelberg Game. The two servers have different payoffs. Each server maximizes its payoff. Decisions are made sequentially.

3. ***ICU Stackelberg (IS)***. The ICU chooses its LOS first using the response function of the SDU. The SDU determines its LOS to maximize its payoff function.

4. ***SDU Stackelberg (SS)***. The SDU chooses its LOS first using the response function of the ICU. The ICU determines its LOS to maximize its payoff function.

Our optimization problem is given as

$$\begin{aligned} \max_{l_i} \quad & S_i = R(l_i) - C(w_i) \quad i = 1, 2 \\ \text{s.t.} \quad & \sum_{j=1}^{2} l_j \leq l, l_j \geq 0 \\ & \lambda \leq \frac{1}{\sum_{j=1}^{2} l_j}, \quad j = 1, 2 \end{aligned} \tag{4.1}$$

where S_i is the payoff of server i. and l is the LOS when only one server performed all the services. The first part of the payoff $R(l_i)$, represents the service benefit as a function of the LOS l_i at the server, and the second part $C(w_i)$, is the penalty associated with the waiting time, w_i, before service. At the ICU, this corresponds to waiting in queue to join the system, and at the SDU, this corresponds to waiting at the first server to join the second server. The expected wait and length-of-stays are obtained as a characteristic of the queue using queuing theory formulas for the M/M/1 queue. The first constraint assumes that the sum of all the length-of-stays at each of the servers is less or equal to the LOS if only one server did the full service. The second constraint is the steady-state stability requirement of a simple queuing system.

The ICU's payoff function is

$$S_1 = \lambda \left(r_1 l_{ICU} - \frac{c\lambda(l_{ICU} + l_{SDU})^2}{1 - \lambda(l_{ICU} + l_{SDU})} \right) \tag{4.2}$$

where the ICU benefit function, $R(l_{ICU}) = r_1 l_{ICU}$ is a linear function of the LOS with r_1 being the unit service benefit (for the rest of the paper, for simplicity and without loss of generality, we assume $r_1 = 1$ and measure the SDU benefit against it, and the queue penalty function,

$C(w_1) = cW_q$ is a linear function of the waiting time $W_q = \frac{\lambda l^2}{1-\lambda l}$ with c being the unit wait time cost.

The SDU's payoff function is

$$S_2 = \lambda\left(r_2 l_{SDU} - \frac{c l_{ICU} l_{SDU}}{l_{ICU} + l_{SDU}}\right) \tag{4.3}$$

with similar benefit function, $R(l_{SDU}) = r_2 l_{SDU}$, (for the rest of the paper, we assume $(0 \leq r_2 = r \leq r_1 = 1)$ and queue penalty function, $C(w_2) = cW_2$. The waiting time, W_2, is the expected overstay time at the first server due to the second server's business. This happens when the first server finishes service first and the second does not. Since each of the two servers serves according to an exponential distribution with parameters $\mu_i = \frac{1}{l_i}, i = 1, 2$, we can show that the probability that the first server finishes first is given by

$$\begin{aligned} P(T_1 < T_2) &= P(argmin_{i\in\{1,2\}}\{T_1, T_2\} = 1) \\ &= \frac{\mu_1}{\mu_1 + \mu_2} \\ &= \frac{l_{SDU}}{l_{ICU} + l_{SDU}}, \end{aligned} \tag{4.4}$$

where $\mu_i, i = 1, 2$ is the service rate at server i. Table 4.1 presents the description of each variable in our model.

Table 4.1: Model variables

Variable	Description
λ	Arrival rate into the system
l	LOS needed for full recovery
l_{ICU}	ICU LOS
l_{SDU}	SDU LOS
c	Unit loss time penalty
r_1	ICU benefit
r_2	SDU benefit
W_1	Expected waiting time in Queue
W_2	Expected overstay time

4.4 Results and Discussions

4.4.1 Equilibrium Length of Service Decisions and Payoffs

We first solve the following optimization problem:

$$\max_{l} \quad S_u = \lambda\left(l - \frac{c\lambda l^2}{1-\lambda l}\right) \tag{4.5}$$

to obtain the optimal average LOS, l^*, if the system is a single ICU station. The payoff function of the ICU alone system in Eq. 4.5 is illustrated in Fig. (4.2).

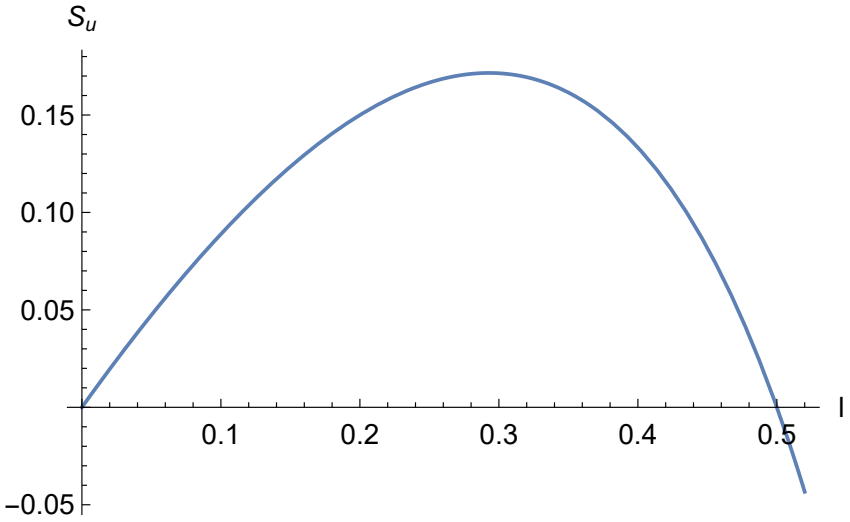


Figure 4.2: Concave payoff function of the system with only one station (ICU) as a function of the LOS (l) when $r = 1, c = 1$. The LOS is in a unit length-of-stay and the payoff function is measure in system service effectiveness.

The concavity of the utility as a function of the LOS l guarantees the existence of a unique maximal value for the optimal average LOS l^*. When the steady-state condition, ($\lambda l_{ICU} < 1$), is fulfilled. The optimal average LOS is obtained as

$$l^* = \frac{1}{\lambda}\left(1 - \sqrt{\frac{c}{1+c}}\right). \tag{4.6}$$

Analysis of the one station ICU system's payoff from Eq. 4.5 shows that the utility function is a decreasing function of the queue cost, c, a concave function of the LOS, l, and the demand, λ, and an increasing function of the in the ICU benefit, r. Thus, a high arrival rate negatively affects the payoff value due to the queue in front of the server by causing wait. The queue cost can force the server to set a lower service length that will increase the service rate but is not effective as many will not be served efficaciously. To avoid the loss of utility to the

whole system, due to the cost of waiting patients in a queue, as the length-of-stay of the current patient increases, a second station (SDU) with lower service cost is created to care for those patients after a given amount of time in the ICU. As the two servers in tandem either cooperate or compete for the patient's LOS, we now derive analytical equilibrium average length-of-stays solutions for those two stations under each power structure game scenario. Then illustrate them using numerical examples.

Cooperative Decision (CP)

Under the cooperation game, there is no competition between the servers. The two servers are assumed to be managed by one central planner. This manager has to choose the LOS of both servers cooperatively and optimize the payoff function of the entire system formed by the ICU and the SDU. The manager may choose to give all services to one server or divide the service among them. One server may have a larger service time compared to the other. Under cooperation, the following optimization problem is considered:

$$\begin{aligned} \max_{l_{ICU}, l_{SDU}} \quad & S_c = \lambda\left(l_{ICU} + r l_{SDU} - \frac{c\lambda(l_{ICU}+l_{SDU})^2}{1-\lambda(l_{ICU}+l_{SDU})} - \frac{c l_{ICU} l_{SDU}}{l_{ICU}+l_{SDU}}\right) \\ \text{s.t.} \quad & l_{ICU} \geq 0, l_{SDU} \geq 0 \\ & \lambda < \frac{1}{l_{ICU}+l_{SDU}}. \end{aligned} \tag{4.7}$$

S_c is the sum of the ICU and SDU's payoffs. The ICU and SDU's reaction functions given the LOS of the other server are derived as:

$$\frac{\partial S_c}{\partial l_{ICU}} = \lambda\left(\frac{c l_{ICU} l_{SDU}}{(l_{ICU}+l_{SDU})^2} - \frac{c l_{SDU}}{l_{ICU}+l_{SDU}}\right) + \lambda\left(1 - \frac{c\lambda^2 (l_{ICU}+l_{SDU})^2}{(1-\lambda(l_{ICU}+l_{SDU}))^2} - \frac{2c\lambda(l_{ICU}+l_{SDU})}{1-\lambda(l_{ICU}+l_{SDU})}\right) = 0 \tag{4.8}$$

$$\frac{\partial S_c}{\partial l_{SDU}} = \lambda\left(-\frac{c\lambda^2 (l_{ICU}+l_{SDU})^2}{(1-\lambda(l_{ICU}+l_{SDU}))^2} - \frac{2c\lambda(l_{ICU}+l_{SDU})}{1-\lambda(l_{ICU}+l_{SDU})}\right) + \lambda\left(r - \frac{c l_{ICU}}{l_{ICU}+l_{SDU}} + \frac{c l_{SDU} l_{ICU}}{(l_{ICU}+l_{SDU})^2}\right) = 0 \tag{4.9}$$

From Eqs. (4.8) and (4.9), the ICU and SDU's equilibrium average length-of-stays are derived as:

$$l_{ICU}^{CP*} = \left(\frac{c+r-1}{2c\lambda}\right)\left(1 - \sqrt{\frac{4c^2}{3c^2+2c(r+1)-(r-1)^2}}\right), \text{ and} \tag{4.10}$$

$$l_{SDU}^{CP*} = \left(\frac{c-r+1}{2c\lambda}\right)\left(1-\sqrt{\frac{4c^2}{3c^2+2c(r+1)-(r-1)^2}}\right). \tag{4.11}$$

The sum of the length-of-stays in (4.10) and (4.11) is the total LOS under cooperation, given by

$$l^{CP} = \frac{1}{\lambda}\left(1-\sqrt{\frac{4c^2}{3c^2+2c(r+1)-(r-1)^2}}\right). \tag{4.12}$$

And also, by substituting Eqs. (4.10) and (4.11) in the payoff functions in Eqs. (4.2) and (4.3), the payoffs are derived as follows:

$$S_{ICU}^{CP} = \left(1-\sqrt{\frac{4c^2}{3c^2+2c(r+1)-(r-1)^2}}\right)\left(\frac{2c^2+c+r-1}{2c}-\frac{1}{2}\left(3c^2+2c(r+1)-(r-1)^2\right)\right) \tag{4.13}$$

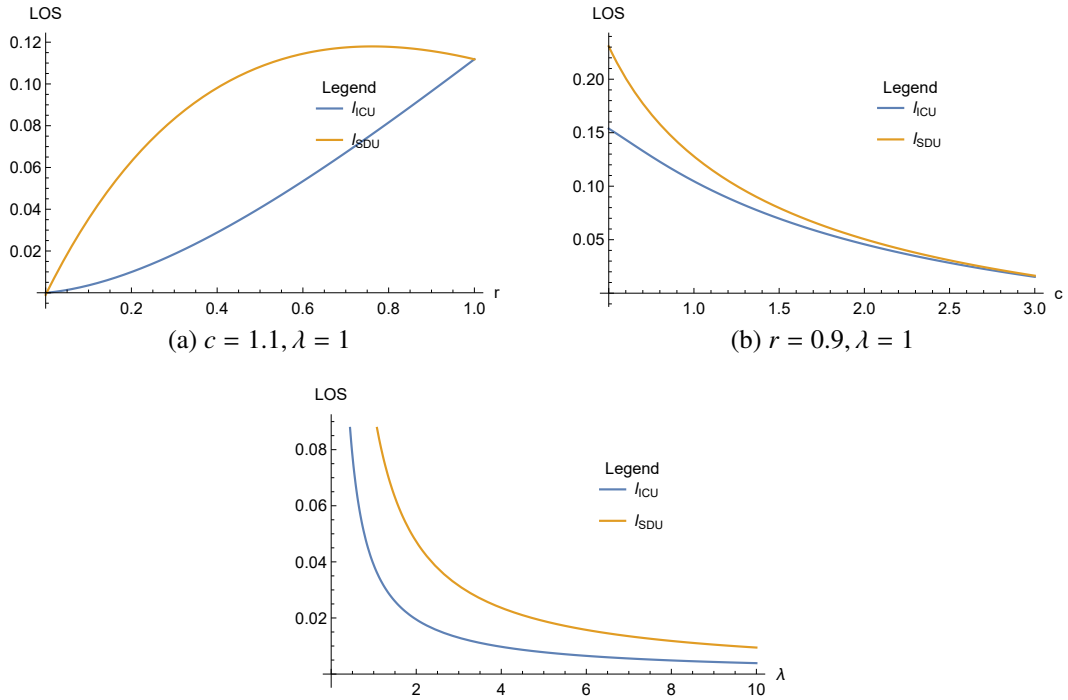
$$S_{SDU}^{CP} = \left(1-\sqrt{\frac{4c^2}{3c^2+2c(r+1)-(r-1)^2}}\right)\left(\frac{1-(c-r)^2}{4c}\right) \tag{4.14}$$

with the full system payoff shown in Eq. A.2 in the appendix.

Lemma 4.4.1 *Under cooperation,*
$l_{ICU}^{CP}, l_{SDU}^{CP} > 0$ *and* $l_{SDU}^{CP} > l_{ICU}^{CP}$, *if* $(0 < c \leq 1$ *and* $1-c < r < 1)$
or $(1 < c < 4$ *and* $(1-\sqrt{c})^2 < r < 1)$
and $S_{SDU}^{CP} > S_{ICU}^{CP}$, *if* $0 < c < 4$ *and* $(1-\sqrt{c})^2 < r < 1$.

Lemma 4.4.1 states that, under cooperation game, at equilibrium, the LOS (or burden of care) at the SDU is longer than that of the ICU, as shown in Fig. 4.3(a),(b) and (c). Both stations share the burden of care when the ICU service benefit is four times less than the queuing cost. The cost can be at most four times the ICU benefit, otherwise, both servers' LOS will have a negative LOS, which means, its service would not be appropriate and violates the constraints of the modell. Fig. 4.3(a) shows that when the SDU benefit approaches that of the ICU, there is a transfer of the burden of care from the SDU to the ICU. Conversely, when the cost and/or the demand increases, both server's length-of-stays decreases exponentially.

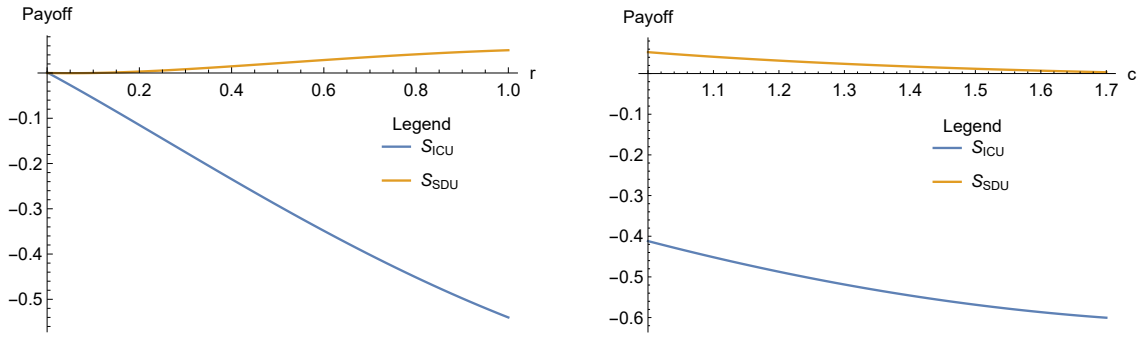
As shown in Fig.4.4(a) and (b), under cooperation game, the payoff of the SDU is greater than that of the ICU, $S_{SDU}^{CP} > S_{ICU}^{CP}$. This can partially be explained by the fact that under the cooperative game, patients spend less time in the ICU, i.e., $l_{SDU}^{CP} > l_{ICU}^{CP}$. Fig.4.4(a) shows that increasing the SDU reward reduces the ICU payoff, making the ICU perhaps unnecessary. This



(a) $c = 1.1, \lambda = 1$ (b) $r = 0.9, \lambda = 1$

(c) $r = 0.9, c = 1.1$

Figure 4.3: length-of-stay at the ICU and SDU (l_{ICU} in blue and l_{SDU} in orange) under the cooperation game as a function of (a) SDU benefit, (b) Lost time Cost, and (c) Arrival rate.



(a) $c = 1.1, \lambda = 1$ (b) $r = 0.9, \lambda = 1$

Figure 4.4: Payoffs at the ICU and SDU (S_{ICU} in blue and S_{SDU} in orange) under the cooperation game as a function of (a) SDU benefit and (b) Lost time Cost.

supports the assumption and the reality of the SDU reward being lesser than that of the ICU. Under cooperation, it will be better to have a unique station if the cost is lower than the reward but if the cost is higher than the reward, then the two servers are beneficial. Fig.4.4(b) shows, as expected that as the cost increases, the payoffs decrease.

Simultaneous Decision (ST)

As hospitals become congested, coordination efforts can fall apart rather quickly. So under simultaneous decision-making, servers make their decision individually and simultaneously as they compete for the patient's LOS. The two servers are assumed to be managed independently, each manager chooses its service length in a non-cooperative manner to maximize its payoff function. The servers' reaction functions are obtained as a derivative of Eqs. (4.2) and (4.3) as:

$$\Gamma_{ICU}^{ST} = \lambda\left(1 - \frac{c\lambda^2\left(l_{ICU} + l_{SDU}\right)^2}{\left(1 - \lambda\left(l_{ICU} + l_{SDU}\right)\right)^2} - \frac{2c\lambda\left(l_{ICU} + l_{SDU}\right)}{1 - \lambda\left(l_{ICU} + l_{SDU}\right)}\right) \tag{4.15}$$

$$\Gamma_{SDU}^{ST} = \lambda\left(r - \frac{cl_{ICU}}{l_{ICU} + l_{SDU}} + \frac{cl_{SDU}l_{ICU}}{\left(l_{ICU} + l_{SDU}\right)^2}\right) \tag{4.16}$$

The elements of the Hessian of this system are given in Appendix Eq. (A.3). Under simultaneous game, the equilibrium length-of-stays are obtained as:

$$l_{ICU}^{ST*} = \frac{1}{\lambda}\left(1 - \sqrt{\frac{c}{1+c}}\right)\sqrt{\frac{r}{c}}, \text{ and} \tag{4.17}$$

$$l_{SDU}^{ST*} = \frac{1}{\lambda}\left(1 - \sqrt{\frac{c}{1+c}}\right)\left(1 - \sqrt{\frac{r}{c}}\right). \tag{4.18}$$

The total LOS under the simultaneous decision is obtained as

$$l^{ST} = \frac{1}{\lambda}\left(1 - \sqrt{\frac{c}{1+c}}\right). \tag{4.19}$$

The stations' payoffs under simultaneous decision are given as

$$S_{ICU}^{ST} = \left(1 - \sqrt{\frac{c}{c+1}}\right)\left(2c + \sqrt{\frac{r}{c}}\right) - \sqrt{\frac{c}{c+1}}, \tag{4.20}$$

$$S_{SDU}^{ST} = \left(1 - \sqrt{\frac{c}{c+1}}\right)\left(2r - (r+c)\sqrt{\frac{r}{c}}\right), \text{ and} \tag{4.21}$$

$$S^{ST} = \left(1 - \sqrt{\frac{c}{c+1}}\right)\left(\sqrt{\frac{(c+2r)^2}{c+1}} - \sqrt{\frac{r(c+r-1)^2}{c(c+1)}} - \sqrt{c}\right). \tag{4.22}$$

Lemma 4.4.2 *Under simultaneous decision game,*

1. $l_{ICU}^{ST} > l_{SDU}^{ST} > 0$, *if* $r < c < 4r$ *and* $0 < r < 1$.

2. $S_{ICU}^{ST} > S_{SDU}^{ST}$, *if* $1 < c$ *and* $r_{ls} < r < 1$.

where r_{ls} is the solution of the Equation 4.23 close to zero.

$$\begin{aligned} X^6 + X^5(4-4c) + X^4\left(14c^2 - 4c + 6\right) + X^3\left(-24c^3 + 10c^2 + 4c + 4\right) + \\ X^2\left(17c^4 + 10c^2 + 4c + 1\right) + X\left(-4c^5 - 2c^4 - 8c^3 - 2c^2\right) + c^4 = 0 \end{aligned} \tag{4.23}$$

Lemma 4.4.2 states and it is shown in Fig.4.5(a), (b) and (c) that under simultaneous decision, the ICU takes most of the burden of care and both servers are needed when the cost is higher than the SDU reward. The ICU only gives up LOS when the SDU reward is small and its cost is high. In Fig. 4.5(a), it is shown that the ICU LOS in blue increased while that of the SDU in orange decreased when the SDU benefit increased. This transfer of the burden of care from the SDU to the ICU from low benefit to high benefit is counter-intuitive. Fig. 4.5(b) proves that when the cost is higher than the SDU reward, the ICU LOS is higher than the SDU LOS. We also observe a partial transfer of the burden of care from the ICU to the SDU when the cost increases, yet the ICU LOS remains higher. Thus the ICU remains the provider of most of the patient's LOS in the system under simultaneous decision. From Fig. 4.5(c), increasing the demand on the other hand does not affect the leader in terms of care.

Fig.(4.6) shows that the SDU payoff imay prove to be negative, i.e., though the SDU provides a positive care, it does not benefit itself but benefits the system. In the simultaneous decision game, the SDU does not benefit itself even though it contributes to the lessening of the burden of care. As expected, the payoffs increase as the SDU reward increase but that of the SDU remain low (see, Fig.4.6(a)). Also, when the cost increased, the payoffs decrease (See, Fig.4.6(b)). In Fig. 4.6(c) we can clearly observe that the payoffs are independent of the demand.

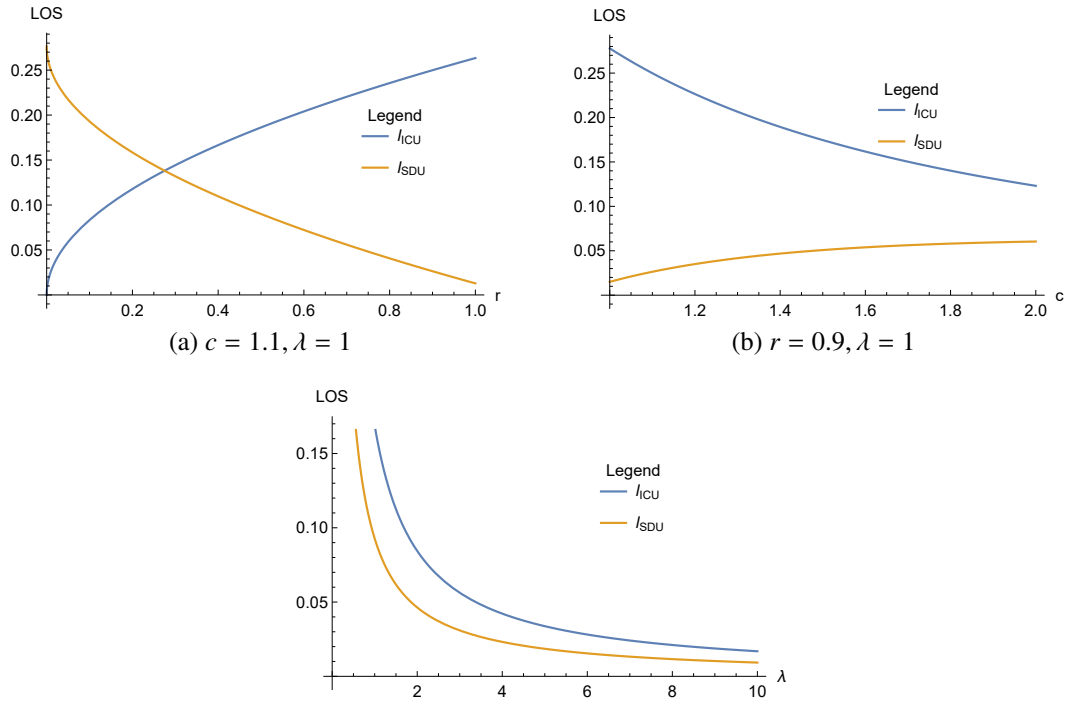
ICU Stackelberg (IS)

Under the ICU Stackelberg (IS) assumption, the ICU as the leader makes its LOS decision first then the SDU reacts to the ICU's decision. The equilibrium is computed backwardly starting from the SDU's reaction function. The SDU's reaction function is derived from Eq. (4.3) as follows:

$$\Gamma(l_{SDU}) = l_{ICU}\left(\sqrt{\frac{c}{r}} - 1\right), \tag{4.24}$$

where l_{ICU} and l_{SDU} represent the ICU and the SDU LOS respectively.

This form of the SDU reaction function implies that the LOS of the SDU is proportional to that of the ICU. From Eq . (4.24), the ICU's Nash equilibrium LOS decision can be derived as



(a) $c = 1.1, \lambda = 1$

(b) $r = 0.9, \lambda = 1$

(c) $r = 0.9, c = 1.1$

Figure 4.5: length-of-stays at the ICU and SDU (l_{ICU} in blue and l_{SDU} in orange) under simultaneous decision as a function of (a) SDU benefit, (b) Lost time Cost,and (c) Arrival rate.

the following maximization problem:

$$\max_{l_{ICU}} \left\{ \frac{\lambda l_{ICU} \left(\lambda l_{ICU} \left(c^2 + \sqrt{cr} \right) - r \right)}{\lambda l_{ICU} \sqrt{cr} - r} \right\}. \tag{4.25}$$

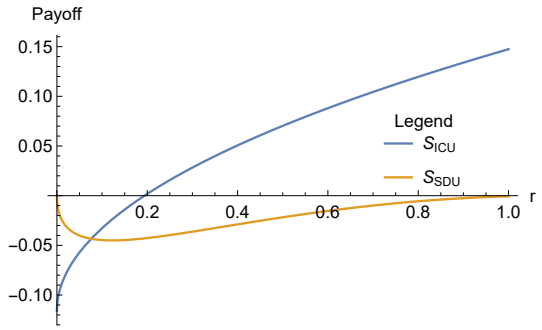
The equilibrium ICU LOS under the ICU Stackelberg is obtained as:

$$l_{ICU}^{IS} = \frac{r(c^2 + \sqrt{cr}) - cr\sqrt{c^2 + \sqrt{cr}}}{\lambda(c^2\sqrt{cr} + cr)}. \tag{4.26}$$

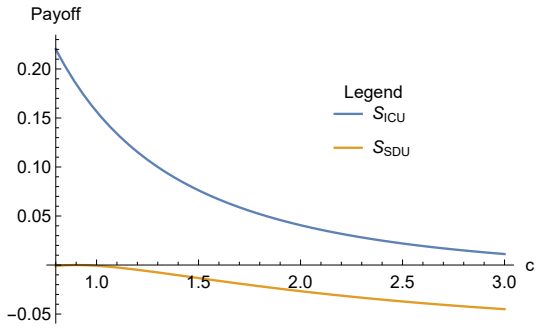
Back substitution in Eq. 4.24 gives the SDU LOS equilibrium as follows:

$$l_{SDU}^{IS} = \left(\sqrt{\frac{c}{r}} - 1 \right) \left(\frac{r(c^2 + \sqrt{cr}) - cr\sqrt{c^2 + \sqrt{cr}}}{c\lambda r + c^2\lambda\sqrt{cr}} \right), \tag{4.27}$$

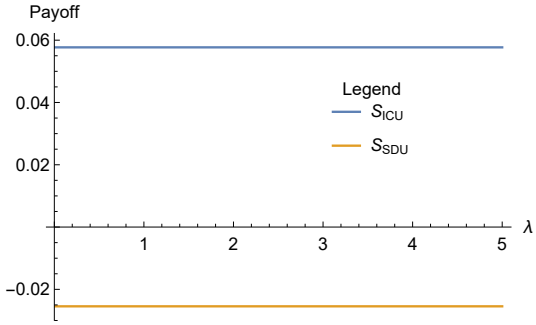
and the total length-of-stay in the system



(a) $c = 1.1, \lambda = 1$



(b) $r = 0.9, \lambda = 1$



(c) $r = 0.9, \lambda = 1$

Figure 4.6: Payoffs at the ICU and SDU (S_{ICU} in blue and S_{SDU} in orange) respectively under simultaneous decision as a function of (a) SDU benefit and (b) Lost time Cost.

$$l^{IS} = \sqrt{\frac{c}{r}}\left(\frac{r(c^2 + \sqrt{cr}) - cr\sqrt{c^2 + \sqrt{cr}}}{c\lambda r + c^2\lambda\sqrt{cr}}\right). \tag{4.28}$$

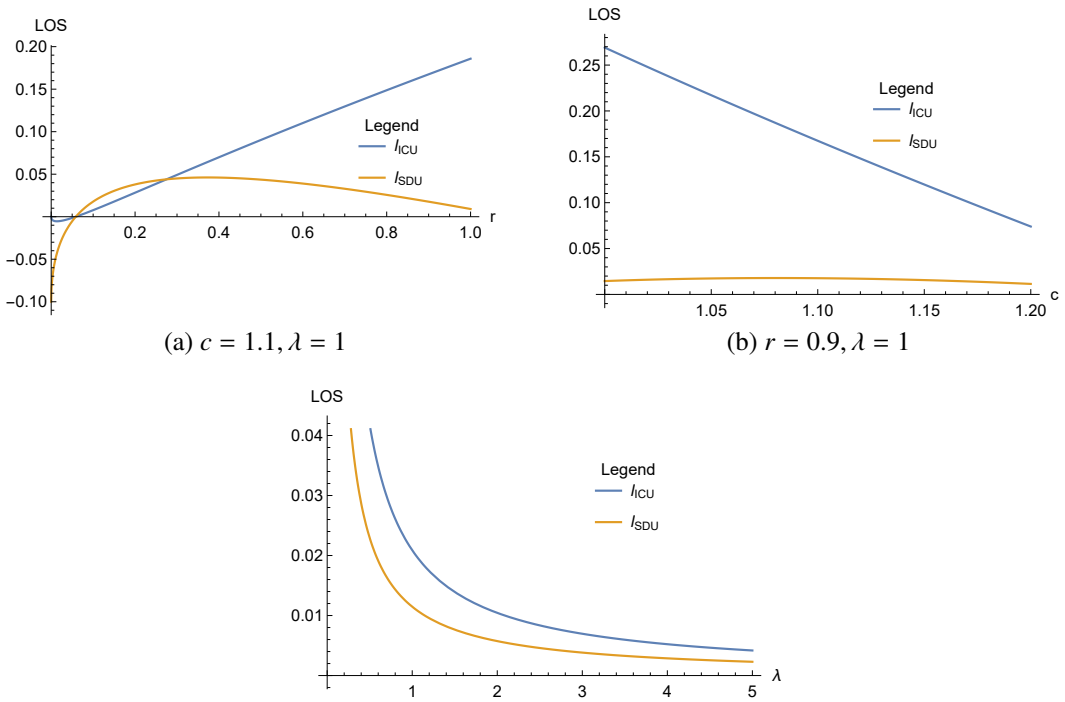
The payoffs under the ICU Stackelberg game are displayed in Appendix Eqs. (A.5) and (A.6).

Lemma 4.4.3 *Under the ICU Stackelberg game,*

1. $l^{IS}_{ICU} > l^{IS}_{SDU} > 0$, *if* $(0 < c < 1$ *and* $\frac{c}{4} < r < c)$, *or* $(1 < c < c_{iu}$ *and* $\frac{c}{4} < r < 1)$.

From Lemma 4.4.3 it is noted again that, for the system to require both servers, the cost must be higher than the ICU benefit and the SDU reward must be lower than the ICU benefit. The ICU has the highest burden of care when $\frac{c}{4} < r < 1$ while the SDU leads when $0 < r < \frac{c}{4}$. From Fig. 4.7(a), increasing the SDU benefit increases the ICU LOS in blue while that of the SDU in orange has a concave form. From Fig. 4.5(b) higher cost reduces the LOS at both servers. From 4.5(c), increasing the demand decreases the length-of-stay exponentially. From Fig. (4.8), the SDU payoff can be mostly negative . It does not benefit itself but contributes to

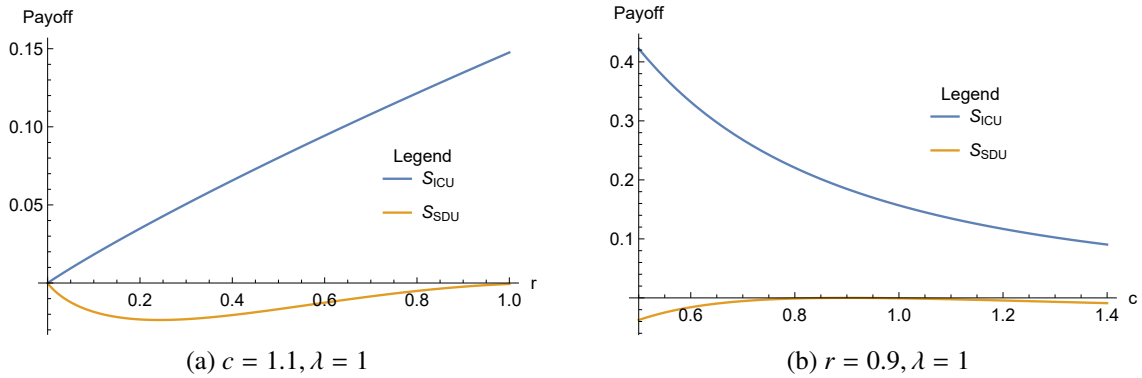
the overall care of the system. Increasing the SDU reward increases the payoff at both servers (Fig. 4.8(a)). Increased cost decreased payoff at both servers (Fig. 4.8(b)).



(a) $c = 1.1, \lambda = 1$ (b) $r = 0.9, \lambda = 1$

(c) $r = 0.9, c = 1.1$

Figure 4.7: length-of-stays at the ICU and SDU (l_{ICU} in blue and l_{SDU} in orange) under the ICU Stackelberg game as a function of (a) SDU benefit, (b) Lost time Cost, and (c) Arrival rate.



(a) $c = 1.1, \lambda = 1$ (b) $r = 0.9, \lambda = 1$

Figure 4.8: Payoffs at the ICU and SDU (S_{ICU} in blue and S_{SDU} in orange) under the ICU Stackelberg game as a function of (a) SDU benefit and (b) Lost time Cost.

SDU Stackelberg (SS)

Under the SDU Stackelberg (SS) game, the SDU becomes the leader and the ICU the follower. In this case, the SDU takes the ICU's reaction functions into account for its own LOS decisions. The ICU's reaction function is derived as

$$\Gamma(l_{ICU}) = \frac{1}{\lambda}\left(1 - \sqrt{\frac{c}{1+c}}\right) - l_{SDU}. \tag{4.29}$$

The SDU exploits the ICU's reaction function (Eq.(4.29)) by setting optimal LOS in its payoff maximization problem in Equation (4.3)). Solving this maximization problem provides the equilibrium SDU LOS as

$$l_{SDU}^{SS} = \left(\frac{c-r}{2c\lambda}\right)\left(1 - \sqrt{\frac{c}{1+c}}\right). \tag{4.30}$$

Back substitution of (4.30) in Eq. (4.29) gives the ICU equilibrium LOS

$$l_{ICU}^{SS} = \left(\frac{c+r}{2c\lambda}\right)\left(1 - \sqrt{\frac{c}{1+c}}\right). \tag{4.31}$$

The total system LOS under the SDU Stackelberg is obtained as

$$l^{SS} = \left(\frac{1}{\lambda}\right)\left(1 - \sqrt{\frac{c}{1+c}}\right). \tag{4.32}$$

The corresponding payoffs of the stations under the SDU Stackelberg game are obtained as follows:

$$S_{ICU}^{SS} = \frac{4c^2 + c + r}{2c} - \frac{\left(4c^2 + 3c + r\right)}{2c}\sqrt{\frac{c}{c+1}} \tag{4.33}$$

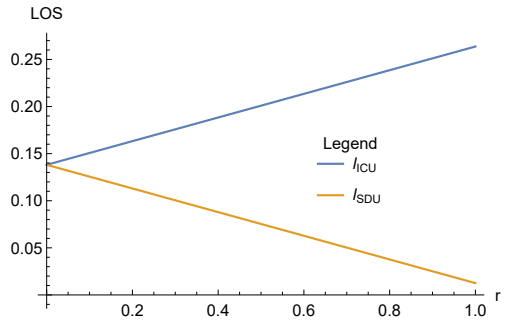
$$S_{SDU}^{SS} = \frac{(c-r)^2}{4c}\left(\sqrt{\frac{c}{c+1}} - 1\right) \tag{4.34}$$

$$S^{SS} = \left(\frac{7c^2 + 2c(r+1) - (r-2)r}{4c}\right) - \left(\frac{7c^2 + 2c(r+3) - (r-2)r}{4c}\right)\sqrt{\frac{c}{c+1}} \tag{4.35}$$

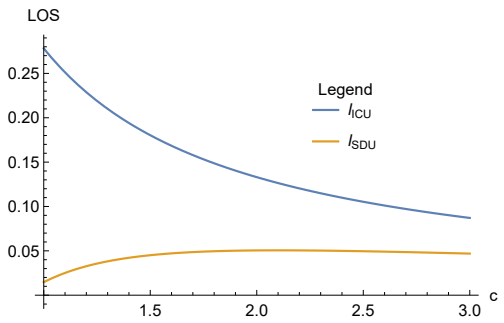
Lemma 4.4.4 *Under SDU Stackelberg game,*

1. $l_{ICU}^{SS} > l_{SDU}^{SS} > 0$, *if* $0 < r < 1$ *and* $r < c$.

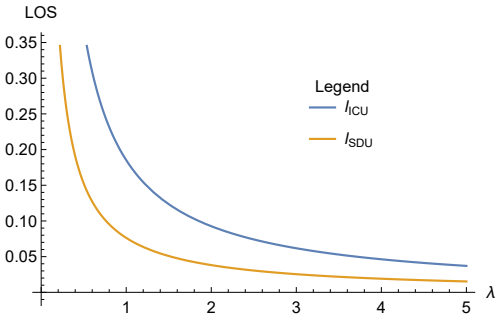
2. $S_{ICU}^{SS} > S_{SDU}^{SS}$, *if* $0 < r < 1$ *and* $0 < c$.



(a) $c = 1.1, \lambda = 1$

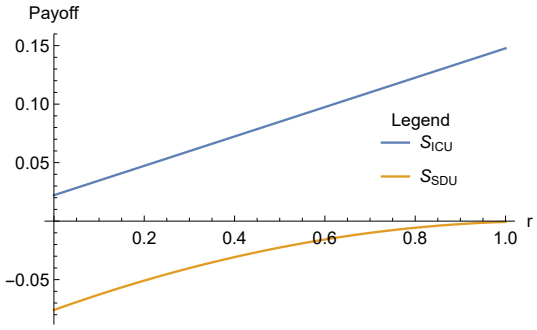


(b) $r = 0.9, \lambda = 1$

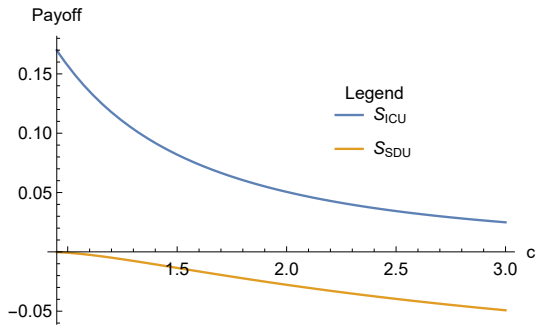


(c) $r = 0.9, c = 1.1$

Figure 4.9: length-of-stays at the ICU and SDU (l_{ICU} in blue and l_{SDU} in orange) respectively under SDU Stackelberg game as a function of (a) SDU benefit, (r), (b) Lost time Cost, (c), and (c) Arrival rate (λ).



(a) $c = 1.1, \lambda = 1$



(b) $r = 0.9, \lambda = 1$

Figure 4.10: Payoffs at the ICU and SDU (S_{ICU} in blue and S_{SDU} in orange) respectively under SDU Stackelberg game as a function of (a) SDU benefit, and (b) Lost time Cost.

Lemma 4.4.4 states that, for the system to require both servers, the unit cost of queuing must be higher than the SDU benefit and the SDU reward must be lower than that of the ICU. This is shown numerically on Fig. 4.9(a),(b) and (c). Under the SDU Stackelberg decision, ICU LOS is always larger than that of the SDU. The ICU has the highest-burden of care when $0 < r < 1$ and $c > r$. From Fig. 4.9(a), and counter-intuitively, increasing the SDU benefit increases the

ICU LOS in blue while that of the SDU in orange decreases. Increasing SDU benefit increases the level of the burden of care at the ICU while decreasing that of the SDU. From Fig. 4.9(b) higher cost reduces the LOS at both servers and the ICU transfers some of its burden to the SDU. From 4.5(c), increasing the demand decreases the length-of-stay exponentially. As seen in Fig. 4.9(c), the demand keeps the burden of care at both servers decreasingly proportional. The SDU payoff can be mostly negative (see Fig.4.10), that is, it does not benefit itself. But given the positive LOS, it benefits the whole system. Increased SDU reward increases the payoff at both servers. Increased cost decreased payoff at both servers.

For the remainder of the paper, we use superscripts CP, ST, IS, and SS to denote that the corresponding quantities are for the CP (Cooperative), ST (Simultaneous decision), IS (ICU Stackelberg), and SS (SDU-Stackelberg) cases, respectively. Apart from the cooperative game, CP, the rest of ST, IS, and SS are competitive games. In the rest of the paper, competition or competitive games refer to these three games.

4.4.2 Further Results and Implications

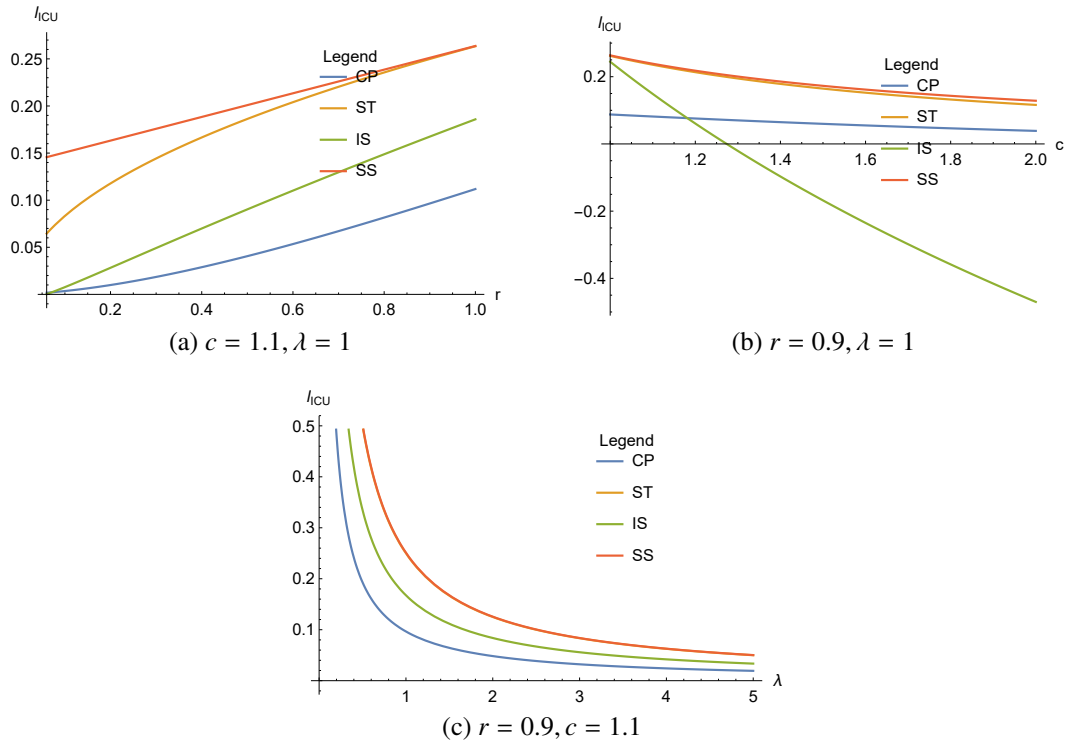
In Subsection 4.4.1, we obtained equilibrium length-of-stays and corresponding payoffs at the two servers under various competitions. In this section, we analyze and discuss the implications of these results on the game structures and compare them to the system with one server. We start with a general comparison between the power structures at both servers. First, we compare the length-of-stays among the different game structures at each server. Then we analyze the whole system and compare the system with two servers to the system with only one server. Due to the complexity of the analytical results, an exhaustive analytical comparison is challenging. To ease the discussion, numerical results are summarized graphically to provide a visual illustration. The discussion here is based on the assumption that the unit cost of waiting in a queue is greater than that of the unit rewards, ($c > 1 > r > 0$). In the figures, when comparing the length-of-stays or the payoffs of four structures at the ICU or the SDU, the Cooperative (CP) line is blue, the Simultaneous decision, (ST), is yellow, ICU Stackelberg, (IS), is green, SDU-Stackelberg, (SS), is orange, and the unique server system lines are in purple.

ICU length-of-stays

In general, Fig.(4.11) shows that when ($c > 1 > r > 0$),

$$l^{SS} \geq l^{ST} \geq l^{IS} \geq l^{CP}. \tag{4.36}$$

In Fig.4.11(a), it is observed that when c and λ are constant, the ICU LOS increases with respect to the SDU benefit under all game structures. The SS consistently gives the highest



(a) $c = 1.1, \lambda = 1$

(b) $r = 0.9, \lambda = 1$

(c) $r = 0.9, c = 1.1$

Figure 4.11: ICU length-of-stays under the various power structure (CP (Cooperative), ST (Simultaneous decision), IS (ICU Stackelberg), and SS (SDU-Stackelberg)) as a function of (a) SDU benefit, (b) cost, and (c) Arrival rate.

burden of care to the ICU followed by the ST, then the IS and finally the CP. The competitive games tend to give a higher burden of care to the ICU. We note also that when the SDU benefit is equal to the ICU benefit, the SS and the ST are equivalent.

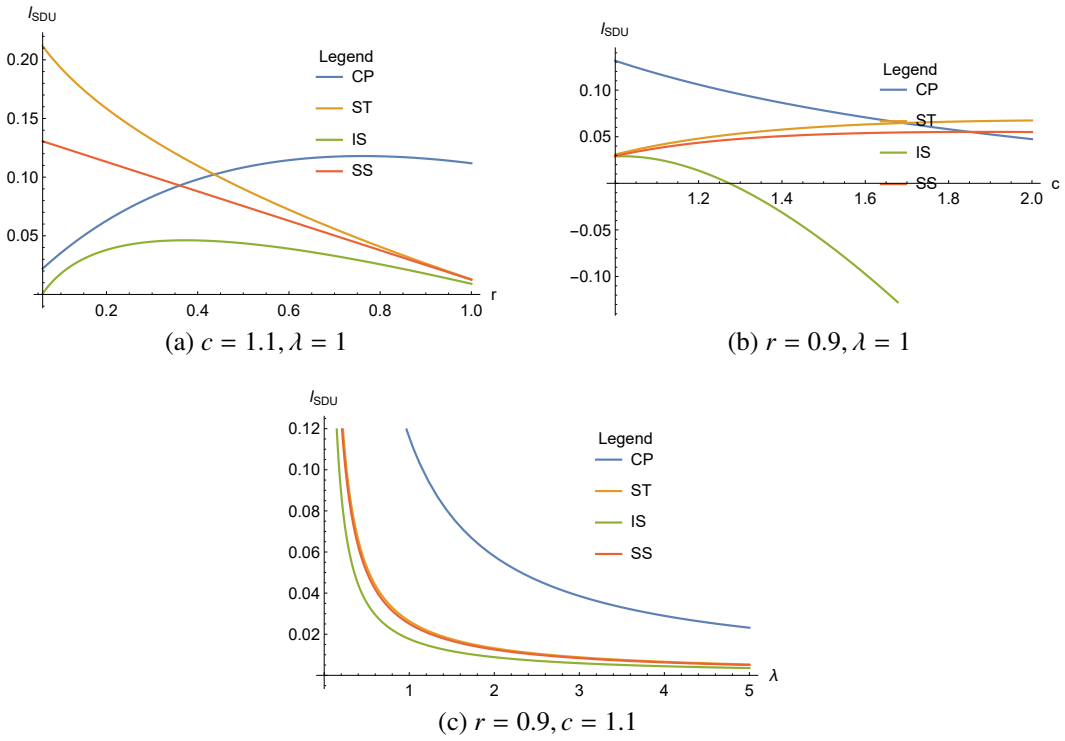
In Fig.4.11(b), it is observed that as expected, when r and λ are constant, the ICU LOS decreases with respect to the cost under all game structures. The IS has a sharper decline while the SS and the ST behave similarly. When the cost is small, close to the ICU reward, the IS gives the highest-burden of care to the ICU followed by the SS, the ST and finally the CP. When the cost is high, the SS followed by the ST has the highest-burden of care in the ICU. When the cost and the ICU benefit are equal, the SS burden and the IS burden are equivalent.

Fig.4.11(c) shows that increasing the arrival rate does not change the structure with the highest burden of care. Therefore, there is no transfer of the weight of the burden of care when $(c > 1 > r > 0)$ and the rate of arrival increases.

SDU's length-of-stays

Fig.(4.12) shows that when the SDU's LOS increases only under cooperation. In general, at the SDU,

$$l^{CP} \geq l^{ST} \geq l^{SS} \geq l^{IS}. \tag{4.37}$$



(a) $c = 1.1, \lambda = 1$ (b) $r = 0.9, \lambda = 1$

(c) $r = 0.9, c = 1.1$

Figure 4.12: SDU length-of-stays under the various power structures (CP (Cooperative), ST (Simultaneous decision), IS (ICU Stackelberg), and SS (SDU-Stackelberg)) as a function of (a) SDU benefit, (b) Queue cost, and (c) Arrival rate.

Under competition, the ST and the SS have a monotonic decrease while the IS is concave. All three decreased to zero where the SDU reward, r, equal the ICU reward. When the ICU reward is equal to the SDU reward, under competition, the SDU's LOS is 0. That is the SDU is often dominated by competition structures. When the reward is small, the ST has the highest-burden of care for the SDU followed by the SS, then the CP and finally the IS. When the reward is high, the CP provides the highest-burden of care followed by the ST, then the SS and finally the IS. The IS consistently allocates a lower burden of care to the SDU.

From Fig.4.12(b), when the cost alone varies, the cooperation game's burden of care decreases while that of the competition increases except the IS. As seen under the SDU benefit, the IS has the lowest burden of care in the SDU followed by the SS, then the ST and finally the CP. When the cost of overstay is equivalent to the reward at the SDU, the SDU LOS is zero under competition.

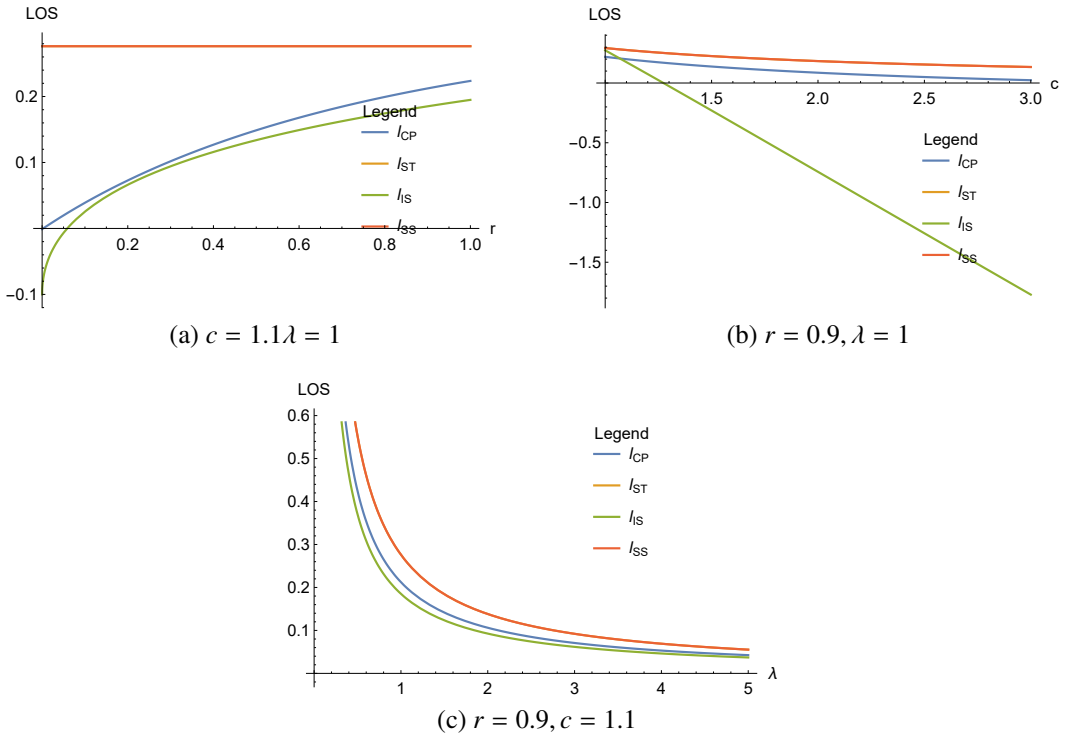
Change in arrival rate only affect change in the length-of-stays not in the dominance of the game structure. This is shown in Fig.4.12(c).

System's Total LOS

In this section, we compare the system's total LOS under all the game structures. In general, when $c > 1 > r > 0$, the system's burden of care can be summarised as follows:

$$l^{SS} = l^{ST} > l^{CP} > l^{IS}.$$

when the l^{ICU} represents the LOS in the system with a unique server.



(a) $c = 1.1\lambda = 1$

(b) $r = 0.9, \lambda = 1$

(c) $r = 0.9, c = 1.1$

Figure 4.13: ICU length-of-stays under the various power structures (CP (Cooperative), ST (Simultaneous decision), IS (ICU Stackelberg), and SS (SDU-Stackelberg)) as a function of (a) SDU benefit, (b) Queue cost, and (c) Arrival rate.

From the three plots in Fig.4.13, we observe that when the cost and the arrival rate are fixed and the reward varies, the system's total LOS under ST, SS and the unique ICU system are equivalent and constant. Increasing the SDU reward does not change the total length-of-stay in the system, only the two servers' LOS have inverse variation. These three configurations have the highest total LOS. The ST, the SS and the unique ICU system, therefore, tend to keep patients longer in the system, increasing the total burden of care. When the cost is equal to the ICU reward then the IS's burden of care is equal to those of the ST, SS and ICU, otherwise, it is higher. The structure with the lowest burden of care is the IS, followed by the CP, then the three mentioned earlier (See Fig. 4.13(c)).

Increasing the SDU benefit increases the system's burden of care under the CP and the IS, while under the SS, the ST and the unique ICU server system, the LOS is constant. Under the SS and the ST, the ICU and the SDU behave like they are playing a constant sum game. In this constant game, they share the unique ICU LOS. Increasing the cost decreases the burden of care and the time spent in the system. The system hastens the service in other to reduce the accumulation of patients in the overstay state.

ICU's Payoff

The ICU's payoff is simply the difference between the benefit of the average time used in the ICU bed and the average penalty to queuing time. Generally, the competitive games' ICU payoffs increase with the SDU benefit, while they decrease in cost. ICU payoff under the co-operation decreases with both SDU benefit and cost. The Payoff is invariant with an increased arrival rate, which is independent of the arrival rate. Generally, the SS dominates at the ICU with the highest payoff, followed by the IS, then the ST game and finally the CP game (see also Fig. 4.14):

$$S_{ICU}^{SS} > S_{ICU}^{ST} > S_{ICU}^{IS} > S_{ICU}^{CP} \tag{4.38}$$

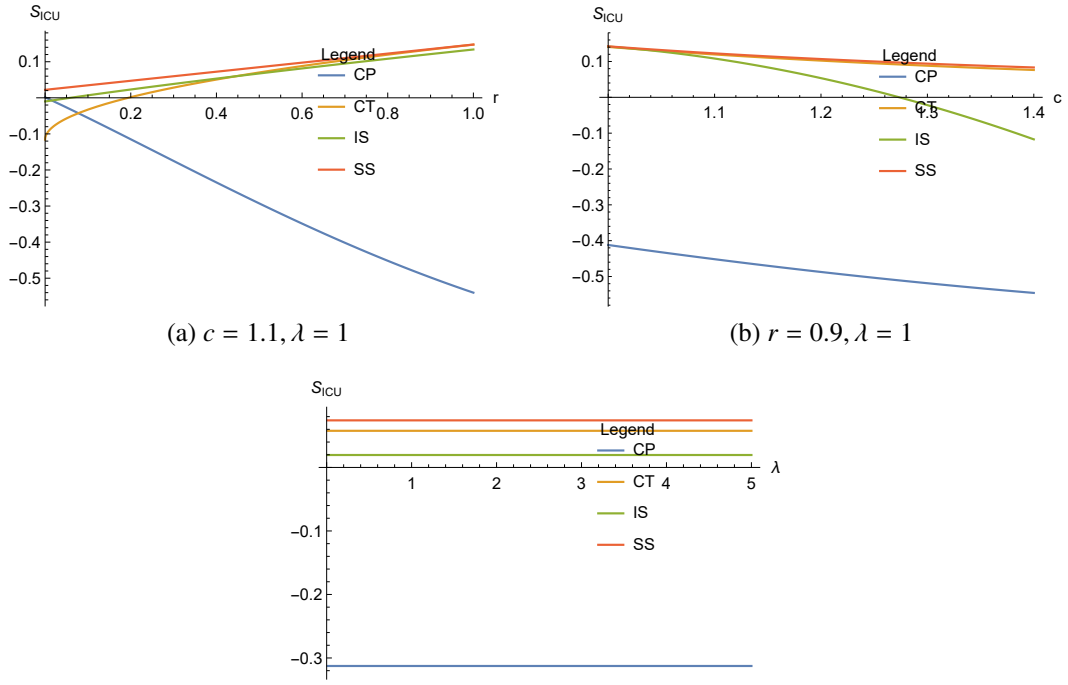
SDU's Payoff

The SDU's payoff is the difference between the SDU's benefit per service time and the overstay per service time penalty. Generally, the competition games' payoff at the SDU are negative. Though the cooperation's payoff increases when the SDU benefit increases, it decreases with cost. The CP has the highest payoff followed by the IS game, then the ST and finally the SS game (see Fig. 4.15).

$$S_{SDU}^{CP} > S_{SDU}^{IS} > S_{SDU}^{ST} > S_{SDU}^{SS} \tag{4.39}$$

System's Payoff

In the whole system, increasing the SDU benefit increases the payoff of every game structure except the CP game (See Fig.4.16(a)). However, increasing the cost decreases the payoffs of every game (See Fig.4.16(b)). In Fig.4.16(c), when the SDU benefit is less than the ICU benefit and the cost higher than the ICU reward, the competitive games are the most beneficial. When the SDU reward is higher than that of the ICU and the cost is higher than the rewards, the competition games dominate. When the Cost is lower than the ICU reward, the SS and the ST



(a) $c = 1.1, \lambda = 1$

(b) $r = 0.9, \lambda = 1$

(c) $r = 0.9, c = 1.1$

Figure 4.14: ICU's Payoff under the various power structures (CP (Cooperative), ST (Simultaneous decision), IS (ICU Stackelberg), and SS (SDU-Stackelberg)) as a function of (a) SDU benefit, (b) Cost, and (d) arrival rate.

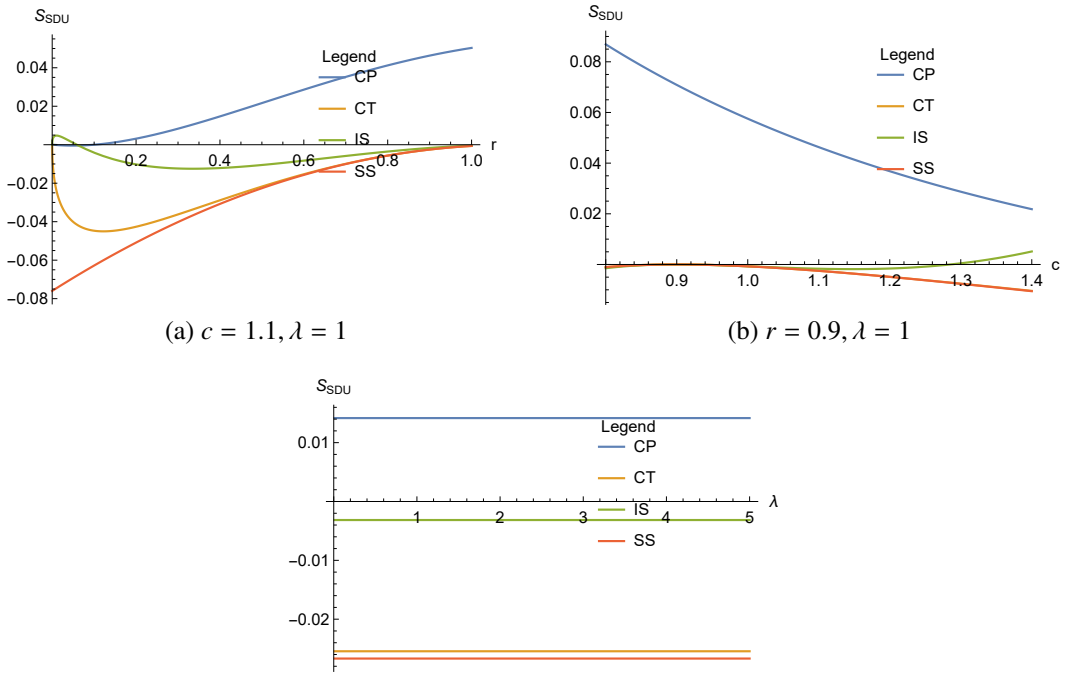
dominate. In general, under the assumption that $c > 1 > r > 0$, cooperation is dominated while the unique system dominates.

There is no direct incentive as a leader as neither the ICU nor the SDU Pareto dominates each other under the ICU and SDU Stackelberg respectively. However, all parties are worse off when no one assumes the leadership since the CP and the ST payoffs are the lowest for both servers. This implies that the system is better off when one of the units is in the leading position.

Comparing Game Structures

The game structure with the smallest burden of care and the highest reward is preferable. The ratio analysis method will be used to compare the performance of the game structures. The ratio payoff per LOS denoted by P produces information on the relationship between one input (LOS) and one output (payoff). That is, efficiency is defined as the number of output units per unit of input. Figs. (4.18), (4.17) and (4.19) are the payoff per LOS plots.

In general, from Figs. 4.17 it can be observed that competition games have the highest ratio in the ICU while cooperation's payoff has the lowest and is negative. The ICU Stackelberg has



(a) $c = 1.1, \lambda = 1$

(b) $r = 0.9, \lambda = 1$

(c) $r = 0.9, c = 1.1$

Figure 4.15: SDU's Payoff under the various power structures (CP (Cooperative), ST (Simultaneous decision), IS (ICU Stackelberg), and SS (SDU-Stackelberg)) as a function of (a) SDU benefit, (b) Cost, and (d) arrival rate.

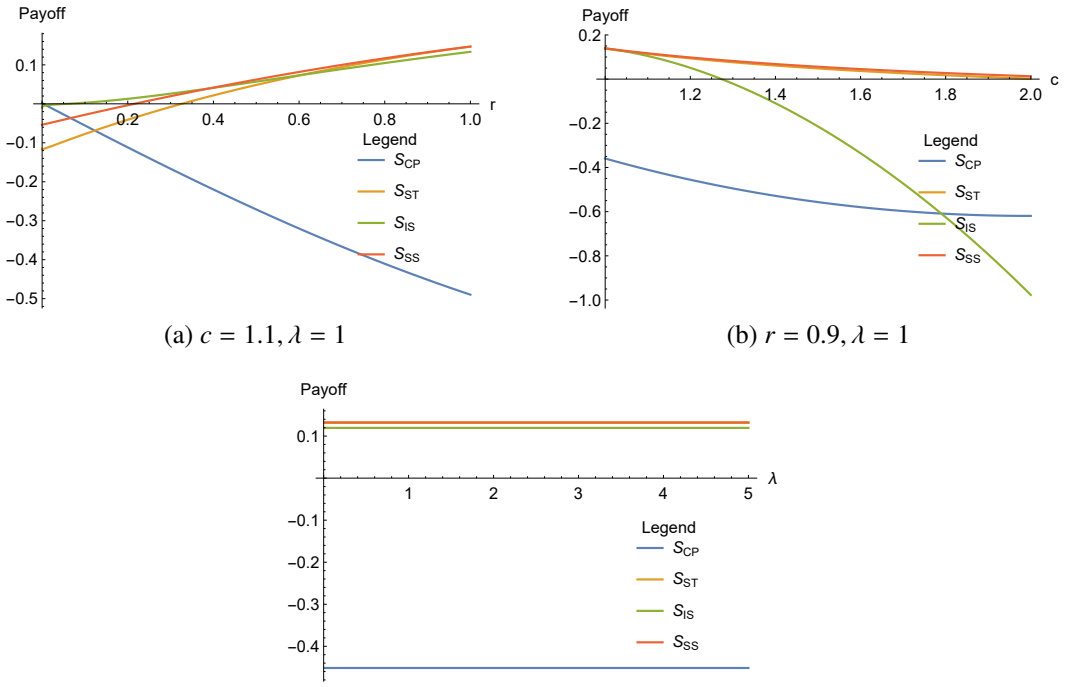
the greatest ratio, followed by the SDU Stackelberg, then the simultaneous decision and finally the cooperation: (see Fig. (4.17)).

$$P_{ICU}^{IS} > P_{ICU}^{SS} > P_{ICU}^{ST} > P_{ICU}^{CP} \tag{4.40}$$

In particular, Figs. 4.17(a) shows that at the ICU, as the SDU reward increases, the ratio increases but that of cooperation decreases to negative. Figs. 4.17(b) shows that as the queueing cost increases, only the ICU Stackelberg's ratio increases and the rest decreases with that of cooperation being negative. Figs. 4.19(c) shows that when the demand increases the competition games have a steady ratio increase and cooperation has a decreasing ratio.

In the SDU, in general, from Figs. 4.18 we can observe that the cooperative game has the highest payoff per LOS in the SDU, while competition payoffs are negative. The cooperation has the greatest payoff per LOS, followed by the ICU Stackelberg, the simultaneous decision and the SDU Stackelberg: (see Fig. (4.18)).

$$P_{SDU}^{CP} > P_{SDU}^{IS} > P_{SDU}^{ST} > P_{SDU}^{SS} \tag{4.41}$$



(a) $c = 1.1, \lambda = 1$

(b) $r = 0.9, \lambda = 1$

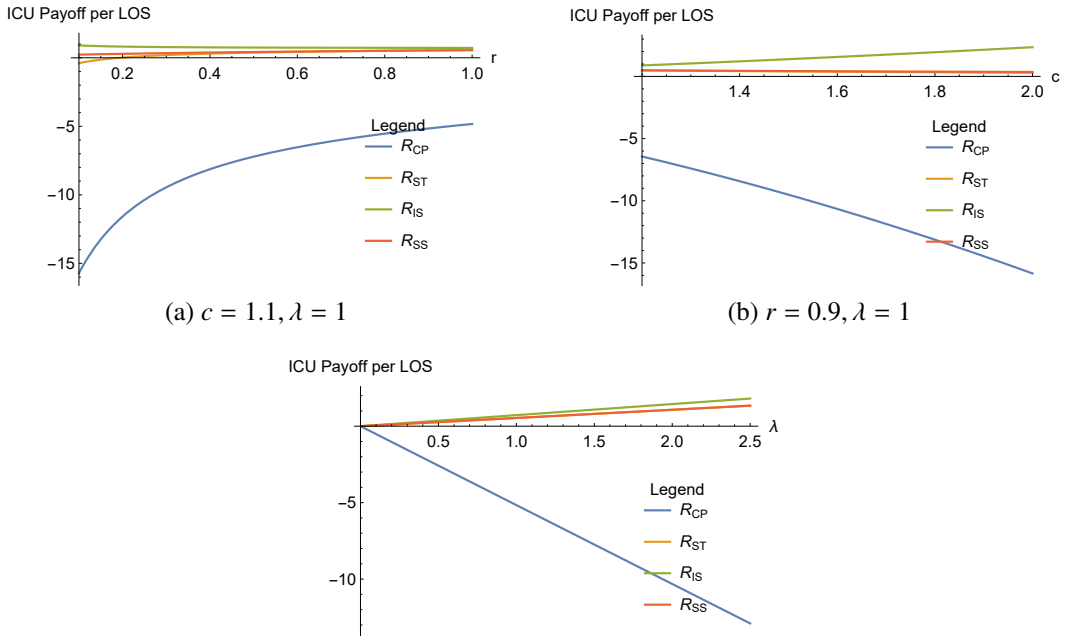
(c) $r = 0.9, c = 1.1$

Figure 4.16: Full System's Payoff under the various power structures (ICU, (Unique server system), CP (Cooperative), ST (Simultaneous decision), IS (ICU Stackelberg), and SS (SDU-Stackelberg)) as a function of (a) SDU benefit, (b) Cost, and (d) arrival rate.

In particular, Figs. 4.17(a) shows that at the ICU, as the SDU reward increases, the payoffs per LOS increase but that of cooperation is negative. Figs. 4.17(b) shows that as the queueing cost increases, only the ICU Stackelberg's payoff per LOS increases and the rest decreases with that of cooperation being negative. Figs. 4.19(c) shows that when the demand increases the competition games have a steady payoff per LOS increase and cooperation has a decreasing payoff per LOS.

Figs. 4.19(a) shows that as the SDU reward increases, competitions payoff per LOS increases but that of cooperation decreases. When the SDU benefit is equal to the ICU benefit, the SDU Stackelberg's payoff per LOS is equivalent to the simultaneous decision's payoff per LOS. Figs. 4.19(b) shows that as the queueing cost increases, only the ICU Stackelberg's payoff per LOS increases and the rest decreases. Figs. 4.19(c) shows that when the demand increases the competition games have an increasing payoff per LOS and cooperation has a decreasing payoff per LOS. In general, the ICU Stackelberg has the greatest payoff per LOS, followed by the SDU Stackelberg, then the simultaneous decision and finally the cooperation: (see Fig. (4.19)).

$$P^{IS} > P^{SS} > P^{ST} > P^{CP} \tag{4.42}$$



(a) $c = 1.1, \lambda = 1$

(b) $r = 0.9, \lambda = 1$

(c) $r = 0.9, c = 1.1$

Figure 4.17: ICU Payoff per LOS under the various power structures (ICU, (Unique server system), CP (Cooperative), ST (Simultaneous decision), IS (ICU Stackelberg), and SS (SDU-Stackelberg)) as a function of (a) SDU benefit, (b) Cost, and (d) arrival rate.

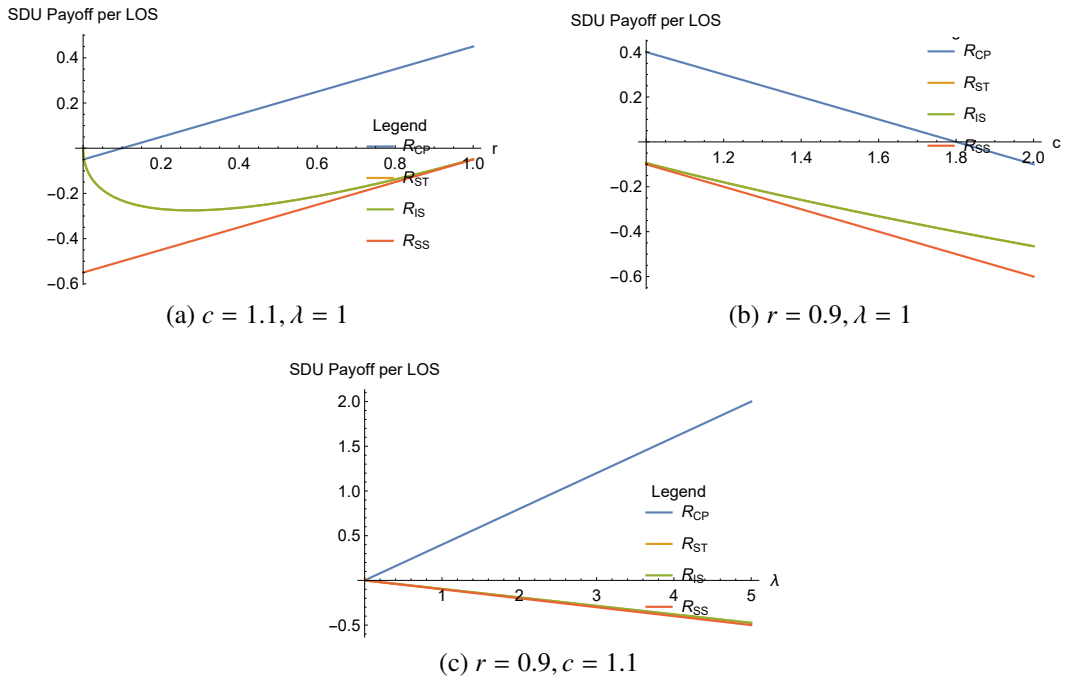
Proposition 4.4.5 *Under the assumption that the SDU reward is less than that of the ICU, the SDU is needful and useful when the unit cost of staying in a queue or overstaying the ICU is higher than the unit benefit at the ICU.*

From Lemmas 4.4.1, 4.4.2, 4.4.3, and 4.4.4, we established that the LOS under all the game structures at the SDU is positive if the queueing unit cost per time is higher than the unit utility per time. Recovering patients have a moderate risk compared to arriving patients therefore the SDU benefit is lower than that of the ICU. And patients that are in service at the ICU have a lower risk compared to arriving patients without any care. This may help explain the fact that the cost is higher than the reward received.

Proposition 4.4.6 *When the SDU benefit is less than that of the ICU and the costs higher than the reward, the ICU Stackelberg provides the highest payoff per unit LOS.*

4.5 Conclusion and Recommendations

While previous studies address customers' decision to join a queue and the servers' decision on pricing to attract customers in parallel servers, in the proposed ICU/SDU system, the servers



(a) $c = 1.1, \lambda = 1$

(b) $r = 0.9, \lambda = 1$

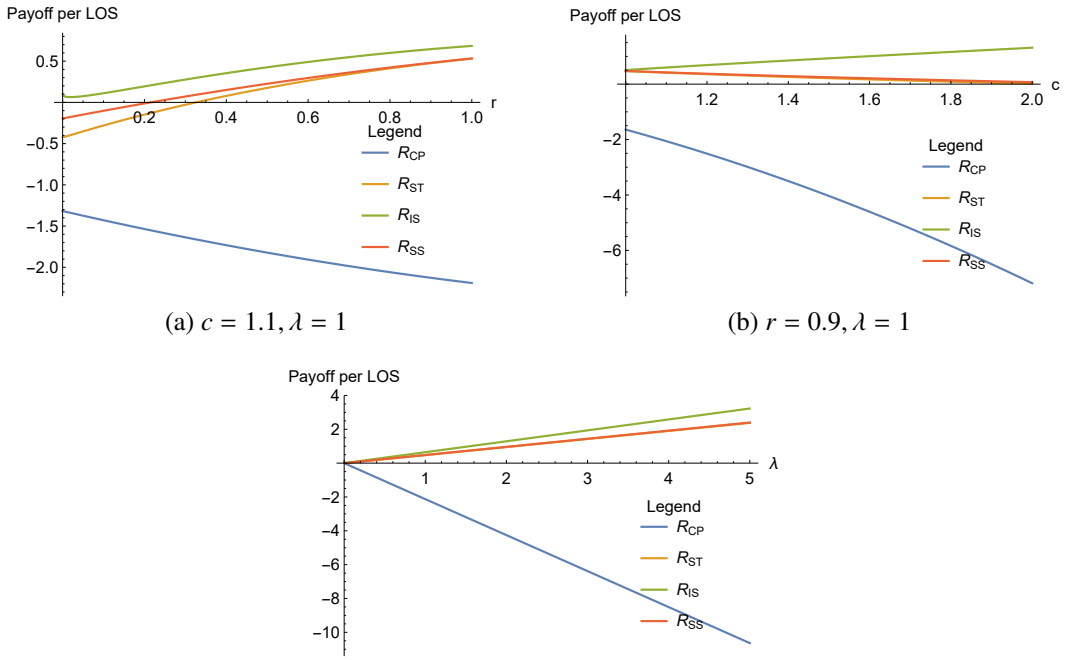
(c) $r = 0.9, c = 1.1$

Figure 4.18: SDU Payoff per LOS under the various power structures (ICU, (Unique server system), CP (Cooperative), ST (Simultaneous decision), IS (ICU Stackelberg), and SS (SDU-Stackelberg)) as a function of (a) SDU benefit, (b) Cost, and (d) arrival rate.

are in a series, and there is no decision on the patient's side. Instead, the hospital is assumed to be a monopoly, and the servers do not compete for the patients but for the time to spend with a patient (LOS).

We considered two traditional forms of power structures: competition and cooperation. We studied competition under three directions (two Stackelberg games and one simultaneous game) and compared the equilibrium payoffs and length-of-stays under four power structures. Furthermore, we determine the conditions for which each structure is feasible. For the ICU, the SDU Stackelberg games provide the highest payoff and the highest burden of care (LOS). For the SDU, the Cooperation tends to provide the highest payoff. The SDU Stackelberg game produced the highest payoff for the whole system under all cases while the simultaneous decision game yielded the lowest reward to the overall system. It is interesting to remark that no leader has a leading privilege under all circumstances.

One may assume that a leader's payoff should generally be larger than that of the follower; however, our results revealed that payoffs depend on the demand and the substitutability in both ICU Stackelberg and SDU Stackelberg games. When the simultaneous game is played, the payoff worsens off at both servers, thus in the system as a whole. Consequently, the simultaneous decision game is the least to be recommended in such a system. Moreover, even though none



(a) $c = 1.1, \lambda = 1$

(b) $r = 0.9, \lambda = 1$

(c) $r = 0.9, c = 1.1$

Figure 4.19: Full System's Payoff per LOS under the various power structures (ICU, (Unique server system), CP (Cooperative), ST (Simultaneous decision), IS (ICU Stackelberg), and SS (SDU-Stackelberg)) as a function of (a) SDU benefit, (b) Cost, and (d) arrival rate.

of the leaders have a leading advantage in all conditions, in the Stackelberg games, the patients spend most of their LOS at the ICU. While this study adds to the growing literature on queuing games, some of the limitations may be present. The formulation of the payoff functions is simplified, albeit already challenging to solve. There's also the lack of patients' decision to join the queue at the ICU, while there is no loss of patients due to other circumstances, i.e. all patients go through both servers in tandem. The use of the M/M/1 queue system, while useful, provides only the "last bed" scenario, and the constant demand rate may be a simplification of a system that may have its arrival rate somewhat service-dependent.

Chapter 5

Invasive Mechanical Ventilation Duration Prediction using Survival Analysis

Abstract

Invasive mechanical ventilation is one of the leading life support machines in the intensive care unit (ICU). By identifying the predictors of ventilation time upon arrival, important information can be gathered to improve decisions regarding capacity planning and patient care.

In this study, first-day ventilated patients' ventilation time was analyzed using survival analysis. The probabilistic behaviour of the ventilation time duration was analyzed and the predictors of the ventilation time duration were determined based from available first day covariates.

A retrospective analysis on ICU ventilation time in Ontario was performed with data about ICU patients obtained from the Critical Care Information System (CCIS) in Ontario between July 2015 and December 2016. As part of the procedure for inclusion, a patient must be connected to the invasive ventilator upon arrival to the ICU. Parametric survival methods were used to characterize ventilation time and determined covariates associated with ventilation time. Parametric and non-parametric methods were used to determine predictors of ventilation duration of the first-day ventilated patients.

A total of 128 030 patients visited the ICUs between July 2015 and December 2016. 51 966 (40.59%) patients received invasive mechanical ventilation on arrival. The analysis of the duration of ventilation suggested that the log-normal distribution provided a better fit for the ventilation time, whereas the log-logistic Accelerated Failure Time model best describes the association between the covariates and the duration of ventilation. ICU site, admission source, admission diagnosis, scheduled admission, scheduled surgery, referral physicians, central venous line

treatment, arterial line treatment, intracranial pressure monitor treatment, extra-corporeal membrane oxygen treatment, intraaortic balloon pump treatment, other interventions, age group, pre-ICU LOS, and MODS score were significant predictors of the ICU ventilation time.

The results show substantial variability in ICU ventilation duration for different ICUs, patient's demographic, underlying condition, and underline mechanical ventilation as an important driver of ICU stay.

The prediction performance of the proposed model showed that both the model and the data can be used to predict individual patient's ventilation time and provide insight into the predictors of ventilation time.
Keywords: Mechanical ventilation, Ventilation duration, Survival analysis, ICU

5.1 Introduction

Mechanical ventilation (MV) is defined as the use of a breathing support machine that takes over the breathing process in patients who cannot breathe properly on their own. MV is one of the life support alternatives the Intensive Care Unit (ICU) provides that differentiate it from other hospital units. There are two forms of MV: invasive and non-invasive. Non-invasive ventilation (NIV) is the delivery of oxygen via a face mask. According to Hyzy and McSparron [73], invasive mechanical ventilation (IMV) is "the delivery of positive pressure to the lungs via an endotracheal or tracheostomy tube." During IMV, a ventilation machine (also called a ventilator) forces a predetermined mixture of air (i.e., oxygen and other gases) into the central airways that then flows into the alveoli [73, 103]. A patient may need a ventilator when there is a low oxygen level in the blood or severe shortness of breath from an infection such as pneumonia, SARS or COVID-19. Over 20 million patients worldwide per year use mechanical ventilation [164, 3].

Since the outbreak of the COVID-19 pandemic, one of the most globally cited causes for the inability for ICUs to manage patients appropriately is the lack of ventilators [16, 52, 124, 15, 75]. However, the concern about an insufficient supply of ICU beds and ventilators to handle critically ill patients is old. COVID-19 sparked a debate on when and how ventilators should be used within the ICU. Before COVID-19, there have been very few publications about modelling the use, demand, and practice of ventilation on ICU patients. Kacmarek [77] gave a detailed history of the evolution of these human breathing aids in medicine. Ventilators are not built-in to the ICU beds like other organ support machines. Nevertheless, as many as 90% of ICU patients required ventilation [124]. Previous papers in the literature associate ICU ventilators use to the ICU bed use, but not all patients in the ICU use ventilators, and their use is patients' state-dependent.

From a managerial perspective, it is important to know the distribution of the time patients are reliant upon IMV to predict ventilation demand. In this case, survival models are often used to study the time to stop the invasive ventilation. The application of survival analysis is extensive in health care. However, the application of survival models on ventilation time is less common. Nevertheless, up to our knowledge, no study has analyzed ventilation time using AFT survival models.

Most of the studies on ventilation time in the literature were interested in classifying the duration of mechanical ventilation into two categories: prolonged versus short ventilation time [161, 55, 46, 1, 56, 162, 10]. Prolonged mechanical ventilation (PMV) was defined differently within each study. Estenssoro et al. [55] described it as being mechanical ventilation for longer than 21 days, [46] define PMV as greater than seven days, [162] define it as patients receiving ventilation for longer than three days, and [92] describe it as greater than a day. Logistic regression, linear regression and machine learning methods are the primary tools used in the literature to model ventilation time and identify significant predictors of PMV.

Dimopoulou et al. [46] investigated PMV in patients with blunt thoracic trauma and found that advancing age, the severity of head injuries, and bilateral thoracic injuries were significant in predicting PMV. Trouillet et al. [162] gathered data on patients undergoing cardiac surgery. They found that a post-operative score could be used to identify patients eligible for rapid weaning of ventilation on day three, which reduces the need for PMV. Légaré et al. [92] took a group of coronary artery bypass grafting patients, identified the predictors of PMV, and found that intra-operative complications significantly impact those patients who required prolonged mechanical ventilation. Trouillet et al. [163] investigated the outcomes of two groups of severely ill patients who required mechanical ventilation. One group received an early percutaneous tracheotomy and the other received prolonged intubation. Upon comparing the two treatments, it was found that early tracheotomy provided no benefit in terms of mortality rates or length-of-stay. Esteban et al. [54] found that factors at the start of mechanical ventilation and complications of critical illness influence the outcome of patients receiving mechanical ventilation.

Logistic regression is used to model the probability of an event, and, therefore, it cannot be used to predict ventilation time. Moreover, when studies used linear regression, the results were often unreliable. Seneff et al. [143] analyzed an individual patient's duration of mechanical ventilation using linear regression and found an R^2 of 0.18. Aung et al. [10] used multiple linear regression to identify variables independently associated with the duration of mechanical ventilation and obtained an R^2 of 0.235.

Abujaber et al. [1] and Sayed et al. [140] used machine learning models but could not find direct relationships between the predictors and the duration of mechanical ventilation.

Abujaber et al. [1] started with logistic regression and built upon this by creating Artificial Neural Networks, Support Vector Machines, Random Forests, and Decision Trees to predict prolonged mechanical ventilation. Ultimately, Support Vector Machines gave the best prediction technique in this study, with an accuracy of 0.79. Sayed et al. [140] attempted to predict the duration of mechanical ventilation using machine learning models. When using the Light Gradient Boosting Machine, it was found that predictors gathered before the third ICU day could be used, allowing for mechanical ventilation to be predicted earlier than other machine learning models, such as random forest and extreme gradient boosting.

In this work, we predict ventilation duration in the ICU using patient information available on arrival (day 1) using survival analysis techniques. We identified the distribution of ventilation duration and evaluated whether differences exist in terms of first-day treatments, ICU sites, admission source, referral physician, patient category, sex, NEMS, and MODS. Our study demonstrates the importance of patient information available at arrival when predicting the risk of ventilation duration amongst ICU patients. Thus, the importance of these results in ventilation planning and management.

5.2 Methods

5.2.1 Study Design and Data Collection

This is a retrospective study designed to predict the time distribution of ICU patients that were connected to IMV on arrival using information available on arrivals, such as Demographics, First-day treatments, NEMS and MODS scores. The outcome is to determine the predictors of ventilation time after the first day and predict the duration of ventilation. Survival analysis was performed using a large dataset of variables obtained from the Critical Care Information System (CCIS) Ontario database.

The CCIS dataset contains information from July 2015 to December 2016. The data has forty-five variables. Priestap et al. [128] used the CCIS dataset to predict ICU mortality and provide detailed information on the data collection procedure, the variables and the ICUs in the CCIS database. The following subset covariates on patients' arrival are used in our study: *Basic Monitoring, Central Venous Line, Arterial Line, Intracranial Pressure Monitor, Dialysis, Extra-corporeal Membrane Oxygen, Intra-aortic Balloon Pump, Other Interventions Within this Unit, Interventions Outside this Unit, the nursing workload proxy by the First-day NEMS score, demographic information (Age, Sex), the MODS Score, the admission sources, the admitting diagnosis, referral physician specialist, and patient category*. For model external validation, data from London Health Science Center (LHSC) are also used in this study. The

Table 5.1: First-day ventilation frequency

Ventilation status	Count	Percentage
Mechanical: Invasive Ventilation	51 966	40.59
Mechanical: Non-Invasive Ventilation	8 917	6.96
No Ventilation	28 650	22.38
Supplementary Ventilator Care	38 497	30.07
Total	128 030	100

LHSC datasets contains information from January 2020 to May 2021.

The data were pre-processed to remove duplicates, transform variables and create new variables needed for the analysis. The data contain records of 128 030 patients admitted into the ICUs. 40.59% (51 966) of those patients received IMV on arrival. Table 5.1 presents the summary of the patient's ventilation status on arrival. From the patients that received IMV upon arrival, we excluded those with missing information and obtained 49 703 (i.e., 38.82% of total ICU patients). Further, we excluded those with ventilation time greater than 60 and we used the 49 467 (99.53 % of patients connected on arrival) remaining. Table 5.2 shows the sex and censoring distribution of patients' information used in this research. Censored patients include those discharged to the Complex Continuing Care Facility, other hospitals, the Level 3 Unit, and Outside ICU while on the ventilator.

Table 5.2: Sex and censor status distribution of used data

Ventilation status	Count	Percentage
Female	18 185	36.76
Male	31 282	63.24
Censored	1 407	2.84
Uncensored	48 060	97.16
Total	49 703	100

This study was approved by the Research Ethics Review Committee at King's University College at Western University, Principal Investigator, Prof. Felipe F. Rodrigues. De-identified data and restrict access and storage of the data is mantained per ethics protocol guidelines.

5.2.2 Statistical analysis

Descriptive analysis

Descriptive statistical analysis was performed, reporting the following measures for continuous variables: mean, standard deviation, skewness, kurtosis, and quartiles. For the modelling,

the continuous variables were grouped into categories. Categorical variables are reported as counts and percentages.

Non-parametric survival analysis

Kaplan-Meier analysis was used to construct non-parametric survival curves based on patients' age and sex. The log-rank (Mantel-Cox) test and the Kruskal-Wallis test were used for multiple comparisons between sub-groups.

Parametric survival analysis

Different classical distributions were fitted to the observed ventilation time to identify the probability distribution function (PDF) that best fit ventilation time. To assess goodness-of-fit, P-P plots, as well as three regularly used goodness-of-fit tests: Kolmogorov-Smirnov, Anderson-Darling, and Chi-Squared were employed. Appropriate maximum likelihood estimates of the parameters were obtained with their respective 95% confidence limits based on the probability distribution function. Parametric Accelerated Failure Time (AFT) modelling was done by randomly dividing the dataset a "training" and a "testing" set (training with 70% of the observations and testing set with 30%). The training set was used for modelling and the test set was used for model prediction performance. External data of different years gathered from the London Health Science Center (LHSC) were used to validate the model.

All covariates included in the analysis were obtained on arrival and are included based on their availability, clinical relevance, statistical significance, and possible association with ICU LOS or mortality in the literature. We followed the variable selection approach as outlined in Collett [39]. This approach fits a univariate model for each covariate, identifying significant predictors, then, a multivariate model is fit with all significant univariate predictors, eliminating insignificant variables using backwards selection. Graphical methods, the likelihood ratio, AIC and BIC criteria were used to compare and select the AFT models (Exponential, Weibull, Log-normal, and Log-logistic). The validity of the model was ascertained using external data. All tests presented are two-sided, and a p-value < 0.05 is considered significant.

Prediction performance

To assess prediction performance, we applied the best fitted model on the test set and considered the following metrics: Mean Squared Error (MSE), Mean Absolute Error (MAE), Percent bias (PBIAS), and Nash-Sutcliffe efficiency (NES). These metrics are calculated as follows.

$$MSE = \frac{1}{N_T} \sum_{i=1}^{N_T} (y_i - \hat{y}_i)^2,$$

$$MAE = \frac{1}{N_T} \sum_{i=1}^{N_T} |y_i - \hat{y}_i|,$$

$$PBIAS = \frac{\sum_{i=1}^{N_T}(y_i - \hat{y}_i)}{\sum_{i=1}^{N_T} y_i} \times 100, \text{ and}$$

$$NES = 1 - \frac{\sum_{i=1}^{N_T}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N_T}(y_i - \bar{y}_i)^2},$$

where N_T is the total number of observations in the test set, y_i and $\hat{y}_i$ are respectively the observed and predicted time on ventilation for the ith observation in the test set.

Statistical software

The data were analyzed using using R version 4.1.2. Statistical modelling employed the *glmnet*, *flexsurv*, *SurvRegCensCov*, *survival*, and *surminer* R packages.

5.3 Results

5.3.1 Descriptive Analysis

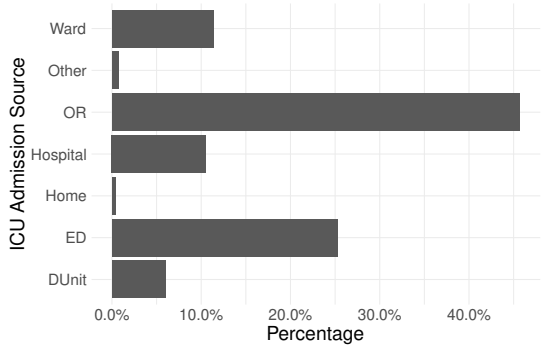
Table 5.3 summarizes the number and proportion of patients that received each of the various treatments at ICU arrival. The most common were basic monitoring (99.93 % of patients), an arterial line (81.76 %), and a central venous line (72.95 %). Variation in treatment patterns showed that 98.34% had no intracranial pressure monitoring, 97.13% had no dialysis, 99.73% had no extracorporeal membrane oxygen, 98.59% had no intra aortic balloon pump 67.38 % had no other interventions within this unit and 78.15% had no interventions outside this unit.

Table 5.3: Distribution of treatments IMV patients received

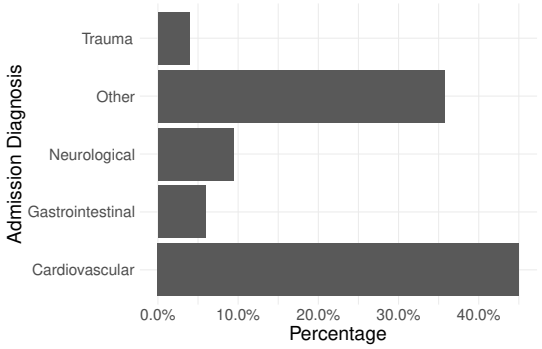
Treatment	No	Yes
Basic Monitoring	33 (0.06 %)	49 434 (99.93%)
Arterial Line	9 065 (18.33%)	40 402 (81.67 %)
Central Venous Line	13 383 (27.05 %)	36 084 (72.95%)
Other Interventions Within this Unit	33 331 (67.38 %)	16 136 (32.62%)
Interventions Outside this Unit	38 658 (78.15 %)	10 809 (21.85 %)
Dialysis	48 046 (97.13 %)	1 421 (2.87 %)
Intracranial Pressure Monitor	48 648 (98.34%)	819 (1.66%)
Intra Aortic Balloon Pump	48 771 (98.59%)	696 (1.41%)
Extracorporeal Membrane Oxygen	49 331 (99.73%)	136 (0.27%)

Fig 5.1 shows the bar plot of the admission sources, admitting diagnosis, referral physician specialist, and patient category. The admission sources included the ward (5 607, 11.33%), downstream units (i.e., Level 2 and Level 3 units) (2 971, 6.00%), the emergency department (ED) (12 507, 25.28%), home (191, 0.39%), hospital outside and within (5 210, 10.53%), the

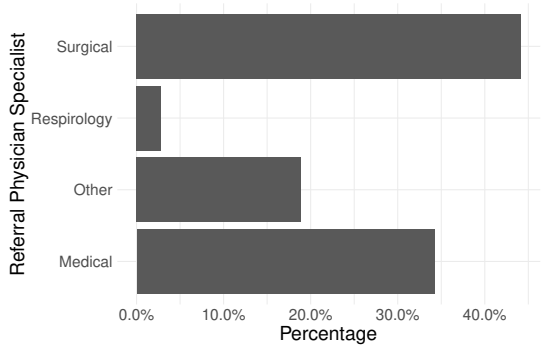
operation room (OR) (22 608, 45.70%), and other sources (373, 0.75%). The other sources include complex continuing care facilities, rehabilitation centers, outside provinces, and others. ICU admitting diagnosis were categorized as Cardiovascular (22 269, 45.00%), Gastrointestinal (2 920, 5.90%), Neurological(4 652, 9.40), Trauma (1 927, 3.90%), and Other (17699, 35.78). Other diagnosis includes patients with the following diseases: Genitourinary, Metabolic, Endocrine, Musculoskeletal, Skin, Oncology, Haematology and Other. Referral physician specialists were grouped into medical (16930, 34.22%), respiratory (1 364, 2.76%), surgical (21 848, 44.17%), and other (9 325, 18.85%). Other referral physician specialist includes Dermatology, Psychiatry, Oncology, Haematology, Ophthalmology, Orthopaedic, and others.



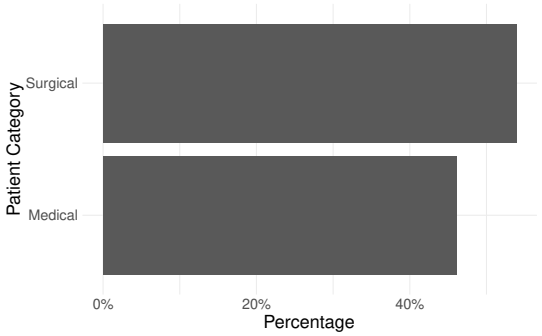
(a) ICU admission source (other source includes Complex Continuing Care Facility - Within and Outside, Inpatient - Rehab, Outside province, Rehab Facility - Within and Outside, and Other - Outside and Within))



(b) ICU admission diagnosis (Other diagnosis includes patients with the following diseases: Genitourinary, Metabolic, Endocrine, Musculoskeletal, Skin, Oncology / Haematology and Other.)



(c) Referral physician specialist. (Other referral physician specialist includes: Dermatology, Psychiatry, Oncology, Haematology, Ophthalmology, Orthopaedic, and other.)



(d) Patient category

Figure 5.1: Variables levels.

Table 5.6 provides the descriptive statistics for the continuous variables. Table 5.5 provides

the descriptive statistics of ventilation time under various patient categories. On average, ventilation time was 4.57 (sd = 6.57) days, NEMS score was 29.09 (sd = 6.85), and MODS score was 5.57 (sd = 3). Table B.2 tabulates the baseline characteristics of the number of events under each category and the result of the Log-rank test, which compares the differences in ventilation times between the independent groups of each covariate.. For ease of readability, the results of the Log-rank test are shown in Table 5.7.

5.3.2 Non-parametric Analysis

As a preliminary analysis, we conducted a non-parametric analysis of the entire dataset using the Kaplan-Meir method. Figure 5.2 shows the Kaplan-Meier (KM) curve of ICU ventilation time. The KM curve shows the unconditional probability that a subject will experience the event beyond time t but does not indicate the proportion of subjects surviving to time t. In our case, survival means becoming independent of ventilation by time t.

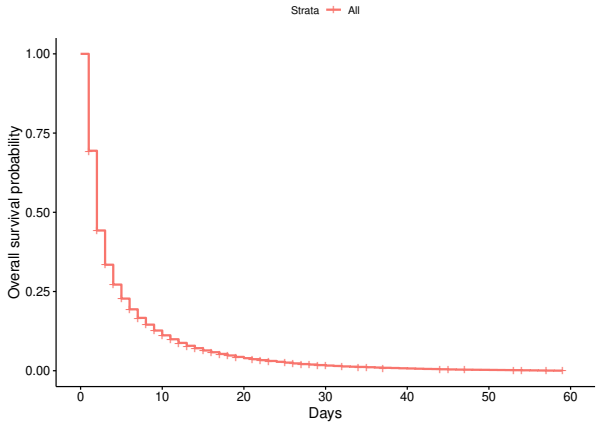


Figure 5.2: Ventilation time KM survival curve.

Table 5.4 tabulates selected KM survival probabilities at some specific times and their associated confidence interval. From Fig 5.2 and Table 5.4, the probability that a person will remain connected to the IMV longer than one day is approximately 0.69. That is 31% get disconnected from mechanical ventilation after the first day. The probability of patients remaining connected beyond 10 days is about 0.112, indicating 88.8% of ICU patients get disconnected after ten days of IMV.

There were 18 185 (36.76 %) females with a mean ventilation time of 4.84 (sd = 6.82) and 31 282 (63.23%) males with mean ventilation time 4.41 (sd = 6.41). There was a significant difference found between the two sexes (χ^2 p-value $< 2e-16$). Investigating age, 4 355 patients (8.80%) aged 18 to 39 had a mean ventilation time of 4.91 (sd = 6.47), 38 009 (76.84%) aged 40 to 79 with a mean ventilation time of 4.54 (sd = 6.61), and 7 103 aged 80 and above with

Table 5.4: Selected survival estimates from the KM curve.

t	n.risk	n.event	n.censor	surv (95% CI)
1	49 467	15 125	646	0.694 (0.698, 0.690)
7	9 100	1 270	28	0.167 (0.170, 0.163)
10	5 879	698	20	0.112 (0.115, 0.109)
30	791	68	2	0.016 (0.018, 0.015)
50	119	12	0	0.003 (0.003, 0.002)
55	56	12	0	0.001 (0.001, 0.001)

n.risk, n.event, n.censor, and surv are number at risk, number of events, number censored, and Survival probability at time t.

a mean ventilation time of 4.51 (sd = 6.36). There was a significant difference found between the age groups (χ^2 p-value = $3e-08$). The admission sources with the highest ventilation times included the Other 373 (0.75%), a mean ventilation time of 7.20 (sd = 8.52) and downstream 2 971 (6.00%), a mean ventilation time of 6.93 (sd = 8.16). There was a significant difference found between the sources of admission (χ^2 p-value $< 2e-16$). There was also a significant difference found between diagnosis, referral physician, and patient category groups with a χ^2 p-value $< 2e-16$. Patients with trauma (1 927 (3.90%) had a mean ventilation time of 6.34 (sd = 7.13)) and had the highest average ventilation time and Cardiovascular patients (22269 (45.02%) had a mean ventilation time of 2.87 (sd = 4.77) and was the lowest. The MODS and NEMS scores were both significantly associated with ventilation time (χ^2 p-value = $3e-08$). Pre-ICU hospital stay was significantly associated with ventilation time (χ^2 p-value $< 2e-16$). The MODS and NEMS scores were both significantly associated with ventilation time (χ^2 p-value $< 2e-16$). Admission and surgery schedule were both significantly associated with ventilation time (χ^2 p-value $< 2e-16$). Among the recorded treatment received at arrival, only the basic monitoring had no significant association with ventilation time (χ^2 p-value =0.8). This can be explained by the fact that the majority of the patients (99.93%) received basic monitoring. This covariate will be removed in further analysis.

In Fig 5.3 and Fig 5.4, the log-rank test of the difference shown in Table 5.7 is confirmed with the distinction in the ventilation time of the various category for each covariate.

Table 5.5: Descriptive statistics of ventilation time under various patient categories.

Variables		Count	Mean	sd	Skew	Kurt	$Q_{0.25}$	$Q_{0.5}$	$Q_{0.75}$
Vent days		49467	4.57	6.57	3.65	17.02	1	2	5
(Sex)	Female	18185	4.84	6.82	3.59	16.35	1	2	5
	Male	31282	4.41	6.41	3.69	17.39	1	2	4

(NEMS)	0 – 22	629	4.79	7.21	3.75	17.09	1	2	5
	23 – 29	11823	4.03	6.06	4.05	21.11	1	2	4
	≥ 30	37015	4.74	6.70	3.55	16.00	1	2	5
(Age group)	0 – 39	4355	4.91	6.47	3.30	14.18	2	2	5
	40 – 79	38009	4.54	6.61	3.67	17.06	1	2	5
	≥ 80	7103	4.51	6.36	3.77	18.65	1	2	5
(MODS)	1	4389	2.80	4.66	5.44	39.28	1	1	2
	1 – 4	14185	4.02	6.09	4.22	23.00	1	2	4
	5 – 8	16070	4.89	6.75	3.40	14.63	1	2	5
	9 – 12	6861	5.94	7.73	3.00	11.10	2	3	7
	≥ 13	7962	4.71	6.52	3.55	16.28	1	2	5
Source	Downstream	2971	6.93	8.16	2.57	8.03	2	4	8
	ED	12507	5.50	6.78	3.24	13.79	2	3	6
	Home	191	5.75	6.07	2.06	4.38	2	3	7
	Hospital	5210	6.81	8.08	2.83	9.86	2	4	8
	OR	22608	2.68	4.36	5.91	46.25	1	2	2
	Other	373	7.20	8.52	2.65	8.15	2	4	9
	Ward	5607	6.54	8.40	2.80	9.32	2	3	8
Diagnosis	Cardiovascular	22269	2.87	4.77	5.61	41.35	1	2	2
	Gastrointestinal	2920	4.76	5.97	3.30	13.99	2	3	5
	Neurological	4652	5.47	6.61	2.96	11.55	2	3	6
	Other	17699	6.24	7.88	2.98	10.80	2	3	7
	Trauma	1927	6.34	7.13	2.54	8.77	2	4	8
Referral	Medical	16930	5.91	7.33	3.09	11.94	2	3	7
	Other	9325	5.59	7.24	3.18	12.86	2	3	6
	Respirology	1364	7.94	9.72	2.66	8.00	2	4	9
	Surgical	21848	2.89	4.74	5.22	35.65	1	2	2
Pat Category	Medical	22811	6.10	7.60	3.05	11.58	2	3	7
	Surgical	26656	3.26	5.19	4.68	28.73	1	2	3

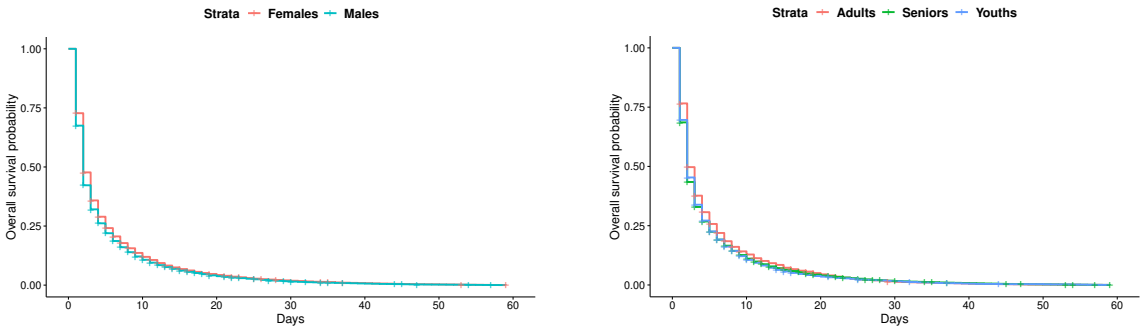
Table 5.6: Descriptive statistics of continuous variables.

Variables		Count	Mean	sd	Skew	Kurt	$Q_{0.25}$	$Q_{0.5}$	$Q_{0.75}$
Day 1 NEMS		49467	35.22	6.69	0.36	-0.36	28	34	39
	0 – 22	629	20.93	0.61	-11.6	153.35	21	21	21
	23 – 29	11823	27.01	0.19	-0.41	64.71	27	27	27

	≥ 30	37015	38.08	5.16	0.86	0.07	34	38	40
Age		49467	63.84	15.88	-0.70	0.20	55.25	66.05	75.29
	18 – 39	4355	29.19	5.95	-0.10	-1.19	24.14	29.47	34.38
	40 – 79	38009	63.85	10.00	-0.42	-0.64	56.75	65.00	71.96
	≥ 80	7103	85.02	3.79	0.93	0.57	81.95	84.23	87.32
Pre ICU LOS		49467	9.05	102.61	25.76	845.44	0	0	2
	≤ 1	35911	0.19	0.40	1.54	0.36	0	0	0
	2 – 7	7661	3.85	1.71	0.48	-1.07	2	4	5
	> 7	5895	69.78	290.10	8.93	100.69	10	15	29
MODS Score		49467	5.57	3	0.42	0.24	4	5	7
	1	4389	0.58	0.49	-0.33	-1.89	0	1	1
	1 – 4	14185	3.23	0.81	-0.43	-1.33	3	3	4
	5 – 8	16070	6.87	0.80	0.23	-1.41	6	7	8
	9 – 12	6861	9.95	1.01	0.67	-0.75	9	10	11
	≥ 13	7962	6.06	2.95	2.51	4.62	5	5	5

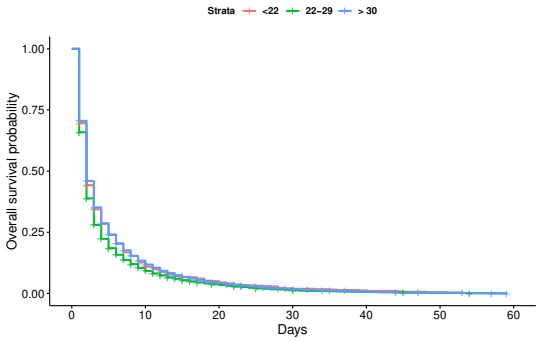
Table 5.7: Log-rank test of equality between groups in covariates

Covariate	χ^2	df	p-value
Pre ICU LOS	262	2	<2e-16
Age group	35	2	3e-08
Sex	77.8	1	<2e-16
ICU Site	3350	65	<2e-16
Admission source	6875	6	<2e-16
Diagnosis	5213	4	<2e-16
Referal physician	5171	3	<2e-16
Patient Category	3929	1	<2e-16
MODS	1107	4	<2e-16
NEMS	157	2	<2e-16
Schedule admission	9579	1	<2e-16
Schedule surgery	8809	1	<2e-16
Basic Monitoring	0.1	1	0.8
Central Venous Line	53.9	1	2e-13
Arterial Line	79.4	1	<2e-16
Intracranial Pressure Monitor	227	1	<2e-16
Dialysis.	194	1	<2e-16
Extra corporeal Membrane Oxygen	79.7	1	<2e-16
Intra Aortic Balloon Pump	22.5	1	2e-06
Other Interventions Within this Unit	878	1	<2e-16
Interventions Outside this Unit	554	1	<2e-16

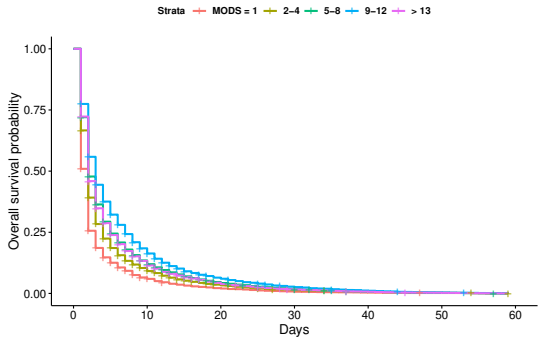


(a) Sex subgroups

(b) Age subgroups



(c) NEMS subgroups



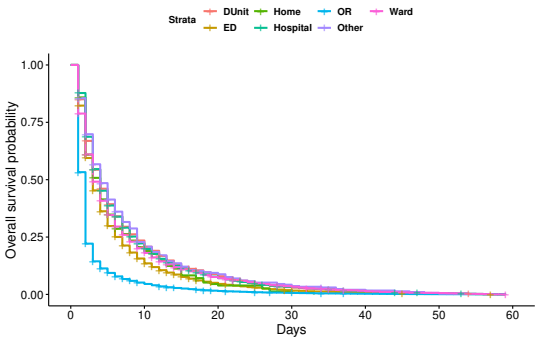
(d) MODS subgroups

Figure 5.3: Kaplan-Meier survival curves

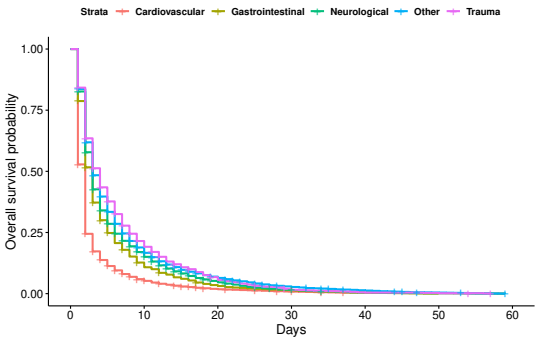
5.3.3 Probabilistic Characterization of ICU Ventilation Time

Here, we performed a parametric analysis of the ventilation time to determine the best distribution that fit the data. Fig 5.5(a) shows the plot of the negative log of the estimated survivor function against time. It shows approximately linear curve trends with minor deviation at the extremes. This suggests that the exponential distribution might be a good candidate. Fig 5.5(b) is the plot of the log of the negative log of the estimated survivor function against log time. It shows an approximately linear trend and suggesting that the Weibull distribution should be considered a good candidate. Fig 5.5(c) shows the plot of the cumulative probabilities versus log time. We observe a concave trend with a faulty linear fit suggesting that the log-normal distribution should be investigated further. Fig 5.5(d) shows the log of the survival probability versus the log of time in black, with a fitted linear model. It shows that the linear trend does not fit appropriately. This implies that the logistic model is not a good fit for the ventilation data.

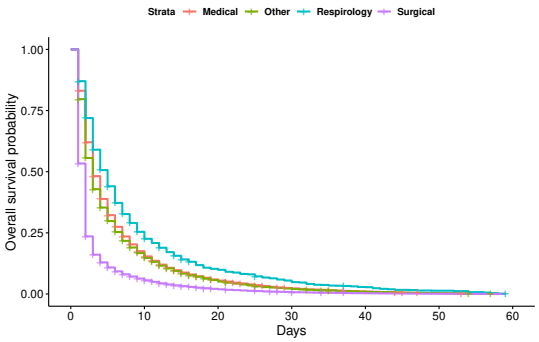
Fig 5.6(a) shows a histogram of the data overlaid with the density plot of the fitted distribution and Fig 5.6(b) contains the PP plot. Table 5.8 tabulates the criteria (Kolmogorov-Smirnov score (K-S), Cramer-von Mises score (C-M)), Anderson-Darling score (A-D), log-likelihood



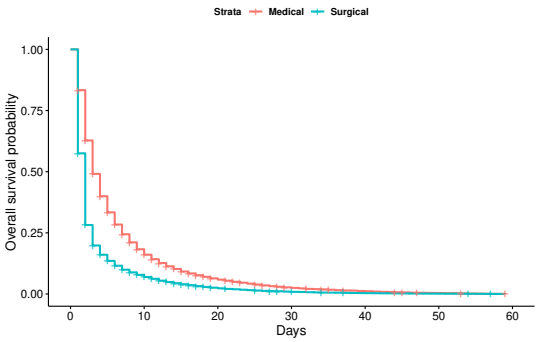
(a) Admission source subgroups



(b) Diagnosis subgroups.

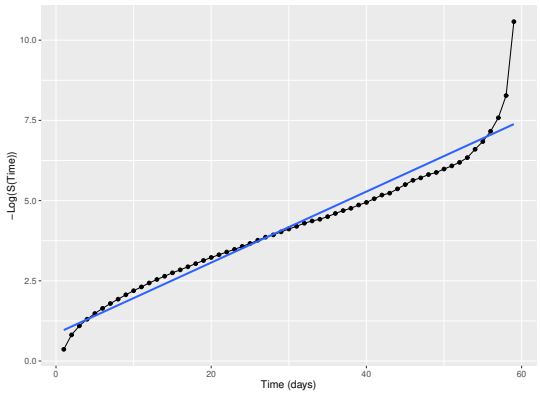


(c) Physician referral subgroups

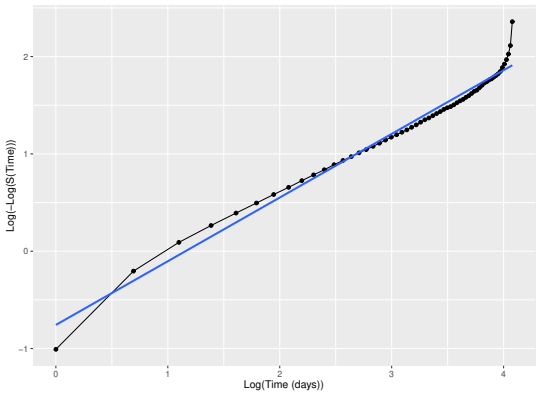


(d) Patient category subgroups.

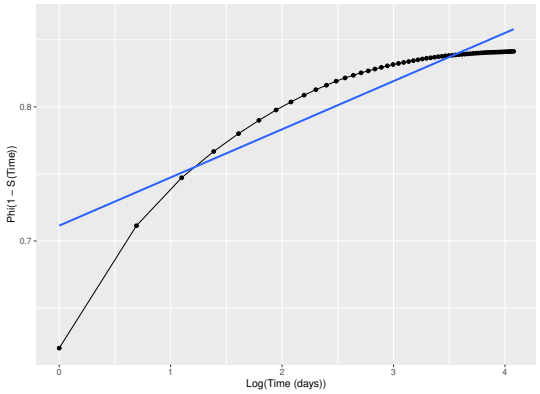
Figure 5.4: Kaplan-Meier survival curves



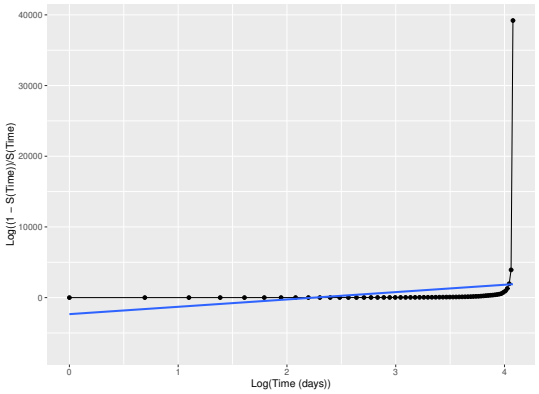
(a) $(-log(S(t))$ vs $t)$



(b) $(log(-log(S(t)))$ vs $log(t))$.



(c) $\Phi(1 - S(t))$ vs $log(t)$

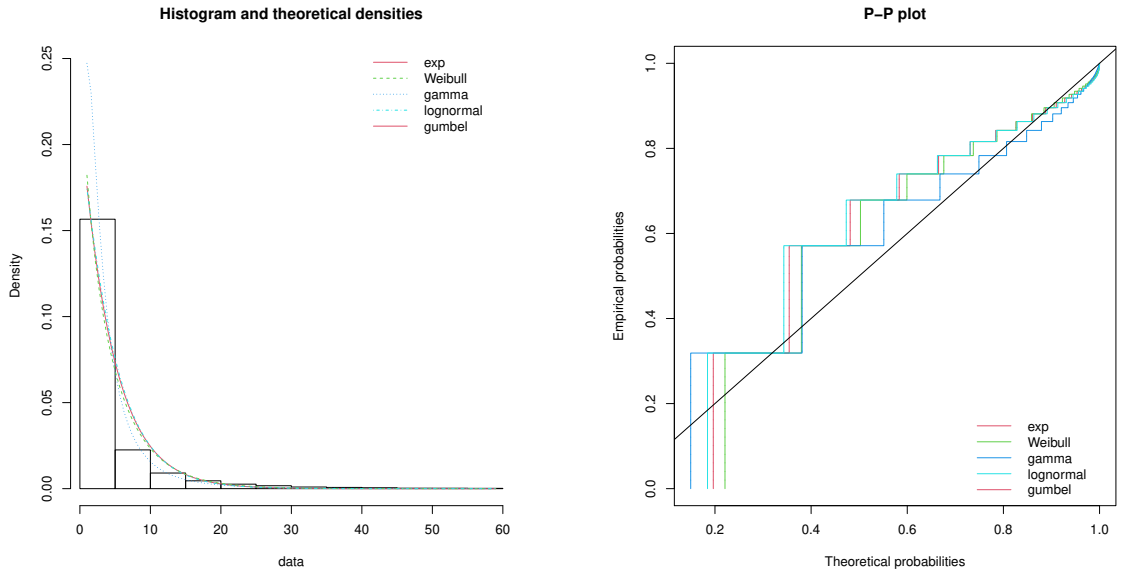


(d) $log(\frac{(1-S(t)}{S(t)})$ vs $log(t)$.

Figure 5.5: Graphical check of AFT assumption for (a) exponential, (b) Weibull, (c) log-logistic, and (d) log-normal distributions

(log-l), the AIC, and the BIC) of the fitted distributions. Based on the criteria, the log-normal probability distribution function was identified as the best distribution for the First-day ventilation time. The maximum likelihood estimates of the shape and scale parameters and their standard deviations are given as $\hat{\mu} = 0.98(0.004)$, and $\hat{\sigma} = 0.94(0.003)$ respectively. The actual form of the probability density function is of the form

$$f(t;\hat{\mu},\hat{\sigma}) = \frac{1}{t\hat{\sigma}\sqrt{2\pi}} exp\left(-\frac{1}{2\hat{\sigma}^2}(log(t)-\mu)^2\right), t > 0. \tag{5.1}$$



(a) Histogram with fitted pdfs.

(b) P-P plot of fitted distribution functions.

Figure 5.6: Estimation of the probability density function for the ventilation time.

Table 5.8: MLE Estimates of ventilation time of all First-day ventilated patients.

Dist	Estimate (sd)	K-S	C-M	A-D	log-l	AIC	BIC
Exp	$\hat{\lambda}$ =0.22 (0.00098)	0.22	521	2929	-124637	249275.3	249284
Weibull	$\hat{k}$ =0.94 0.003, $\hat{\lambda}$ =4.40 (0.02)	0.22	462	2710	-124409	248822	248840
Gamma	$\hat{\sigma}$ =1.06 (0.01), $\hat{\lambda}$= 0.23 (0.002)	0.23	555	3063	-124588	249179	249197
Lognorm	$\hat{\mu}$ = **0.98 (0.004)**, $\hat{\sigma}$ =**0.94 (0.003)**	**0.19**	**299**	**1917**	**-115538**	**231080**	**231098**
Gumbel	$\hat{\alpha}$ =2.46 (0.01), $\hat{\sigma}$ =2.78 (0.01)	0.26	739	4292	-137551	275107	275124

Dist (distribution), sd (standard deviation, KS (Kolmogorov-Smirnov), C-M (Cramer-von Mises), and A-D (Anderson-Darling)

5.3.4 Cox Proportional Hazard Model

In this section, we fitted the Cox-proportional hazard (PH) model and checked the proportional hazard assumption. The Cox PH model has no assumption about the distribution of the event time, however, it assumes that the hazard ratio is constant over time. This assumption was tested for each covariate and globally. The goodness of fit test of proportional hazards assumption tabulated in Table B.3 in the Appendix gives a significant p-value $< 2e - 16$. Therefore the null hypothesis that the proportionality assumption holds is rejected globally. It indicates a lack of proportionality for the hazard function. There is a significant deviation from the proportional hazards assumption for all the variables (p-value < 0.05). By inspecting Figures 5.3 and 5.4, the lines for male and female patients as well as the various age groups were not parallel, confirming that the proportional hazards assumption is not reasonable in this case of stratifying

the data. The PH model is not appropriate for our data and we therefore proceed to model the ventilation time using the AFT model.

5.3.5 Accelerated Failure Time Model

Model Selection

To model the ventilation time, we fit Exponential, Weibull, Log-logistic, and Lognormal AFT models. In each case, we fit the model to all the covariates without Basic monitoring on arrival. From Table 5.9, we assessed each model using Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and the Log-Likelihood from model selection. Here, we identified the Log-logistic model (log-likelihood = -73 400.5, AIC = 147 067, and BIC = 148 191) as a better fit, as it is the model with the smallest criteria. The fitted log-logistic AFT model on all covariates is tabulated in Table B.4 in the Appendix.

Table 5.9: Performance comparison of AFT models on the regional data

Model	$log-lik$	$log-lik^2$	χ^2	AIC	BIC
Exponential	-79951	-85627	11352	160165	161281
Weibull	-79604	-85476	11745	159474	160598
Log-Normal	-73852	-79452	11201	147970	149094
Log-logistics	**-73401**	-79744	**12687**	**147067**	**148191**

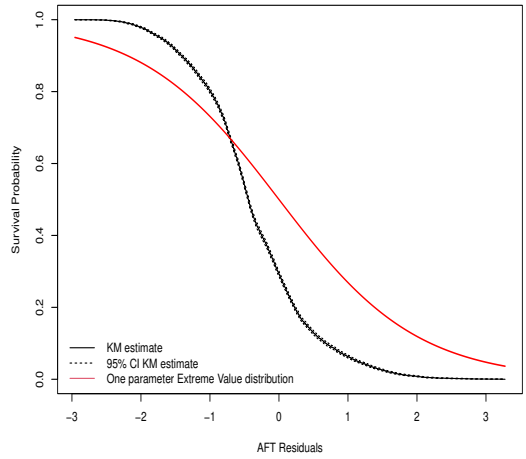
$log-lik$ (log-likelihood), $log-lik^2$ (log-likelihood with intercept only)

We assessed the goodness of fit of each models using the distribution of the Cox-Snell residuals. To do this, we compared the survival estimates of each parametric model with the KM estimates, by plotting the survival probability against the Cox-Snell residuals. We are looking for the survival function to closely follow the KM estimate. From Fig 5.7, the survival function for the Log-logistic model found in Fig 5.7(d) is superimposed with the KM curve, clearly showing that this model approximates the empirical survival better than the other models.

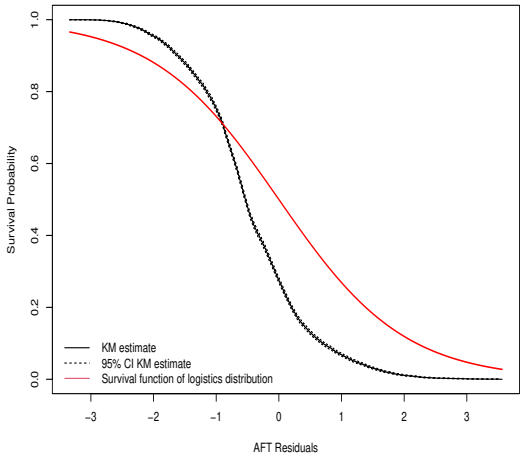
Variable Selection

We followed the variable selection approach outlined in the methods section. Using the backward selection procedure, patient category (p-value > 0.74), dialysis (p-value > 0.57), interventions outside (p-value > 0.10) and sex (p-value > 0.65) were eliminated, resulting in the final model presented in Table 5.10.

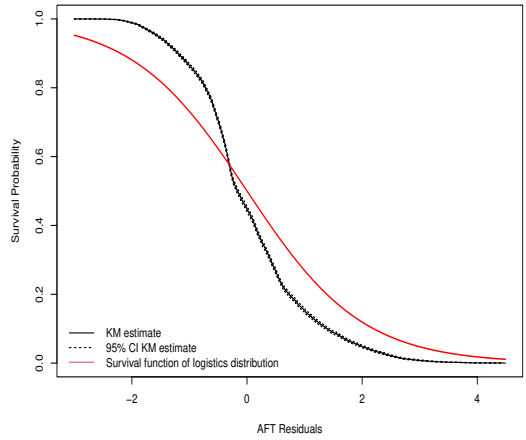
The analysis of the log-logistic AFT model revealed that the model containing the explanatory variables significantly improved the predictive ability of the model with the intercept only, as the likelihood ratio gave a p-value $< 2e - 16$. The overall effect of each of the retained



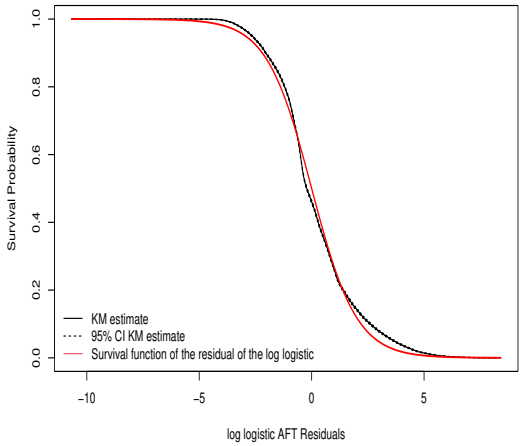
(a) Exponential AFT



(b) Weibull AFT



(c) log-normal AFT



(d) Loglogistics AFT

Figure 5.7: Residual survival plot to assess AFT models' goodness of fit.

covariates on survival time revealed that all had a significant independent effect on IMV time (all likelihood ratio tests resulted in a p-value $< 10e - 5$). Table B.5 tabulates the likelihood ratio test results of variable selection criteria for models fitted to the data using backward elimination.

Table 5.10: Log-logistic AFT model of the training set

	Covariate	$coef$	$L_{95\%}$	$U_{95\%}$	$e^{(coef)}$	$L_{95\%}$	$U_{95\%}$
shape		2.26	2.24	2.28	-	-	-
scale		1.12	0.98	1.29	-	-	-
ICU Site Code							
	3970	(reference)					
	3972	0.30	0.16	0.45	1.36	1.17	1.57
	3985	0.38	0.27	0.49	1.46	1.31	1.62
	3986	0.47	0.34	0.60	1.60	1.40	1.82
	3987	0.19	0.05	0.33	1.21	1.05	1.39
	3996	0.54	0.38	0.70	1.71	1.46	2.01
	4001	0.19	0.08	0.30	1.21	1.08	1.34
	4044	0.33	0.17	0.49	1.39	1.19	1.63
	4045	0.24	0.11	0.36	1.27	1.12	1.44
	4052	0.33	0.21	0.45	1.39	1.23	1.57
	4054	0.42	0.29	0.54	1.52	1.34	1.71
	4056	0.46	0.32	0.59	1.58	1.38	1.80
	4057	0.10	-0.13	0.32	1.10	0.88	1.38
	4063	0.02	-0.09	0.13	1.02	0.91	1.14
	4071	0.07	-0.25	0.40	1.07	0.78	1.49
	4073	0.29	0.17	0.42	1.34	1.18	1.52
	4076	0.58	0.36	0.79	1.78	1.44	2.20
	4079	0.42	0.30	0.53	1.52	1.35	1.70
	4085	0.39	0.27	0.51	1.47	1.31	1.66
	4089	0.43	0.16	0.69	1.53	1.18	2.00
	4090	0.26	0.15	0.37	1.29	1.16	1.44
	4093	-0.28	-0.62	0.07	0.76	0.54	1.07
	4097	0.39	0.24	0.54	1.48	1.27	1.72
	4103	0.17	0.01	0.33	1.19	1.01	1.39
	4107	0.47	0.34	0.60	1.60	1.40	1.82
	4108	0.70	0.50	0.90	2.02	1.65	2.46
	4109	0.25	0.09	0.41	1.28	1.09	1.51
	4110	0.10	-0.03	0.23	1.11	0.97	1.26
	4123	0.30	0.13	0.47	1.35	1.13	1.60
	4130	0.37	0.23	0.51	1.45	1.26	1.66

4131	0.02	-0.16	0.19	1.02	0.85	1.21
4138	0.62	0.49	0.76	1.87	1.63	2.13
4144	0.46	0.31	0.61	1.58	1.36	1.85
4168	0.21	0.02	0.40	1.23	1.02	1.50
4171	0.42	0.30	0.55	1.53	1.34	1.73
4180	0.16	0.05	0.27	1.17	1.05	1.31
4186	0.29	0.03	0.55	1.33	1.03	1.73
4192	0.50	0.36	0.64	1.65	1.43	1.91
4193	0.32	0.06	0.59	1.38	1.06	1.81
4197	0.06	-0.31	0.42	1.06	0.73	1.53
4199	-0.12	-0.31	0.06	0.88	0.74	1.06
4205	0.46	0.35	0.57	1.58	1.42	1.76
4209	0.14	0.01	0.28	1.15	1.01	1.32
4231	0.20	0.09	0.30	1.22	1.10	1.35
4233	0.26	0.14	0.39	1.30	1.15	1.48
4235	0.31	0.16	0.46	1.36	1.17	1.59
4238	0.34	0.19	0.50	1.41	1.21	1.64
4241	0.21	-0.05	0.48	1.24	0.95	1.62
4245	0.27	0.13	0.41	1.31	1.14	1.51
4260	0.25	0.04	0.46	1.29	1.04	1.59
4265	0.57	0.46	0.67	1.76	1.59	1.96
4266	0.54	0.42	0.67	1.72	1.52	1.95
4285	0.18	0.05	0.31	1.20	1.06	1.36
4303	0.48	0.37	0.59	1.61	1.44	1.80
4310	0.27	0.17	0.38	1.31	1.18	1.46
4311	0.25	0.14	0.36	1.28	1.14	1.43
4315	0.38	0.25	0.51	1.46	1.28	1.66
4414	0.36	0.21	0.51	1.43	1.24	1.66
4471	0.27	0.10	0.45	1.32	1.11	1.56
4774	0.35	0.22	0.47	1.41	1.25	1.60
4799	0.46	0.33	0.59	1.59	1.39	1.81
4832	0.32	0.20	0.43	1.37	1.23	1.54
4837	0.43	0.29	0.57	1.53	1.33	1.76
4839	0.72	0.57	0.87	2.05	1.77	2.38
4841	0.44	0.27	0.60	1.55	1.31	1.83
4845	0.45	0.30	0.61	1.57	1.34	1.84

Admission Source						
Down stream units	(reference)					
ED	-0.13	-0.17	-0.09	0.88	0.84	0.92
Home	0.13	-0.01	0.27	1.14	0.99	1.31
Hospital	0.11	0.06	0.16	1.12	1.06	1.17
OR	-0.31	-0.36	-0.27	0.73	0.70	0.77
Other	0.16	0.05	0.27	1.17	1.05	1.31
Ward	-0.13	-0.18	-0.09	0.88	0.84	0.92
Diagnosis						
Cardiovasclar	(reference)					
Gastrointestinal	0.28	0.24	0.32	1.33	1.27	1.38
Neurological	0.31	0.28	0.35	1.37	1.32	1.42
Other	0.40	0.37	0.43	1.49	1.45	1.53
Trauma	0.53	0.47	0.58	1.70	1.61	1.79
Is Scheduled ICU Admission						
No	(reference)					
Yes	-0.22	-0.27	-0.17	0.80	0.76	0.84
Is Scheduled Surgery						
No	(reference)					
Yes	-0.16	-0.20	-0.11	0.86	0.81	0.90
Referral physician specialist						
Medical	(reference)					
Other	-0.01	-0.04	0.02	0.99	0.96	1.02
Respirology	0.15	0.08	0.21	1.16	1.08	1.23
Surgical	-0.08	-0.11	-0.04	0.93	0.89	0.96
Central venous line						
No	(reference)					
Yes	0.17	0.14	0.19	1.18	1.15	1.21
Arterial line						
No	(reference)					
Yes	0.19	0.16	0.21	1.21	1.18	1.24
Intra-cranial pressure monitor						
No	(reference)					
Yes	0.56	0.48	0.63	1.74	1.62	1.88
Extracorporeal membrane oxygen						
No	(reference)					

Yes	0.66	0.48	0.84	1.94	1.62	2.32
Intra aortic balloon pump						
No	(reference)					
Yes	0.34	0.26	0.41	1.40	1.30	1.51
Other Interventions Within this Unit						
No	(reference)					
Yes	0.07	0.05	0.09	1.07	1.05	1.10
Age group						
18 - 39	(reference)					
40 - 79	0.12	0.09	0.15	1.12	1.09	1.16
≥ 80	0.05	0.01	0.09	1.05	1.01	1.09
Pre ICU LOS						
≤ 1 day	(reference)					
2 – 71 day	0.04	0.01	0.06	1.04	1.01	1.07
≥ 7 day	0.15	0.12	0.18	1.16	1.12	1.19
MODS						
1	(reference)					
1-4	0.10	0.07	0.13	1.11	1.07	1.14
5 - 8	0.22	0.19	0.25	1.25	1.21	1.29
9 - 12	0.27	0.23	0.31	1.31	1.26	1.36
≥ 13	0.17	0.14	0.21	1.19	1.15	1.23
NEMS						
0-22	(reference)					
23 - 29	-0.01	-0.09	0.06	0.99	0.92	1.07
≥ 30	0.04	-0.03	0.12	1.04	0.97	1.13

Model predictive performance

To assess the predictive performance of the model, we used a test dataset, and predicted the ventilation time using the log-logistic model. We computed and compared the residuals of the test data with those of the training data. Table 5.11 presents the comparison prediction performance on the training and testing data. The predicted average ventilation duration was 2.96 days for the test data compared with 2.94 days for the training data. There is an insignificant loss of performance from the prediction of the training data to the test data in the quantiles (1.60, 2.75, 3.81 days to 1.59, 2.77, 3.85 days), in the MSE (from 46.81 to 60.21), in the MAE (from 2.95 to 3.00), in the PBIAS (from 0.36 to 0.35), and the NES (from 1.090 to 1.380).

Table 5.11: Table comparing prediction statistics from the test and training dataset.

Statistics	N (%)	Mean	SD	$Q_{0.25}$	$Q_{0.5}$	$Q_{0.75}$	MSE	MAE	PBIAS	NES
Training	34 626 (70%)	2.94	2.83	1.59	2.77	3.85	46.81	2.95	0.36	1.09
Test	14 840 (30%)	2.96	4.55	1.60	2.75	3.81	60.21	3.00	0.35	1.38

Fig 5.8 shows the survival functions of the training (in blue) and testing (in black) residuals superimposed on the baseline KM survival. We observe that the survival curves of the training and testing data are similar to the baseline KM survival. The performance of the training model is revealed on the test data with an insignificant difference and a narrow gap. This indicates that the model has a good ability to predict patients' ventilation duration using First-day observations.

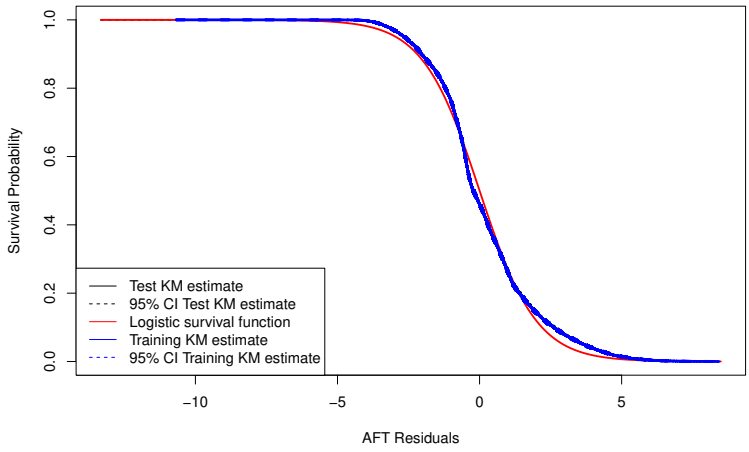
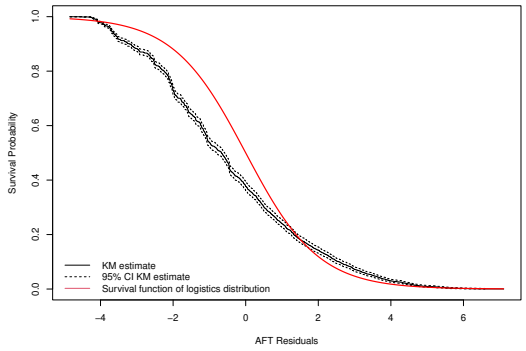


Figure 5.8: Residuals' survival curves: Test (Black), Training (Blue), and Log logistic survival (Red).

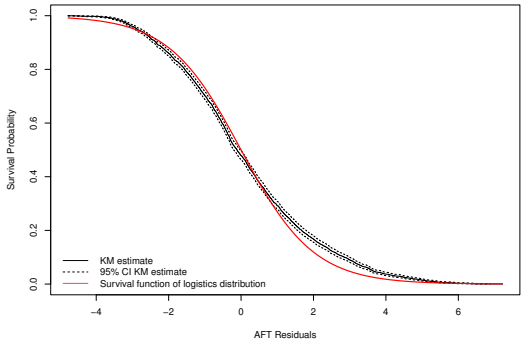
Model Validation

To validate the model's prediction, we used a new dataset from the London Health Sciences Center (LHSC), with data gathered from January 2019 to May 2021. The first step was to calibrate the data. Calibration checks the agreement between observed outcomes and predicted ones. To assess the need for model calibration, a simple linear regression and scatter plot of the observed versus the predicted outcomes is performed. Perfect predictions yield a slope of 1 with an intercept of 0, which is the line of best fit that should divide the first quadrant into two equal parts. A failure could inform a need for a model that considers shrinkage. The prediction performance compares the estimated Kaplan-Meyer survival curve of the predicted residual to

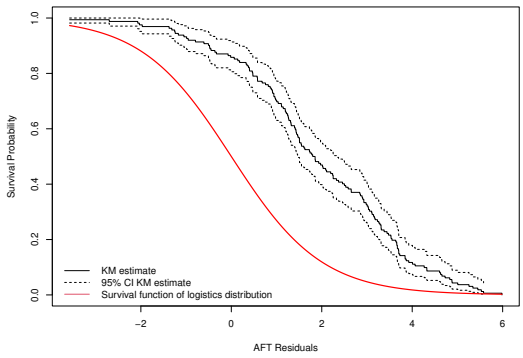
the expected empirical survival curve. This was done on three subgroups: COVID-19 patients, Non-COVID-19 patients, and the whole LHSC data. Fig 5.9(a) is the survival of the residual obtained from the LHSC data from January 2020 to May 2021, including all patients that received IMV on arrival.



(a) LHSC 2019 to 2021 all patients' data



(b) LHSC 2019 to 2021 non-COVID-19 patients' data



(c) LHSC 2019 to 2021 COVID-19 patients' data

Figure 5.9: Residual survival plot using the LHSC data: predicted survival function (black), logo-logistics survival distribution (Red)..

In Fig 5.9(a), the survival curve of the AFT model poorly approximates the expected empirical survival, as there appears to be a major departure between the two. This could be due to a medical condition that has not been present in the previous time interval, in particular, COVID-19. To confirm our assumption about the effect of COVID, we divided the data into two; patients with COVID-19 and patients without COVID-19.

Interestingly, in Fig 5.9(b), we observed a survival function that closely follows the KM estimate, suggesting a good fit for the non COVID patients. However, the survival plot of the residual from the prediction of the COVID patients in Fig 5.9(c) performs poorly. This is

confirmed in Table 5.12, where we numerically compare the model's predictive performance on the three sub-datasets (Covid patients data, Non-COVID patients, and Mixed data).

Table 5.12: Model validation performance on three sub-data set.

Statistics	n	mean	sd	$Q_{0.25}$	$Q_{0.5}$	$Q_{0.75}$	MSE	MAE	PBIAS	NES
Covid Patients	162	4.7	1.11	3.96	4.86	5.28	201.9	10.16	0.67	1.84
Non-Covid patients	3594	3.52	1.19	2.74	3.39	4.08	49.42	3.60	0.37	1.05
All patients	3756	3.57	1.21	2.77	3.43	4.15	56.00	3.89	0.40	1.06

5.4 Discussion

Accurate prediction of ICU resources helps guide therapeutic decision-making, resource allocation, and patient flow management. The number of days on IMV is a major concern of critical care management and costs [41, 21]. However, IMV duration prediction models in the literature were mainly based on the conventional multivariate regression model and the logistics regression, thus, do not incorporate censored observations and are based on classifying patients ventilation time to either a short or long duration [161, 46, 1, 56, 162, 10]. We performed comprehensive survival analysis to predict and determine predictors of ventilation time using the CCIS Ontario dataset. Information obtained at arrival is an important piece in forecasting patient ventilator days.

The Log rank test on the KM curves of the covariates available on the first-day of ventilation show that only basic monitoring was insignificant (p-value = 0.80). This is likely because a very small number did not receive the basic monitoring treatment and therefore do not have the power to rule out a real difference and avoid a type II error (false negative).

The covariate's effect in the AFT model is to accelerate or decelerate the event time, which in this case is the invasive ventilation time. The results of the association are shown in Table 5.10. A convenient way to understand the coefficients better is through interpretation of the time ratio (TR), also called the acceleration factor. The TR for a given covariate is the (natural) exponent of the estimated parameter coefficient (i.e, $exp(\beta)$). A positive coefficient corresponds to a TR greater than 1, while a negative coefficient corresponds to a time ratio less than 1. Correspondingly, a TR greater than one implies that the covariate increases the time to event. An acceleration factor equal to 1 corresponds to no effect on the time to event.

A positive coefficient was observed for the majority of the ICUs (SiteCode) implying that for the majority of the ICUs, the time to event is higher than average. Relatively, the ICU a patient visited was a significant predictor of the patient's ventilation time. Different ICUs had differing TR. Compared to the ICU with code 3970 (use as reference), ICUs with site codes

4057, 4063, 4071, 4093, 4110, 4131, 4197, 4199, and 4241 (9 out of 65) had an insignificant coefficient. That is, their acceleration factor's confidence interval covers 1 and is not significantly different from that of 3970. Moreover, the 56 remaining ICUs had significantly higher TR. Attending those ICUs increases the time a patient spends using a ventilator. This high difference in ventilation time in the ICU may be attributed to the practices and the location of the various ICUs as outlined in Burns et al. [28]. The implication of this to the ICU management is that the ICU managers should learn best practices from the ICUs that seem to have a lower ventilation time.

Comparing patients admitted to the ICU from the downward stream (SDU and Level 3) to those patients admitted from the house (TR = 1.14, CI = (0.99,1.31)), there was no significant difference. However, patients from the ED (TR = 0.88, CI = (0.84 0.92)), OR (TR = 0.73, CI = (0.70, 0.77)), and ward (TR = 0.88, CI =(0.84, 0.92)) had lower odds of longer ventilation time, while patients admitted from hospitals (TR = 1.12, CI = (1.06, 1.17)) and other sources (TR = 1.17, CI = (1.05,1.31)) had higher odds of longer ventilation time within the ICU. Admitted with cardiovascular/cardiac/vascular diagnoses had a higher experience of the event as compared to other etiologies. Gastrointestinal (TR = 1.33 , CI = (1.27 1.38)), Neurological (TR = 1.37 , CI = (1.32 1.42)), Trauma (TR = 1.70 , CI = (1.61 1.79)), and Other diagnosis (TR = 1.49 , CI = (1.45 1.53)) had higher odds to say on the ventilator compared to cardiovascular patients. This could be explained by the founding of Kao et al. [79] using the same data where cardiovascular diagnoses patients had higher mortality as compared to other etiologies [128]. Patients' admission diagnosis types were also significant factors affecting ventilation time. Surgical patients (TR = 0.770, CI = (0.702, 0.843)) had higher odds to leave the ventilator earlier than medical patients.

Scheduled ICU admission (TR = 0.827, CI = (0.784, 0.873)) has a decelerating effect on the ventilation time. This implies that scheduled admission was a significant predictor of patients' ventilation time. Scheduled patients have 17.3% higher odd for shorter ventilation time compared to non-scheduled patients. Scheduled patients have 17.3% higher odds for shorter ventilation times compared to non-scheduled patients. This may be attributed to the fact that generally, scheduled patients are taking elective procedures that support their faster and safer transit through the ventilator, and are thus less likely to remain connected to the ventilator compared to non-scheduled patients. Also, compared to non-scheduled surgery patients, patients with scheduled surgery (TR = 0.823, CI = (0.783, 0.864)) have a 17.7% higher probability of a shorter ventilation time. Referring Physician services such as Cardiology (TR = 1.100, CI = (0.989, 1.220)), Ophthalmology (AF = 1.39, CI = (0.726, 2.64)), and Psychiatry (AF = 1.01, CI = (0.639,1.610)) were non-significant. Referrals from these physicians had no ventilation time effect. They act as the average baseline. However, the rest of the referring physicians'

services had a significantly higher TR with none that were a significantly lower. For example, patients from Trauma (TR = 2.28, CI = (2.12, 2.45)) and Paediatric (TR = 10.100, CI = (2,910, 34.900)) have the longest time on a ventilator. Unequivocally, treatments received on arrival had a significant effect on ventilation time. The likelihood of longer ventilation is much higher for patients who received central venous line (CVL) (TR = 1.18, CI = (1.15, 1.21)), arterial line (AL) (TR = 1.21, CI = (1.18, 1.24)), intracranial pressure monitor (IPM) (TR = 1.74, CI =(1.62, 1.88)), extracorporeal membrane oxygen (EMO) (TR = 1.94, CI = (1.62, 2.32)), Intra Aortic Balloon Pump (IABP) (TR = 1.40, CI = (1.30, 1.51)), and other intervention within the ICU (OIWU) (TR = 1.07, CI =(1.05, 1.10)). These treatments had a decelerating effect on the event. Patients who received these interventions in ICU on arrival were connected for a longer time compared to those who did not. The odds of longer ventilation time are much higher for adults and seniors compared to youths. Patients aged 40–79 (TR = 1.12, CI =(1.09, 1.16)) and those $\geq$ 80 years old (TR = 1.05, CI = (1.01, 1.09)) spent longer time on the ventilator compared to those of patients age $\leq$ 39. This confirms the results by Piotto et al. [126] and Lei et al. [93] who showed that advanced age (more than 60 years) was a significant predictor for IMV. However, we found that patients' sex is not a significant predictor. Longer pre-ICU LOS is associated with a longer ventilation time. Specifically, patients who spend more than 1 day but less than 7 days in the hospital post ICU (TR = 1.04, CI = (1.01, 1.07)), and those who spend more than a week post ICU admission (TR = 1.16, CI =(1.12, 1.19)) are more likely to experience the event later than compared to those who spent less than 1-day post ICU admission. Higher First-day scores in MODS corresponded to increasing time to event (MODS = (1-4), TR = 1.11, CI = (1.07 1.14)) (MODS = (5 - 8), TR = 1.25, CI = (1.21 1.29)), (MODS = (9 - 13), TR = 1.31 , CI = (1.26 1.36)), and (MODS $\geq$ 13, TR = 1.19, CI = (1.15 1.23)). Patients with high MODS scores upon arrival are at a higher risk of longer ventilation connection times compared to those with low scores. Patients with high MODS score on arrival are at high risk of longer connection to the ventilator compared to those with low scores. The First-day NEMS score however had a weak association with the ventilation time (NEMS = (0-22), TR = 0.99 , CI = (0.92 1.07)), (NEMS $\geq$ 30, TR = 1.04 , CI = (0.97 1.13)). In the review by Ghauri et al. [60], results based on logistic regression models indicated no significant effects for the Acute Physiology and Chronic Health Evaluation (APACHE II) on the prediction IMV time of ICU patients. Although one would expect NEMS to be a significant predictor because of the weight of the respirator component, the results show the opposite, where it is not highly associated. However, the presence of the covariate in the model significantly increases the model's prediction performance, as seen in the variable selection.

5.5 Conclusion

This study proposed survival analysis to investigate the duration of invasive mechanical ventilation in ICU patients. Both parametric and non-parametric methods have been explored. Nonetheless, based on the AIC and BIC criteria, the log-logistic AFT model has been retained as the best to predict ventilation time for each ICU patient. ICU site, admission source, admission diagnosis, scheduled admission, scheduled surgery, referral physicians, central venous line treatment, arterial line treatment, intracranial pressure monitor treatment, extra-corporeal membrane oxygen treatment, intra-aortic balloon pump treatment, other interventions, age group, pre ICU LOS, and MODS score were significant predictors of the ICU ventilation time. Even though the data used for the modelling is five years old, it performed well on current non-COVID patients. The prediction performance of the proposed model showed that it can be used to predict future ventilation time duration and provide insight into the predictors of ventilation time.

Our study differs from the previous studies in several ways. First, unlike the studies of [161], [46] , [1], [56], [162], and [10] our study focused entirely on predicting the continuous-time of ventilation time, where we used information gathered on the first day to predict the ventilation time. Additionally, our sample size was larger and we used external validation instead of bootstraps. This gives us more power to detect prediction performances.

Our study has several limitations. Our analysis includes only patients who entered the ICU for one and a half years. Although the patient characteristics of this subgroup are not dissimilar to that of the patients in the validation set, we note that the differing study period could have significant, unrecognizable differences due to the appearance of COVID-19. The heterogeneity of the ICU site may affect our results. Since the models perform differently for different ICUs, ICU site could be considered as a random effect. In that case, each ICU will have its own model. Nonetheless, with the current model, we can compare the acceleration functions of the various ICUs in the province. The survival model used in this research could benefit from the automatic variable selection tools of penalized AFT models. Additionally, other machine learning methodologies, such as the random forest and Support vector machines (SVM) could be used. In addition, nonlinear tree based machine learning algorithms as implemented in libraries such as XGBoost, scikit-learn, LightGBM, and CatBoost with more accuracy estimation could be used. However, the state-of-the-art implementations of such methods for the AFT models are limited.

Chapter 6

Conclusion and Future Work

6.1 Main Contributions

This thesis has studied patients' flow decisions, queuing game interaction, and duration of ventilation in the ICU-SDU system via three research projects. The contributions of each project are outlined as follows.

1. In the first project, we have studied patients flow in the ICU-SDU system. A Markov decision model was used to model patient flow decisions by managing the ICU and the SDU with or without premature ICU step-downs. Using the optimal solution of the model parametrized on real data obtained from LHSC, we simulated the system and have discovered counter-intuitive results. Some of the significant contributions of this study are as follows:

    - The last bed problem in the ICU-SDU flow has been formulated as a Markov decision problem (MDP) to model the patient admission, step-down, and discharge decisions.
    - As a last bed problem, we proposed a reduced state space in the formulation of the MDP model to avoid the problem of high dimensionality, which gave the possibility to zoom in to congestion times and study the situation without blurring the actions in non-congestion times.
    - We implemented the determination of a set of actual actions to be taken when the system is congested instead of the determination of a certain threshold.
    - We optimized the MDP model under two sets of actions in congestion: one rejecting a new patient in need of critical care by default, and the other prematurely

discharging a current critical care patient to the SDU in other to admit the new arriving patient.

- We solved the Markov decision process using linear programming approximation and showed that it is simple, effective, and computationally inexpensive for modelling sequential decision-making processes under congestion.
- We simulated the ICU-SDU system under two sets of actions using the optimal actions obtained from the MDP models.
- Our results counter-intuitively contradict conventional practice. Premature step-down of critical patients done under congestion was found to be a less efficient patient flow policy than rejecting arriving critical patients.

2. The second project is a queuing game model between two servers (ICU and SDU) in tandem without a buffer in between them. In this chapter, our main contributions are as follows.

   - We introduced a queuing game model between two servers in tandem where the first server (ICU) has a queue and the second server (SDU) has no queue.
   - We modelled each unit's objective function as a decentralized integrated system, and as a central planner.
   - We built queueing game models where the servers decide on the treatment efforts in both competition and cooperative games instead of patients choosing to join the queue or renege.
   - We determined the feasibility conditions for each game structure.
   - We found closed form solution for the decisions made by all players in all game structures.
   - We determined the conditions in which the downstream server improves the payoff of the system.
   - Results showed that the best performing structure depends on the KPI metric used.

3. The third project focused on modelling ventilation time in the ICU using survival models. In this project, our main contributions are as follows.

   - We used an extensive data set obtained from the Critical Care Information System of Ontario to identify the distributions that best characterize the ventilation time duration of patients connected to the ventilator on arrival at the ICU.

- We built parametric and non-parametric survival models to infer among several covariates, those that significantly characterize the time to disconnection of invasive mechanical ventilators of patients that started using ventilation at arrival.
- Log-logistic AFT model was found and used as the best model to determine significant predictors, their acceleration function, and the ventilation time of first day ICU patients ventilation time.
- It is important to consider ICU site disparities to improve the analysis of ventilation time and make capacity prediction and inform better decisions.

6.2 Limitations

The findings of this thesis must be seen in light of some limitations.

1. The limitations of the first project are as follows:

   - The first limitation is the assumption regarding the ICU-SDU system capacity and organization. If ICU and SDU capacities and length of stays are equal, our results and conclusions may change. Even though we did an extensive sensitivity analysis, changes to the capacities were not attempted due to the nature of the last bed problem.
   - Further, in some hospitals, the physical structure and the patient flow policy of the ICU and the SDU may differ. Our model was based on LHSC's parameters, which may pose difficulties to ICU-SDU systems at other hospitals.
   - The third limitation concerns the effects of the assumed value for the reward and the cost. Even though cost and reward are robust to small changes, we postulate that they should vary differently in different systems. From a managerial perspective, the value of life analysis (willingness to pay) will be required to better estimate the reward and cost of every action. The value of life is an economic value that can quantify the benefit or loss of avoiding or falling into a fatality.
   - The fourth limitation concerns the data used and the various model parameters estimated, which may be subject to biases and confounding factors. Different ICU-SDU systems with different data may produce different results from a regional distributional perspective.

2. The second project is subject to the following limitations:

- The model we used considered a one-server system at both ICU and SDU and an exponential distribution for inter-arrival times and service times. In practice, ICUs and SDUs have multiple beds, and the number of beds at the ICU differs from those at the SDU.
- The SDU can be used as a buffer between the ICU and the Ward in some hospitals. Therefore, the interaction between SDU and Ward was not considered in the system.
- Finally, the linear payoff used may be too simplistic, and more research is needed to determine the actual structure of the utility of the servers in the system.

3. In the third project,

- The study cohort is the primary limitation to the generalization of the obtained results.
- In addition, it is important to highlight that our results present the mean estimated survival curve for the targeted population and not the individual patient's survival curves.
- Furthermore, the relative risks estimated in the model must be cautiously interpreted as they are based on retrospective records and, therefore, are subject to confounding factors.
- The recent COVID-19 pandemic have changed the nature of mechanical ventilation use in Ontario's ICUs. Therefore, further investigation may be necessary.

6.3 Future Work

We highlight below some extensions of our work for future research.

1. Chapter 3 we proposed an MDP model with a Monte Carlos simulation to optimize patient flow in the ICU.

- One obvious extension is the definition of the objective function at the individual level, instead of the overall health service benefit of the system.
- In addition, further dividing patient's acuity level (e.g., low, medium, high NEMS) may provide further insight so long as it proves to be mathematically tractable.

- This MDP could be reformulated as a re-enforcement learning problem where the algorithm will decide on the actions instead of the user defining the action space of the system.
- Finally, alternative measures of utility may be incorporated into the objective function of the MDP to include other measurement of ICU performance, especially as it concerns the last ICU bed problem.

2. In Chapter 4, we proposed a queuing game between the ICU and the SDU assuming an exponential arrival, an exponential service, and one server per station.

   - From the queuing game perspective, we could consider extending of the queuing game between the ICU and the SDU model to study pooling effects, generalized arrival and service distributions.
   - We could explore the equilibrium state properties of competition and cooperation between more than one downstream unit, eg, ICU-SD-Ward patient flow. The Ward may function as an alternative level of care for the SDU and they may compete for ICU patients.
   - We plan to study service time demand dependence by relaxing the assumption of the exogenous demand.
   - Finally, we envision a model where capacity ad or LOS becomes decisions for each of the units (M/M/C-M/K system or G/G/C-G/K system).

3. The fifth chapter focuses on modelling ventilation duration in the ICU using survival analysis. We can foresee the following extensions.

   - Further studying ventilation time duration prediction by applying penalized ridge, lasso, elastic net regression methods and comparing them to deep learning tools.
   - In addition, we can investigate if more accurate predictions could be obtained by ensemble models as well as nonlinear tree based machine learning algorithms as implemented in libraries such as SuperLEarner, XGBoost, scikit-learn, LightGBM, and CatBoost.
   - We could also combine the logistic regression model with the survival model as a Bayesian model to update predictions of consecutive days of ventilation time.
   - Finally, instead on focusing on the individual ventilation time prediction, we could attempt to predict mechanical ventilation of the current cohort of patients, ie, the utilization of the resources themselves.

Bibliography

[1] Ahmad Abujaber, Adam Fadlalla, Diala Gammoh, Husham Abdelrahman, Monira Mollazehi, and Ayman El-Menyar. Using trauma registry Data to predict prolonged mechanical ventilation in patients with traumatic brain injury: Machine learning approach. *PLoS one*, 15(7):e0235231, 2020.

[2] Daniel Adelman. Dynamic bid prices in revenue Management. *Operations Research*, 55(4):647–661, 2007.

[3] Neill KJ Adhikari, Robert A Fowler, Satish Bhagwanjee, and Gordon D Rubenfeld. Critical Care and the global burden of Critical illness in adults. *The Lancet*, 376(9749): 1339–1346, 2010.

[4] Per K Andersen, Ornulf Borgan, Richard D Gill, and Niels Keiding. *Statistical Models based on Counting Processes*. Springer Science & Business Media, 2012.

[5] David Anderson, Carter Price, Bruce Golden, Wolfgang Jank, and Edward Wasil. Examining the discharge practices of surgeons at a large Medical center. *Health Care Management Science*, 14(4):338–347, 2011.

[6] Shoshana Anily and Moshe Haviv. Cooperation in service systems. *Operations Research*, 58(3):660–673, 2010.

[7] Mor Armony, Carri W Chan, and Bo Zhu. Critical Care in hospitals: When to introduce a step down Unit. *Product Operation Management*, 2013.

[8] Mor Armony, Carri W Chan, and Bo Zhu. Critical Care capacity Management: Understanding the role of a step down Unit. *Production and Operations Management*, 27(5): 859–883, 2018.

[9] Kenneth J Arrow. Uncertainty and the welfare economics of Medical Care. *World Health Organization. Bulletin of the World Health Organization*, 82(2):141, 2004.

[10] MT Aung, D Garner, M Pacquola, S Rosenblum, J McClure, H Cleland, and DV Pilcher. The use of a simple three-level bronchoscopic assessment of inhalation injury to predict in-hospital mortality and duration of mechanical ventilation in patients with burns. *Anaesthesia and Intensive Care*, 46(1):67–73, 2018.

[11] Cristina Azcarate, Laida Esparza, and Fermin Mallor. The problem of the last bed: Contextualization and a new simulation framework for analyzing physician decisions. *Omega*, 96:102120, 2020.

[12] Jie Bai, Andreas Fügener, Jan Schoenfelder, and Jens O Brunner. Operations Research in ICU Management: a literature review. *Health Care Management Science*, 21(1): 1–24, 2018.

[13] Ramji Balakrishnan and Naomi S Soderstrom. The cost of system congestion: Evidence from the HealthCare sector. *Journal of Management Accounting Research*, 12(1):97–114, 2000.

[14] Bacchus Barua, Nadeem Esmail, and Taylor Jackson. *The effect of wait times on mortality in Canada*. Fraser Institute Vancouver, 2014.

[15] Sharo Begley. With ventilators running out, doctors say the machines are overused for Covid-19. *STAT*, 8, 2020.

[16] Jeremy R Beitler, Aaron M Mittel, Richard Kallet, Robert Kacmarek, Dean Hess, Richard Branson, Murray Olson, Ivan Garcia, Barbara Powell, and David S Wang. Ventilator sharing during an acute shortage caused by the COVID-19 pandemic. *American Journal of Respiratory and Critical Care Medicine*, 202(4):600–604, 2020.

[17] Richard Bellman and Stuart Dreyfus. Dynamic programming and the reliability of multicomponent devices. *Operations Research*, 6(2):200–206, 1958.

[18] Dan Bendel and Moshe Haviv. Cooperation and sharing costs in a tandem Queueing network. *European Journal of Operational Research*, 271(3):926–933, 2018.

[19] Dimitri P Bertsekas and John N Tsitsiklis. Neuro-dynamic programming: an overview. In *Proceedings of 1995 34th IEEE Conference on Decision and Control*, volume 1, pages 560–564. IEEE, 1995.

[20] Dimitri P Bertsekas, Dimitri P Bertsekas, Dimitri P Bertsekas, and Dimitri P Bertsekas. *Dynamic Programming and Optimal Control*, volume I. Athena Scientific Belmont, MA, 1995.

[21] Thomas Bice and Shannon S Carson. Acute Respiratory Distress Syndrome: Cost (Early and Long-Term). In *Seminars in Respiratory and Critical Care Medicine*, volume 40, pages 137–144. Thieme Medical Publishers, 2019.

[22] Stephen Boyd, Stephen P Boyd, and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[23] Stephen P Bradley, Arnoldo C Hax, and Thomas L Magnanti. *Applied Mathematical Programming*. Addison-Wesley, 1977.

[24] Kurt R Brekke, Luigi Siciliani, and Odd Rune Straume. Competition and waiting times in hospital markets. *Journal of Public Economics*, 92(7):1607–1628, 2008.

[25] Norman Breslow. Covariance analysis of censored Survival Data. *Biometrics*, pages 89–99, 1974.

[26] Norman Breslow and John Crowley. A large sample study of the life table and product limit estimates under random censorship. *The Annals of Statistics*, pages 437–453, 1974.

[27] James R Broyles, Jeffery K Cochran, and Douglas C Montgomery. A Markov decision process to dynamically match hospital inpatient staffing to demand. *IIE Transactions on HealthCare Systems Engineering*, 1(2):116–130, 2011.

[28] Karen EA Burns, Leena Rizvi, Deborah J Cook, Gerald Lebovic, Peter Dodek, Jesús Villar, Arthur S Slutsky, Andrew Jones, Farhad N Kapadia, and David J Gattas. Ventilator Weaning and Discontinuation Practices for Critically Ill Patients. *Journal of American Medical Association*, 325(12):1173–1184, 2021.

[29] Nancy Cady, Mark Mattes, and Suzanne Burton. Reducing Intensive Care Unit length of stay. A stepdown Unit for first-day heart surgery patients. *The Journal of Nursing Administration*, 25(12):29–35, 1995.

[30] Lucienne TQ Cardoso, Cintia MC Grion, Tiemi Matsuo, Elza HT Anami, Ivanil AM Kauss, Ludmila Seko, and Ana M Bonametti. Impact of delayed Admission to Intensive Care Units on mortality of Critically ill patients: a cohort study. *Critical Care*, 15(1): 1–8, 2011.

[31] Francisco Javier Carmona-Monge, Gloria Mª Rollán Rodríguez, Cristina Quirós Herranz, Sonia García Gómez, and Dolores Marín-Morales. Evaluation of the Nursing

workload through the Nine Equivalents for Nursing Manpower Use Scale and the Nursing Activities Score: a prospective correlation study. *Intensive and Critical Care Nursing*, 29(4):228–233, 2013.

[32] Donald B Chalfin, Stephen Trzeciak, Antonios Likourezos, Brigitte M Baumann, and R Phillip Dellinger. 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.

[33] Chester G Chambers, Maqbool Dada, Shereef Elnahal, Stephanie Terezakis, Theodore DeWeese, Joseph Herman, and Kayode A Williams. Changes to physician processing times in response to clinic congestion and patient punctuality: a retrospective study. *BMJ*, 6(10), 2016.

[34] Carri W Chan, Vivek F Farias, Nicholas Bambos, and Gabriel J Escobar. Optimizing Intensive Care Unit discharge decisions with patient readmissions. *Operations Research*, 60(6):1323–1341, 2012.

[35] Carri W Chan, Linda V Green, Lijian Lu, and Gabriel Escobar. The role of a step-down Unit in improving patient outcomes. Technical report, Citeseer, 2014.

[36] Wuhua Chen, Zhe George Zhang, and Zhongsheng Hua. Analysis of price competition in two-tier service systems. *Journal of the Operational Research Society*, 70(11):1938–1950, 2019.

[37] Vincent Chin, Noelle I Samia, Roman Marchant, Ori Rosen, John PA Ioannidis, Martin A Tanner, and Sally Cripps. A case study in Model failure? COVID-19 daily deaths and ICU bed utilisation predictions in New York state. *European Journal of Epidemiology*, 35(8):733–742, 2020.

[38] Jeffery K Cochran and Aseem Bharti. Stochastic bed balancing of an obstetrics hospital. *Health Care Management Science*, 9(1):31–45, 2006.

[39] David Collett. *Modelling Survival Data in Medical Research*. CRC press, 2015.

[40] David R Cox. Regression Models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2):187–202, 1972.

[41] Joseph F Dasta, Trent P McLaughlin, Samir H Mody, and Catherine Tak Piech. Daily Cost of an Intensive Care Unit Day: the Contribution of Mechanical Ventilation. *Critical Care Medicine*, 33(6):1266–1271, 2005.

[42] Daniela Pucci De Farias and Benjamin Van Roy. The linear programming approach to approximate dynamic programming. *Operations Research*, 51(6):850–865, 2003.

[43] Eric V Denardo. *Dynamic programming: Models and Applications*. Courier Corporation, 2012.

[44] Ed Diener and Eunkook Suh. Measuring quality of life: Economic, social, and subjective indicators. *Social Indicators Research*, 40(1-2):189–216, 1997.

[45] Michelle C Dimitris, Sandro Galea, Julia L Marcus, An Pan, Beate Sander, and Robert W Platt. What has the pandemic revealed about the shortcomings of modern epidemiology? what can we fix or do better? *American Journal of Epidemiology*, 2022.

[46] Ioanna Dimopoulou, Anastasia Anthi, Michalis Lignos, Efstratios Boukouvalas, Evangelos Evangelou, Christina Routsi, Konstantinos Mandragos, and Charis Roussos. Prediction of prolonged ventilatory support in blunt thoracic trauma patients. *Intensive Care Medicine*, 29(7):1101–1105, 2003.

[47] Gregory Dobson, Hsiao-Hui Lee, and Edieal Pinker. A Model of ICU bumping. *Operations Research*, 58(6):1564–1576, 2010.

[48] Molla S Donaldson, Janet M Corrigan, and Linda T Kohn. To err is human: building a safer Health system. 2000.

[49] L Doolan, R Bellomo, G Hart, H Opdam, S Uchino, W Silvester, J Buckmaster, D Goldsmith, and G Gutteridge. A before and after trial of the effect of a high-dependency Unit on postoperative morbidity and mortality. *Critical Care and Resuscitation*, 7(1):16, 2005.

[50] Soumitra R Eachempati, Lynn J Hydo, and Philip S Barie. The effect of an intermediate Care Unit on the demographics and outcomes of a surgical Intensive Care Unit population. *Archives of Surgery*, 139(3):315–319, 2004.

[51] David L Edbrooke, Cosetta Minelli, Gary H Mills, Gaetano Iapichino, Angelo Pezzi, Davide Corbella, Philip Jacobs, Anne Lippert, Joergen Wiis, and Antonio Pesenti. Implications of ICU triage decisions on patient mortality: a cost-effectiveness analysis. *Critical Care*, 15(1):R56, 2011.

[52] NX Elegant. Ventilators are key to preventing coronavirus deaths—but does the world have enough of them? Fortune. Published March 17, 2020, 2020.

[53] Agner Krarup Erlang. The Theory of probabilities and telephone conversations. *Nyt. Tidsskr. Mat. Ser. B*, 20:33–39, 1909.

[54] Andrés Esteban, Antonio Anzueto, Fernando Frutos, Inmaculada Alía, Laurent Brochard, Thomas E Stewart, Salvador Benito, Scott K Epstein, Carlos Apezteguía, and Peter Nightingale. Characteristics and outcomes in adult patients receiving mechanical ventilation: a 28-day international study. *Journal of American Medical Association*, 287 (3):345–355, 2002.

[55] Elisa Estenssoro, Francisco González, Enrique Laffaire, Héctor Canales, Gabriela Sáenz, Rosa Reina, and Arnaldo Dubin. Shock on Admission day is the best predictor of prolonged mechanical ventilation in the ICU. *Chest*, 127(2):598–603, 2005.

[56] Juan B Figueroa-Casas, Alok K Dwivedi, Sean M Connery, Raphael Quansah, Lowell Ellerbrook, and Juan Galvis. Predictive Models of prolonged mechanical ventilation yield Moderate accuracy. *Journal of Critical Care*, 30(3):502–505, 2015.

[57] TR Fleming and DP Harrington. Counting Processes and Survival Analysis, 1991. *John Wiley & Sons, Hoboken, NJ, USA*, 2011.

[58] Jason Gaitonde and Éva Tardos. Stability and learning in strategic Queuing systems. In *Proceedings of the 21st ACM Conference on Economics and Computation*, pages 319–347, 2020.

[59] Hayley B Gershengorn, Carri W Chan, Yunchao Xu, Hanxi Sun, Ronni Levy, Mor Armony, and Michelle N Gong. The impact of opening a Medical step-down Unit on Medically Critically Ill patient outcomes and throughput: a difference-in-differences analysis. *Journal of Intensive Care Medicine*, 35(5):425–437, 2020.

[60] Sanniya Khan Ghauri, Arslaan Javaeed, Khawaja Junaid Mustafa, and Abdus Salam Khan. Predictors of prolonged mechanical ventilation in patients Admitted to Intensive Care Units: A systematic review. *International Journal of Health sciences*, 13(6):31, 2019.

[61] Gotsman and Schrire. Acute Myocardial Infarction—an Ideal Concept of Progressive Coronary Care. 42:829–832, 1968.

[62] Major Greenwood. A report on the natural duration of cancer. *A Report on the Natural Duration of Cancer.*, (33), 1926.

[63] M Tanaka Gutiez and R Ramaiah. Demand versus supply in Intensive Care: an ever-growing problem. *Critical Care*, 18(1):P9, 2014.

[64] Neil A Halpern and Stephen M Pastores. Critical Care Medicine in the United States 2000–2005: an analysis of bed numbers, occupancy rates, payer mix, and costs. *Critical Care Medicine*, 38(1):65–71, 2010.

[65] Neil A Halpern and Stephen M Pastores. Critical Care Medicine beds, use, occupancy and costs in the United States: a Methodological review. *Critical Care Medicine*, 43 (11):2452, 2015.

[66] Andrew D Harding. What can an intermediate Care Unit do for you? *JONA: The Journal of Nursing Administration*, 39(1):4–7, 2009.

[67] Refael Hassin. *Rational Queueing*. CRC press, 2016.

[68] Refael Hassin and Moshe Haviv. *To Queue or not to Queue: Equilibrium behavior in Queueing systems*, volume 59. Springer Science & Business Media, 2003.

[69] Jean-Claude Hennet and Yasemin Arda. Supply chain coordination: A game-theory approach. *Engineering Applications of Artificial Intelligence*, 21(3):399–405, 2008.

[70] Theodore R Holford. Life tables with concomitant information. *Biometrics*, 32(3): 587–597, 1976.

[71] Theodore R Holford. The analysis of rates and of Survivorship using log-linear Models. *Biometrics*, 33(2):299–305, 1980.

[72] Ronald A Howard. Dynamic programming and Markov processes. 1960.

[73] Robert C Hyzy and Jakob I McSparron. Overview of initiating invasive mechanical ventilation in adults in the Intensive Care Unit. 2022.

[74] Care in Canadian ICUs. Canadian Institute for Health Information. 2016.

[75] Karthikeyan Iyengar, Shashi Bahl, Raju Vaishya, and Abhishek Vaish. Challenges and solutions in meeting up the urgent requirement of ventilators for COVID-19 patients. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, 14(4):499–501, 2020.

[76] Yeon-Soo Jang. Development of Admission and discharge criteria in Intensive Care Units. *Korean Journal of Adult Nursing*, 13(2):291–304, 2001.

[77] Robert M Kacmarek. The mechanical ventilator: past, present, and future. *Respiratory Care*, 56(8):1170–1180, 2011.

[78] JD Kalbfleisch and RL Prentice. The Survival analysis of failure time Data. 2nd, 2002.

[79] Raymond Kao, Fran Priestap, and Allan Donner. To develop a regional ICU mortality prediction Model during the first 24 h of ICU Admission utilizing MODS and NEMS with six other independent variables from the Critical Care Information System (CCIS) Ontario, Canada. *Journal of Intensive Care*, 4(1):16, 2016.

[80] Edward L Kaplan and Paul Meier. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53(282):457–481, 1958.

[81] Frank Karsten, Marco Slikker, and Geert-Jan van Houtum. Analysis of resource pooling games via a new extension of the Erlang loss function. *Tech. Rep.*, 2011.

[82] David G Kendall. Stochastic processes occurring in the Theory of Queues and their analysis by the Method of the imbedded Markov chain. *The Annals of Mathematical Statistics*, pages 338–354, 1953.

[83] Marie-Paule Kieny, Timothy Grant Evans, Stefano Scarpetta, Edward T Kelley, Niek Klazinga, Ian Forde, Jeremy Henri Maurice Veillard, Sheila Leatherman, Shamsuzzoha Syed, and Sun Mean Kim. Delivering quality Health services: a global imperative for universal Health coverage. Technical report, The World Bank, 2018.

[84] Song-Hee Kim, Carri W Chan, Marcelo Olivares, and Gabriel Escobar. ICU Admission control: An empirical study of capacity allocation and its implication for patient outcomes. *Management Science*, 61(1):19–38, 2015.

[85] Song-Hee Kim, Carri W Chan, Marcelo Olivares, and Gabriel J Escobar. Association among ICU congestion, ICU Admission decision, and patient outcomes. *Critical Care Medicine*, 44(10):1814–1821, 2016.

[86] John P Klein and Melvin L Moeschberger. *Survival analysis: techniques for censored and truncated Data*, volume 1230. Springer, 2003.

[87] Leonard Kleinrock. *Theory, volume 1, Queueing systems*. Wiley-Interscience, 1975.

[88] William A Knaus, Jack E Zimmerman, Douglas P Wagner, Elizabeth A Draper, and Diane E Lawrence. APACHE-acute physiology and chronic Health evaluation: a physiologically based classification system. *Critical Care Medicine*, 9(8):591–597, 1981.

[89] Andrey Kolobov. Planning with Markov decision processes: An AI perspective. *Synthesis Lectures on Artificial Intelligence and Machine Learning*, 6(1):1–210, 2012.

[90] Simon Lambden, Pierre Francois Laterre, Mitchell M Levy, and Bruno Francois. The SOFA score—development, utility and challenges of accurate assessment in clinical trials. *Critical Care*, 23(1):1–9, 2019.

[91] Elisa T Lee and John Wang. *Statistical Methods for Survival Data analysis*, volume 476. John Wiley & Sons, 2003.

[92] JF Légaré, GM Hirsch, KJ Buth, C MacDougall, and JA Sullivan. Preoperative prediction of prolonged mechanical ventilation following coronary artery bypass grafting. *European Journal of Cardio-thoracic Surgery*, 20(5):930–936, 2001.

[93] Qian Lei, Lei Chen, Yi Zhang, Nengxin Fang, Weiping Cheng, and Lihuan Li. Predictors of prolonged mechanical ventilation after aortic arch surgery with deep hypothermic circulatory arrest plus antegrade selective cerebral perfusion. *Journal of Cardiothoracic and Vascular Anesthesia*, 23(4):495–500, 2009.

[94] Suparerk Lekwijit, Carri W Chan, Linda V Green, Vincent X Liu, and Gabriel J Escobar. The Impact of Step-Down Unit Care on Patient Outcomes After ICU Discharge. *Critical Care Explorations*, 2(5):e0114, 2020.

[95] Phillip D Levin. The process of intensive care triage, 2001.

[96] Jing Li, Ming Dong, and Wenhui Zhao. Admissions optimisation and premature discharge decisions in Intensive Care Units. *International Journal of Production Research*, 53(24):7329–7342, 2015.

[97] Xuanjing Li, Dacheng Liu, Na Geng, and Xiaolei Xie. Optimal ICU Admission Control With Premature Discharge. *IEEE Transactions on Automation Science and Engineering*, 16(1):148–164, 2018.

[98] Frances Lin, Wendy Chaboyer, and Marianne Wallis. A literature review of organisational, individual and teamwork factors contributing to the ICU discharge process. *Australian Critical Care*, 22(1):29–43, 2009.

[99] Michael L Littman, Thomas L Dean, and Leslie Pack Kaelbling. On the complexity of solving Markov decision problems. *arXiv preprint arXiv:1302.4971*, 2013.

[100] Elisa F Long and Kusum S Mathews. The boarding patient: Effects of ICU and hospital occupancy surges on patient flow. *Production and Operations Management*, 27(12): 2122–2143, 2018.

[101] Lijian Lu, Carri Chan, Linda Green, and Gabriel J Escobar. The impact of step-down Unit Care on patient outcomes. *Columbia Business School Research Paper*, (16-74), 2014.

[102] David G Luenberger and Yinyu Ye. *Linear and nonlinear programming*, volume 2. Springer, 1984.

[103] Fang Luo, Djillali Annane, David Orlikowski, Li He, Mi Yang, Muke Zhou, and Guan J Liu. Invasive versus non-invasive ventilation for acute Respiratory failure in neuromuscular disease and chest wall disorders. *Cochrane Database of Systematic Reviews*, (12), 2017.

[104] Xiya Ma and Dominique Vervoort. Critical Care capacity during the COVID-19 pandemic: global availability of Intensive Care beds. *Journal of Critical Care*, 58:96, 2020.

[105] William Matthew Makeham. On the law of mortality and the construction of annuity tables. *The Assurance Magazine, and Journal of the Institute of Actuaries*, 8(6):301–310, 1860.

[106] Alan S Manne. Linear programming and sequential decisions. *Management Science*, 6 (3):259–267, 1960.

[107] Nathan Mantel. Evaluation of Survival Data and Two New Rank Order Statistics Arising in its Consideration. *Cancer Chemother Rep*, 50(3):163–170, 1966.

[108] Sarah Markus. Effect of Emergency Department and ICU Occupancy on Admission Decisions and Outcomes for Critically Ill Patients. *Journal of Emergency Medicine*, 55 (3):451–452, 2018.

[109] KS Mathews, GY Jenq, MA Pisani, and EF Long. Characterizing the flow of short-stay patients in the Intensive Care Unit. In *Society for Medical Decision Making 34th Annual Conference*, 2012.

[110] Kusum S Mathews and Elisa F Long. A conceptual framework for improving Critical Care patient flow and bed use. *Annals of the American Thoracic Society*, 12(6):886–894, 2015.

[111] David R Mcilroy, BD Coleman, and Paul S Myles. Outcomes following a shortage of high dependency Unit beds for surgical patients. *Anaesthesia and Intensive Care*, 34(4): 457–463, 2006.

[112] Reis D. Miranda. The therapeutic intervention scoring system: one single tool for the evaluation of workload, the work process and Management? *Intensive Care Medicine*, 23(6):615, 1997.

[113] Reis D. Miranda, Rui Moreno, and Gaetano Iapichino. Nine equivalents of Nursing manpower use score (NEMS). *Intensive Care Medicine*, 23(7):760–765, 1997.

[114] Seyed M Moghadas, Affan Shoukat, Meagan C Fitzpatrick, Chad R Wells, Pratha Sah, Abhishek Pandey, Jeffrey D Sachs, Zheng Wang, Lauren A Meyers, and Burton H Singer. Projecting hospital utilization during the COVID-19 outbreaks in the United States. *Proceedings of the National Academy of Sciences*, 117(16):9122–9126, 2020.

[115] John F Nash. Equilibrium points in n-person games. *Proceedings of the National Academy of Sciences*, 36(1):48–49, 1950.

[116] Stanley A Nasraway, Ian L Cohen, Richard C Dennis, Michelle A Howenstein, Diana K Nikas, Jonathan Warren, and Suzanne K Wedel. Guidelines on Admission and discharge for adult intermediate Care Units. *Critical Care Medicine*, 26(3):607–610, 1998.

[117] Joseph L Nates, Mark Nunnally, Ruth Kleinpell, Sandralee Blosser, Jonathan Goldner, Barbara Birriel, Clara S Fowler, Diane Byrum, William Scherer Miles, and Heatherlee Bailey. ICU Admission, discharge, and triage guidelines: a framework to enhance clinical Operations, development of institutional policies, and further Research. *Critical Care Medicine*, 44(8):1553–1602, 2016.

[118] Luiz Guilherme Nadal Nunes, Solon Venancio de Carvalho, and Rita de Cássia Meneses Rodrigues. Markov decision process Applied to the control of hospital elective Admissions. *Artificial intelligence in Medicine*, 47(2):159–171, 2009.

[119] American College of Critical Care Medicine of the Society of Critical Care Medicine. Guidelines for ICU Admission, discharge, and triage. *Crit Care Med*, 27(3):633–638, 1999.

[120] Martin J Osborne and Ariel Rubinstein. *A Course in Game Theory*. MIT press, 1994.

[121] Christos H Papadimitriou and John N Tsitsiklis. The Complexity of Markov Decision Processes. *Mathematics of Operations Research*, 12(3):441–450, 1987.

[122] Jonathan Patrick. A Markov decision Model for determining optimal outpatient scheduling. *Health Care Management Science*, 15(2):91–102, 2012.

[123] Jonathan Patrick, Martin L Puterman, and Maurice Queyranne. Dynamic multipriority patient scheduling for a diagnostic resource. *Operations Research*, 56(6):1507–1525, 2008.

[124] Joshua M Pearce. A review of open source ventilators for COVID-19 and future pandemics. *F1000Research*, 9, 2020.

[125] Andreas Perren, Marco Previsdomini, Ilaria Perren, and Paolo Merlani. High accuracy of the nine equivalents of Nursing manpower use score assessed by Critical Care Nurses. *Swiss Medical Weekly*, 142(1314), 2012.

[126] Raquel Ferrari Piotto, Fabricio Beltrame Ferreira, Flávia Cortez Colósimo, Gilmara Silveira da Silva, Alexandre Gonçalves de Sousa, and Domingo Marcolino Braile. Independent predictors of prolonged mechanical ventilation after coronary artery bypass surgery. *Brazilian Journal of Cardiovascular Surgery*, 27(4):520–528, 2012.

[127] Warren B Powell. *Approximate Dynamic Programming: Solving the curses of dimensionality*, volume 703. John Wiley & Sons, 2007.

[128] Fran Priestap, Raymond Kao, and Claudio M Martin. External validation of a prognostic Model for Intensive Care Unit mortality: A retrospective study using the Ontario Critical Care Information System. *Canadian Journal of Anesthesia*, 67(8):981–991, 2020.

[129] Meghan Prin and Hannah Wunsch. The role of stepdown beds in hospital Care. *American Journal of Respiratory and Critical Care Medicine*, 190(11):1210–1216, 2014.

[130] Martin L Puterman. *Markov decision processes: Discrete stochastic dynamic programming*. John Wiley & Sons, 2014.

[131] Richard D Remington and M Anthony Schork. *Statistics with Applications to the biological and Health sciences.* Englewood Cliffs, NJ: Prentice-Hall, Inc, 1970.

[132] Abbas Ali Rezaee, Mohammad Hossein Yaghmaee, and Amir Masoud Rahmani. Optimized congestion Management protocol for HealthCare wireless sensor networks. *Wireless Personal Communications*, 75(1):11–34, 2014.

[133] Boyd F Richards, J Brett Fleming, Chevis N Shannon, Beverly C Walters, and Mark R Harrigan. Safety and cost effectiveness of step-down Unit Admission following elective

neurointerventional procedures. *Journal of neurointerventional surgery*, 4(5):390–392, 2012.

[134] Felipe F. Rodrigues, Gregory S. Zaric, and David A. Stanford. Discrete event simulation Model for planning Level 2 "step-down" bed needs using NEMS. *Operations Research for Health Care*, 17:42–54, 2018.

[135] Thomas L Rodziewicz, Benjamin Houseman, and John E Hipskind. Medical error prevention. 2018.

[136] Sheldon M Ross. *Introduction to Probability Models*. Academic Press, 2014.

[137] GAGE RP. Calculation of Survival rates for cancer. In *Proceedings of the staff meetings. Mayo Clinic*, volume 25, pages 270–286, 1950.

[138] Andrzej Ruszczynski. *Nonlinear Optimization*. Princeton University Press, 2011.

[139] Somayeh Sadat, Hossein Abouee-Mehrizi, and Michael W Carter. Can hospitals compete on quality? *Health Care Management Science*, 18(3):376–388, 2015.

[140] Mohammed Sayed, David Riaño, and Jesús Villar. Predicting Duration of Mechanical Ventilation in Acute Respiratory Distress Syndrome Using Supervised Machine Learning. *Journal of Clinical Medicine*, 10(17):3824, 2021.

[141] Paul J Schweitzer and Abraham Seidmann. Generalized polynomial approximations in Markovian decision processes. *Journal of Mathematical Analysis and Applications*, 110 (2):568–582, 1985.

[142] J Seidel, PC Whiting, and DL Edbrooke. The costs of Intensive Care. *Continuing Education in Anaesthesia, Critical Care & Pain*, 6(4):160–163, 2006.

[143] Michael G Seneff, Jack E Zimmerman, William A Knaus, Douglas P Wagner, and Elizabeth A Draper. Predicting the duration of mechanical ventilation: the importance of disease and patient characteristics. *Chest*, 110(2):469–479, 1996.

[144] Pengyi Shi, Mabel C Chou, JG Dai, Ding Ding, and Joe Sim. Models and insights for hospital inpatient Operations: Time-dependent ED boarding time. *Management Science*, 62(1):1–28, 2016.

[145] Amir Shmueli, Charles L Sprung, and Edward H Kaplan. Optimizing Admissions to an Intensive Care Unit. *Health Care Management Science*, 6(3):131–136, 2003.

[146] John F Shortle, James M Thompson, Donald Gross, and Carl M Harris. *Fundamentals of Queueing Theory*, volume 399. John Wiley & Sons, 2018.

[147] Affan Shoukat, Chad R Wells, Joanne M Langley, Burton H Singer, Alison P Galvani, and Seyed M Moghadas. Projecting demand for critical care beds during covid-19 outbreaks in canada. *Cmaj*, 192(19):E489–E496, 2020.

[148] Luigi Siciliani and Jeremy Hurst. Tackling excessive waiting times for elective surgery: a comparative analysis of policies in 12 OECD countries. *Health Policy*, 72(2):201–215, 2005.

[149] Olivier Sigaud and Olivier Buffet. *Markov Decision Processes in Artificial Intelligence*. John Wiley & Sons, 2013.

[150] Soraia Aparecida da Silva, Reginaldo Aparecido Valácio, Flávia Carvalho Botelho, and Carlos Faria Santos Amaral. Reasons for discharge delays in teaching hospitals. *Revista de saude publica*, 48(2):314–321, 2014.

[151] Gary Smith and Mick Nielsen. Criteria for Admission. *British Medical Association.*, 318(7197):1544–1547, 1999.

[152] Charles L Sprung, Marion Danis, Gaetano Iapichino, Antonio Artigas, Jozef Kesecioglu, Rui Moreno, Anne Lippert, J Randall Curtis, Paula Meale, and Simon L Cohen. Triage of Intensive Care patients: identifying agreement and controversy. *Intensive Care Medicine*, 39(11):1916–1924, 2013.

[153] A von Stackelberg. Dolichopodidae. *Die Fliegen der Palaearktischen Region*, 4(5), 1934.

[154] Henry T Stelfox, Andrea Soo, Daniel J Niven, Kirsten M Fiest, Hannah Wunsch, Kathryn M Rowan, and Sean M Bagshaw. Assessment of the safety of discharging select patients directly home from the Intensive Care Unit: a multicenter population-based cohort study. *Journal of the American Medical Association Internal Medicine*, 178(10):1390–1399, 2018.

[155] K Strand and H Flaatten. Severity scoring in the ICU: A review. *Acta Anaesthesiologica Scandinavica*, 52(4):467–478, 2008.

[156] Jianguo Sun. *The Statistical Analysis of Interval-censored Failure Time Data*, volume 3. Springer, 2006.

[157] Richard S Sutton and Andrew G Barto. *Introduction to reinforcement learning*, volume 135. MIT Press Cambridge, 1998.

[158] Judith Timmer and Werner Scheinhardt. How to share the cost of cooperating Queues in a tandem network? In *2010 22nd International Teletraffic Congress (lTC 22)*, pages 1–7. IEEE, 2010.

[159] Judith Timmer and Werner Scheinhardt. Cost sharing of cooperating Queues in a Jackson network. *Queueing Systems*, 75(1):1–17, 2013.

[160] Judith Timmer and Werner Scheinhardt. Customer and cost sharing in a Jackson network. *International Game Theory Review*, 20(03):1850002, 2018.

[161] Gilles Troche and Pierre Moine. Is the duration of mechanical ventilation predictable? *Chest*, 112(3):745–751, 1997.

[162] Jean-Louis Trouillet, Alain Combes, Elisabeth Vaissier, Charles-Edouard Luyt, Alexandre Ouattara, Alain Pavie, and Jean Chastre. Prolonged mechanical ventilation after cardiac surgery: outcome and predictors. *The Journal of Thoracic and Cardiovascular Surgery*, 138(4):948–953, 2009.

[163] Jean-Louis Trouillet, Charles-Edouard Luyt, Marguerite Guiguet, Alexandre Ouattara, Elisabeth Vaissier, Ralouka Makri, Ania Nieszkowska, Pascal Leprince, Alain Pavie, and Jean Chastre. Early percutaneous tracheotomy versus prolonged intubation of mechanically ventilated patients after cardiac surgery: a randomized trial. *Annals of Internal Medicine*, 154(6):373–383, 2011.

[164] Martin Urner, Peter Jüni, Bettina Hansen, Marian S Wettstein, Niall D Ferguson, and Eddy Fan. Time-varying intensity of mechanical ventilation and mortality in patients with acute Respiratory failure: a registry-based, prospective cohort study. *The Lancet Respiratory Medicine*, 8(9):905–913, 2020.

[165] Peter T VanBerkel and John T Blake. A comprehensive simulation for wait time reduction and capacity planning Applied in general surgery. *Health Care Management Science*, 10(4):373–385, 2007.

[166] John Von Neumann and Oskar Morgenstern. *Theory of Games and Economic Behavior (commemorative edition)*. Princeton University Press, 2007.

[167] Ljiljana Vuković. Assessment of Nurses' Workload in Intensive Care Unit by Use of Scoring Systems. *Croatian Nursing Journal*, 4(1):59–71, 2020.

[168] Tao Wang, Michael Bowling, and Dale Schuurmans. Dual representations for dynamic programming and reinforcement learning. In *2007 IEEE International Symposium on Approximate Dynamic Programming and Reinforcement Learning*, pages 44–51. IEEE, 2007.

[169] Hannah Wunsch, Derek C Angus, David A Harrison, Olivier Collange, Robert Fowler, Eric AJ Hoste, Nicolette F de Keizer, Alexander Kersten, Walter T Linde-Zwirble, and Alberto Sandiumenge. Variation in Critical Care services across North America and Western Europe. *Critical Care Medicine*, 36(10):2787–e8, 2008.

[170] Hannah Wunsch, Hayley Gershengorn, and Damon C Scales. Economics of ICU organization and Management. *Critical Care Clinics*, 28(1):25–37, 2012.

[171] Mohammad Hossein Yaghmaee, Nazbanoo Farzaneh Bahalgardi, and Donald Adjeroh. A prioritization based congestion control protocol for HealthCare monitoring Application in wireless sensor networks. *Wireless Personal Communications*, 72(4):2605–2631, 2013.

[172] Yinlian Zeng, Xiaoqiang Cai, Lianmin Zhang, and Jun Li. A core cost allocation for capacity transfer among M/M/1 Queueing systems. In *2016 13th International Conference on Service Systems and Service Management (ICSSSM)*, pages 1–4. IEEE, 2016.

[173] Yinlian Zeng, Lianmin Zhang, Xiaoqiang Cai, and Jun Li. Cost sharing for capacity transfer in cooperating Queueing systems. *Production and Operations Management*, 27 (4):644–662, 2018.

[174] Christopher W Zobel and William T Scherer. An empirical study of policy convergence in Markov decision process value iteration. *Computers & Operations Research*, 32(1): 127–142, 2005.

Appendix A

Complementary on Chapter 4

A.1 Payoff functions

A.1.1 Cooperation

Elements of the Hessian matrix

$$
\begin{aligned}
\frac{\partial^2 S_c}{\partial^2 l_1} &= \lambda\left(-\frac{2c\lambda^3\,(l_1+l_2)^{\,2}}{(1-\lambda\,(l_1+l_2))^{\,3}}-\frac{4c\lambda^2\,(l_1+l_2)}{(1-\lambda\,(l_1+l_2))^{\,2}}-\frac{2c\lambda}{1-\lambda\,(l_1+l_2)}\right)+\lambda\left(\frac{2cl_2}{(l_1+l_2)^{\,2}}-\frac{2cl_1l_2}{(l_1+l_2)^{\,3}}\right)\\
\frac{\partial^2 S_c}{\partial^2 l_2} &= \lambda\left(-\frac{2c\lambda^3\,(l_1+l_2)^{\,2}}{(1-\lambda\,(l_1+l_2))^{\,3}}-\frac{4c\lambda^2\,(l_1+l_2)}{(1-\lambda\,(l_1+l_2))^{\,2}}-\frac{2c\lambda}{1-\lambda\,(l_1+l_2)}\right)+\lambda\left(\frac{2cl_1}{(l_1+l_2)^{\,2}}-\frac{2cl_1l_2}{(l_1+l_2)^{\,3}}\right)\\
\frac{\partial^2 S_c}{\partial l_1\partial l_2} &= \lambda\left(-\frac{2c\lambda^3\,(l_1+l_2)^{\,2}}{(1-\lambda\,(l_1+l_2))^{\,3}}-\frac{4c\lambda^2\,(l_1+l_2)}{(1-\lambda\,(l_1+l_2))^{\,2}}-\frac{2c\lambda}{1-\lambda\,(l_1+l_2)}\right)\\
&\quad+\lambda\left(-\frac{c}{l_1+l_2}+\frac{cl_1}{(l_1+l_2)^{\,2}}+\frac{cl_2}{(l_1+l_2)^{\,2}}-\frac{2cl_1l_2}{(l_1+l_2)^{\,3}}\right)\\
\frac{\partial^2 S_c}{\partial l_2\partial l_1} &= \lambda\left(-\frac{2c\lambda^3\,(l_1+l_2)^{\,2}}{(1-\lambda\,(l_1+l_2))^{\,3}}-\frac{4c\lambda^2\,(l_1+l_2)}{(1-\lambda\,(l_1+l_2))^{\,2}}-\frac{2c\lambda}{1-\lambda\,(l_1+l_2)}\right)\\
&\quad+\lambda\left(-\frac{c}{l_1+l_2}+\frac{cl_1}{(l_1+l_2)^{\,2}}+\frac{cl_2}{(l_1+l_2)^{\,2}}-\frac{2cl_1l_2}{(l_1+l_2)^{\,3}}\right)
\end{aligned}
\tag{A.1}
$$

Full System payoff under cooperation

$$
\begin{aligned}
S_{\mathrm{CP}} &= \frac{2(2c(r-1)+c(3c-r)(c+r)-r+1)}{4c}\left(2c\sqrt{\frac{1}{3c^2+2c(r+1)-(r-1)^2}}-1\right)\\
&\quad+\frac{\left(1-(c-r)^2\right)}{4c}\left(1-2\sqrt{\frac{c^2}{3c^2+2c(r+1)-(r-1)^2}}\right)
\end{aligned}
\tag{A.2}
$$

A.1.2 Simultaneous Decision

Elements of the Hessian matrix

$$\begin{aligned}\Delta S = \Bigg\langle & \lambda\left(-\frac{2c\lambda^3\left(l_1+l_2\right)^2}{\left(1-\lambda\left(l_1+l_2\right)\right)^3}-\frac{4c\lambda^2\left(l_1+l_2\right)}{\left(1-\lambda\left(l_1+l_2\right)\right)^2}-\frac{2c\lambda}{1-\lambda\left(l_1+l_2\right)}\right),\\ & \lambda\left(-\frac{2c\lambda^3\left(l_1+l_2\right)^2}{\left(1-\lambda\left(l_1+l_2\right)\right)^3}-\frac{4c\lambda^2\left(l_1+l_2\right)}{\left(1-\lambda\left(l_1+l_2\right)\right)^2}-\frac{2c\lambda}{1-\lambda\left(l_1+l_2\right)}\right),\\ & \lambda\left(-\frac{c}{l_1+l_2}+\frac{cl_1}{\left(l_1+l_2\right)^2}+\frac{cl_2}{\left(l_1+l_2\right)^2}-\frac{2cl_1l_2}{\left(l_1+l_2\right)^3}\right),\\ & \lambda\left(\frac{2cl_1}{\left(l_1+l_2\right)^2}-\frac{2cl_1l_2}{\left(l_1+l_2\right)^3}\right)\Bigg\rangle\end{aligned} \tag{A.3}$$

Full System payoff under Simultaneous decision

$$S^{ST}=\sqrt{\frac{r}{c}}-\sqrt{\frac{c}{c+1}}\left(\sqrt{\frac{r}{c}}+2c+1\right)+\left(\sqrt{\frac{c}{c+1}}-1\right)\left(r\left(\sqrt{\frac{r}{c}}-2\right)+\sqrt{cr}\right)+2c \tag{A.4}$$

A.1.3 ICU Stackelberg Decision

System payoff under ICU Stackelberg

$$S^{IS}_{ICU}=-\frac{r\left(-c\sqrt{c^2+\sqrt{cr}}+c^2+\sqrt{cr}\right)\left(-c^3\sqrt{c^2+\sqrt{cr}}+c^2r\sqrt{\frac{c}{r}}-c\left(r\sqrt{\frac{c}{r}}\sqrt{c^2+\sqrt{cr}}+r\right)+c^4+r\sqrt{\frac{c}{r}}\sqrt{cr}\right)}{c\left(c\sqrt{cr}+r\right)\left(c^2\left(\sqrt{cr}-r\sqrt{\frac{c}{r}}\right)+c\left(r\sqrt{\frac{c}{r}}\sqrt{c^2+\sqrt{cr}}+r\right)+r\left(-\sqrt{\frac{c}{r}}\right)\sqrt{cr}\right)} \tag{A.5}$$

$$S^{IS}_{SDU}=-\frac{r\left(-2r\sqrt{\frac{c}{r}}+c+r\right)\left(-c\sqrt{c^2+\sqrt{cr}}+c^2+\sqrt{cr}\right)}{c\left(c\sqrt{cr}+r\right)} \tag{A.6}$$

A.1.4 SDU Stackelberg Decision

Full System payoff under SDU Stackelberg

$$S^{SS}=\frac{1}{2}-\frac{3}{2}\sqrt{\frac{c}{c+1}}\frac{\left(7c^2+2(c+1)r-r^2\right)}{4c}\left(1-\sqrt{\frac{c}{c+1}}\right) \tag{A.7}$$

Appendix B

Complementary on Chapter 5

Table B.1: Selected estimates from the K-M curve.

Time	n.risk	n.event	n.censor	surv	UCI	LCI
1	49,467	15,125	646	0.694	0.698	0.690
2	33,696	12,213	273	0.443	0.447	0.438
3	21,210	5,179	129	0.335	0.339	0.330
4	15,902	2,979	60	0.272	0.276	0.268
5	12,863	2,092	39	0.228	0.231	0.224
6	10,732	1,594	38	0.194	0.197	0.190
7	9,100	1,270	28	0.167	0.170	0.163
8	7,802	998	33	0.145	0.149	0.142
9	6,771	866	26	0.127	0.130	0.124
10	5,879	698	20	0.112	0.115	0.109
15	3,249	326	8	0.064	0.066	0.062
20	1,963	176	0	0.040	0.041	0.038
25	1,258	110	6	0.026	0.027	0.024
30	791	68	2	0.016	0.018	0.015
35	528	41	2	0.011	0.012	0.010
40	334	26	0	0.007	0.008	0.006
45	201	25	1	0.004	0.005	0.004
50	119	12	0	0.003	0.003	0.002
55	56	12	0	0.001	0.001	0.001
59	10	9	0	0.000	0.000	0.000

n.risk (number at risk at time t), n.event (number of events at time *t*), surv (Survival probability), UCI (upper confident interval limit), LCI (lower confident interval limit).

Table B.2: Description of the Log rank uni-variate test

Covariates		Counts	events
Pre ICU LOS (χ^2 = 262 on 2 df, p <2e-16)			
	<=1	35911	34925
	<8	7661	7430
	>7	5895	5705
Age group (χ^2 = 35 on 2 df, p= 0.00000003)			
	18-35	4355	4209
	36-64	38009	36914
	65+	7103	6937
Sex (χ^2 = 77.8 on 1 df, p <2e-16)			
	Female	18185	17696
	Male	31282	30364
ICU Site (χ^2 = 3350 on 65 df, p <2e-16			)
	3970	359	349
	3972	296	293
	3985	2417	2353
	3986	530	526
	3987	398	387
	3996	261	254
	4001	2352	2292
	4044	291	251
	4045	612	604
	4052	882	870
	4054	786	774
	4056	476	468
	4057	80	76
	4063	1596	1578
	4071	31	29
	4073	586	568
	4076	127	123
	4079	1127	1101
	4085	813	802
	4089	58	57
	4090	2881	2856

4093	27	27
4097	270	261
4103	196	193
4107	559	546
4108	126	123
4109	228	221
4110	504	490
4123	159	145
4130	409	400
4131	170	163
4138	540	524
4144	281	278
4168	127	124
4171	660	653
4180	1563	1530
4186	57	55
4192	334	327
4193	62	62
4197	29	28
4199	138	131
4205	2516	2474
4209	427	423
4231	4112	3995
4233	661	655
4235	302	278
4238	242	236
4241	58	52
4245	352	348
4260	127	123
4265	3567	3508
4266	786	776
4285	642	625
4303	2570	2088
4310	3215	3198
4311	1441	1428
4315	489	482

	4414	417	406
	4471	180	170
	4774	754	739
	4799	622	620
	4832	1265	1255
	4837	430	413
	4839	399	392
	4841	230	224
	4845	265	260
Admission source (χ^2 = 6875 on 6 df, p <2e-16)			
	Downstream Unit	2971	2864
	ED	12507	12217
	Home	191	181
	Hospital	5210	5000
	OR	22608	21927
	Other	373	358
	Ward	5607	5513
Diagnosis (χ^2 = 5213 on 4 df, p <2e-16)			
	Cardiovascular	22269	21517
	Gastrointestinal	2920	2866
	Neurological	4652	4557
	Other	17699	17255
	Trauma	1927	1865
Physician Referral (χ^2 = 5171 on 3 df, p <2e-16)			
	Medical	16930	16457
	Other	9325	9098
	Respirology	1364	1316
	Surgical	21848	21189
Patient Category (χ^2 = 3929 on 1 df, p <2e-16)			
	Medical	22811	22189
	Surgical	26656	25871
MODS (χ^2 = 1107 on 4 df, p <2e-16)			
	<=1	4389	4070
	1 – 4	14185	13704
	5 – 8	16070	15787
	9 – 13	6861	6721

	>13	7962	7778
NEMS (χ^2 = 157 on 2 df, p <2e-16)			
	<22	629	613
	C2	11823	11462
	C3	37015	35985
Schedule admission (χ^2 = 9579 on 1 df, p <2e-16)			
	No	32488	31629
	Yes	16979	16431
Schedule surgery (χ^2 = 8809 on 1 df, p <2e-16)			
	No	30670	29858
	Yes	18797	18202
Schedule (χ^2 = 0.1 on 1 df, p= 0.8)			
	No	33	31
	Yes	49434	48029
CentralVenousLine (χ^2 = 53.9 on 1 df, p= 2e-13)			
	No	13383	12996
	Yes	36084	35064
ArterialLine (χ^2 = 79.4 on 1 df, p <2e-16)			
	No	9065	8822
	Yes	40402	39238
IntracranialPressureMonito (χ^2 = 227 on 1 df, p <2e-16)			
	No	48648	47253
	Yes	819	807
Dialysis. (χ^2 = 194 on 1 df, p <2e-16			)
	No	48046	46668
	Yes	1421	1392
ExtracorporealMembraneOxygen (χ^2 = 79.7 on 1 df, p <2e-16)			
	No	49331	47934
	Yes	136	126
IntraAorticBalloonPump (χ^2 = 22.5 on 1 df, p= 0.000002)			
	No	48771	47390
	Yes	696	670
OtherInterventionsWithinthisUnit (χ^2 = 878 on 1 df, p <2e-16)			
	No	33331	32357
	Yes	16136	15703

InterventionsOutsidethisUnit (χ^2 = 554 on 1 df, p <2e-16)			
	No	38658	37531
	Yes	10809	10529

Table B.3: Test of Cox PH assumption

Covariate	χ^2	df	$p-$value
Site Code	1156.267	65	$< 2e-16$
Admission Source	847.177	6	$< 2e-16$
Diagnosis	851.135	4	$< 2e-16$
Is Scheduled ICU Admission	953.906	1	$< 2e-16$
Is Scheduled Surgery	884.831	1	$< 2e-16$
Patient Category	609.514	1	$< 2e-16$
Physician Specialist	688.851	3	$< 2e-16$
Central Venous Line	29.441	1	$5.8e-08$
Arterial Line	0.206	1	0.6502
Intracranial Pressure Monitor.	69.925	1	$< 2e-16$
Dialysis	33.588	1	$6.8e-09$
Extra-corporeal Membrane Oxygen	7.091	1	0.0077
Intra-Aortic Balloon Pump	21.868	1	2.9e-06
Other Interventions Within this Unit	113.692	1	$< 2e-16$
Interventions Outside this Unit	186.073	1	$< 2e-16$
Age	27.107	1	1.9e-07
Gender	8.828	1	0.0030
Age group	7.381	2	0.0250
Pre LOS	28.340	2	7.0e-07
MODS Cat	145.383	4	$< 2e-16$
NEMS Cat	21.775	2	1.9e-05
GLOBAL	2552.177	101	$< 2e-16$

χ^2 (Chi-square value), df (degree of freedom)

Table B.4: Survival distribution of log-logistic AFT model for CCIS data

Covaraite	est	L95%	U95%	exp(est)	L95%	U95%
shape	2.26	2.24	2.28	NA	NA	NA
scale	1.11	0.96	1.27	NA	NA	NA
SiteCode 3972	0.31	0.16	0.46	1.36	1.18	1.58
SiteCode 3985	0.39	0.28	0.50	1.48	1.32	1.64
SiteCode 3986	0.48	0.34	0.61	1.61	1.41	1.84
SiteCode 3987	0.22	0.08	0.36	1.25	1.09	1.43
SiteCode 3996	0.55	0.39	0.71	1.73	1.48	2.03
SiteCode 4001	0.20	0.09	0.31	1.22	1.09	1.36
SiteCode 4044	0.34	0.18	0.50	1.40	1.20	1.64
SiteCode 4045	0.25	0.12	0.37	1.28	1.13	1.45
SiteCode 4052	0.35	0.23	0.47	1.41	1.25	1.59
SiteCode 4054	0.43	0.31	0.55	1.53	1.36	1.73
SiteCode 4056	0.49	0.35	0.62	1.62	1.42	1.86
SiteCode 4057	0.10	-0.12	0.33	1.11	0.89	1.39
SiteCode 4063	0.03	-0.08	0.14	1.03	0.92	1.15
SiteCode 4071	0.07	-0.25	0.40	1.08	0.78	1.49
SiteCode 4073	0.31	0.19	0.44	1.36	1.20	1.55
SiteCode 4076	0.57	0.36	0.78	1.76	1.43	2.18
SiteCode 4079	0.43	0.31	0.55	1.54	1.37	1.73
SiteCode 4085	0.40	0.28	0.52	1.49	1.32	1.68
SiteCode 4089	0.44	0.17	0.70	1.55	1.19	2.02
SiteCode 4090	0.27	0.16	0.38	1.31	1.18	1.46
SiteCode 4093	-0.29	-0.64	0.06	0.75	0.53	1.06
SiteCode 4097	0.41	0.25	0.56	1.50	1.29	1.75
SiteCode 4103	0.19	0.03	0.35	1.21	1.03	1.42
SiteCode 4107	0.46	0.33	0.59	1.58	1.39	1.80
SiteCode 4108	0.71	0.51	0.91	2.04	1.67	2.49
SiteCode 4109	0.26	0.10	0.42	1.30	1.11	1.53
SiteCode 4110	0.11	-0.01	0.24	1.12	0.99	1.27
SiteCode 4123	0.30	0.13	0.47	1.35	1.13	1.60
SiteCode 4130	0.37	0.23	0.51	1.45	1.26	1.66
SiteCode 4131	0.04	-0.14	0.21	1.04	0.87	1.24
SiteCode 4138	0.63	0.50	0.77	1.88	1.65	2.15

SiteCode 4144	0.47	0.31	0.62	1.59	1.37	1.86
SiteCode 4168	0.22	0.03	0.42	1.25	1.03	1.52
SiteCode 4171	0.43	0.31	0.56	1.54	1.36	1.75
SiteCode 4180	0.17	0.06	0.28	1.19	1.06	1.33
SiteCode 4186	0.26	0.00	0.52	1.30	1.00	1.69
SiteCode 4192	0.51	0.37	0.65	1.66	1.44	1.92
SiteCode 4193	0.35	0.08	0.62	1.42	1.09	1.86
SiteCode 4197	0.05	-0.32	0.42	1.06	0.73	1.53
SiteCode 4199	-0.12	-0.30	0.06	0.89	0.74	1.07
SiteCode 4205	0.47	0.36	0.58	1.60	1.43	1.78
SiteCode 4209	0.15	0.01	0.28	1.16	1.01	1.33
SiteCode 4231	0.21	0.10	0.32	1.23	1.11	1.37
SiteCode 4233	0.28	0.15	0.40	1.32	1.16	1.49
SiteCode 4235	0.33	0.17	0.48	1.39	1.19	1.62
SiteCode 4238	0.37	0.21	0.52	1.44	1.24	1.68
SiteCode 4241	0.22	-0.05	0.49	1.25	0.96	1.63
SiteCode 4245	0.28	0.14	0.42	1.33	1.15	1.53
SiteCode 4260	0.29	0.07	0.50	1.33	1.08	1.65
SiteCode 4265	0.58	0.47	0.69	1.79	1.60	1.99
SiteCode 4266	0.55	0.43	0.68	1.74	1.54	1.96
SiteCode 4285	0.19	0.07	0.32	1.21	1.07	1.37
SiteCode 4303	0.49	0.38	0.60	1.63	1.46	1.82
SiteCode 4310	0.29	0.18	0.39	1.33	1.20	1.48
SiteCode 4311	0.26	0.15	0.37	1.30	1.16	1.45
SiteCode 4315	0.37	0.24	0.50	1.45	1.27	1.66
SiteCode 4414	0.37	0.22	0.52	1.44	1.25	1.67
SiteCode 4471	0.28	0.11	0.45	1.32	1.11	1.57
SiteCode 4774	0.35	0.23	0.48	1.42	1.26	1.61
SiteCode 4799	0.46	0.34	0.59	1.59	1.40	1.81
SiteCode 4832	0.33	0.22	0.45	1.40	1.24	1.57
SiteCode 4837	0.44	0.30	0.58	1.55	1.35	1.79
SiteCode 4839	0.73	0.58	0.88	2.08	1.79	2.41
SiteCode 4841	0.44	0.27	0.60	1.55	1.31	1.83
SiteCode 4845	0.47	0.31	0.62	1.60	1.36	1.87
Admission Source ED	-0.13	-0.17	-0.08	0.88	0.84	0.92
Admission Source Home	0.13	-0.01	0.27	1.14	0.99	1.31

Admission Source Hospital	0.11	0.06	0.16	1.12	1.07	1.17
Admission Source OR	-0.31	-0.36	-0.26	0.73	0.70	0.77
Admission Source Other	0.15	0.04	0.26	1.17	1.04	1.30
Admission Source Ward	-0.13	-0.18	-0.08	0.88	0.84	0.92
Diagnosis Gastrointestinal	0.28	0.24	0.33	1.33	1.27	1.38
Diagnosis Neurological	0.31	0.27	0.35	1.37	1.31	1.42
Diagnosis Other	0.40	0.37	0.43	1.49	1.45	1.53
Diagnosis Trauma	0.52	0.47	0.58	1.69	1.59	1.79
Is Scheduled ICU Admission Yes	-0.22	-0.27	-0.17	0.80	0.77	0.85
Is Scheduled Surgery Yes	-0.16	-0.21	-0.11	0.85	0.81	0.90
Patient Category Surgical	0.01	-0.04	0.06	1.01	0.96	1.06
Physician Specialist Other	-0.01	-0.04	0.03	0.99	0.96	1.03
Physician Specialist Respirology	0.15	0.09	0.22	1.16	1.09	1.24
Physician Specialist Surgical	-0.08	-0.13	-0.03	0.92	0.88	0.97
Central Venous Line Yes	0.17	0.14	0.19	1.18	1.16	1.21
Arterial Line Yes	0.19	0.16	0.21	1.20	1.17	1.24
Intracranial Pressure Monitor Yes	0.55	0.48	0.63	1.74	1.61	1.87
Dialysis Yes	0.01	-0.04	0.07	1.01	0.96	1.07
Extra corporeal Membrane Oxygen Yes	0.67	0.49	0.85	1.95	1.63	2.33
Intra Aortic Balloon Pump Yes	0.33	0.25	0.40	1.39	1.29	1.50
Other Interventions in this Unit Yes	0.07	0.05	0.09	1.07	1.05	1.10
Interventions Outside this Unit Yes	0.02	0.00	0.04	1.02	1.00	1.04
Gender Male	0.00	-0.01	0.02	1.00	0.99	1.02
Age group2	0.12	0.08	0.16	1.13	1.08	1.18
Age group3	0.05	-0.01	0.12	1.05	0.99	1.12
Pre LOSC2	0.04	0.02	0.06	1.04	1.02	1.07
Pre LOSC3	0.15	0.12	0.18	1.16	1.12	1.19
MODS CatC2	0.11	0.07	0.14	1.11	1.08	1.15
MODS CatC3	0.22	0.19	0.26	1.25	1.21	1.29
MODS CatC4	0.27	0.23	0.31	1.31	1.26	1.36
MODS CatC5	0.18	0.14	0.21	1.19	1.15	1.24
NEMS CatC2	-0.01	-0.09	0.06	0.99	0.91	1.06
NEMS CatC3	0.04	-0.04	0.11	1.04	0.96	1.12

Table B.5: Variable selection criteria for models fitted to the data using backward elimination.

Variable removed	df	Deviance	Pr(>Chi)
Site Code	65	1267.19	2.80E-222
Admission Source	6	452.51	1.42E-94
Diagnosis	4	889.69	2.86E-191
IsScheduledICUAdmission	1	75.50	3.65E-18
IsScheduledSurgery	1	38.93	4.38E-10
PatientCategory (PC)	1	0.10	0.7498
PC + PhysicianSpecialist	4	44.34	5.46E-09
PC + CentralVenousLine	2	184.81	7.40E-41
PC + ArterialLine	2	193.28	1.07E-42
PC + IntracranialPressureMonitor	2	210.01	2.50E-46
PC + Dialysis (D)	2	0.40	0.8200
PC + D + ExtracorporealMembraneOxygen	3	51.27	4.28E-11
PC + D + IntraAorticBalloonPump	3	77.06	1.31E-16
PC + D + OtherInterventionsWithinthisUnit	3	42.75	2.78E-09
PC + D + Interventions Outside (IO)	3	3.06	0.3826
PC + D + IO + Gender	4	3.24	0.5179
PC + D + IO + Gender + Age_group	6	77.82	1.00E-14
PC + D + IO + Gender + Pre_LOS	6	100.58	1.90E-19
PC + D + IO + Gender + MODS	8	273.48	1.79E-54
PC + D + IO + Gender + NEMS	6	26.86	0.0002

Curriculum Vitae

Name:	Yawo Kobara
Post-Secondary Education and Degrees:	B.Sc. Actuarial Sciences (2011 - 2015) University for Development Studies, Tamale, Ghana M.Sc. Mathematical Sciences (2016 - 2017) Africa Institute of Mathematical Sciences, Mbour, Senegal M.Sc. Financial Engineering (2018 - 2020) WorldQuant University, New Orleans, LA, USA Ph.D. Statistics (2017 - 2022) Western University, London, ON, Canada
Honours and Awards:	Prudential Insurance OS Award (2015) Next Einstein Fellowship Award (2016) NSERC Postgraduate Scholarship (2017-2022) Rotman Post-Doctoral Opportunity Program Award (2022)
Related Work Experience:	Teaching Assistant (2015 - 2016) University for Development Studies SS2141 Instructor (Jan 2020 - April 2020) The University of Western Ontario Research and Teaching Assistant (2017 - 2022) The University of Western Ontario Contract Assistant Instructor (2021 - In progress) WorldQuant University

Publications:

Published Papers

- Akpan, I. J., Aguolu, O. G., Kobara, Y. M., Razavi, R., Akpan, A. A., and Shanker, M. (2021). **Association Between What People Learned About COVID-19 Using Web Searches and Their Behavior Toward Public Health Guidelines: Empirical Infodemiology Study.** Journal of medical Internet research, 23(9), e28975.

- Kobara, Y. M., Pehlivanoglu, C., and Okigbo, O. J. (2021). **A Linear Process Approach to Short-term Trading Using the VIX Index as a Sentiment Indicator** (Preprint at https://www.preprints.org/manuscript/202107.0673/v1).

- Kobara Yawo, Salifu Katara and Abdul-Rahaman Issahaku (2020). **Statistical Analysis and Vector Auto Regressive Model for Minimum and Maximum Temperature**, International journal of advanced research and publication, ISSN: 2456 9992.

Submitted Papers

- Yawo M. Kobara, Felipe F. Rodrigues, Camila E. P. de Souza, and David A. Stanford, **ICU Discharge Policy under Down Stream Congestion** (Resubmitted to INFOR: Information Systems and Operational Research on November 30, 2021).

- Akpan, I. Justice and Yawo Kobara, **An evaluation and science mapping of data science techniques and algorithms in women's health studies** (Submitted to Journal of Medical Internet Research)

- Yawo M. Kobara, Felipe F. Rodrigues, Camila E. P. de Souza, and David A. Stanford **ICU/SDU Queuing Game with Length-of-stay Decisions** (Submitted to European Journal of Operation Research on December 30, 2021).

- Yawo M. Kobara, Felipe F. Rodrigues, Camila E. P. de Souza, and David A. Stanford, Analysis of ICU Ventilator Use and Demand (Submitted to Journal of Critical Care, Elsevier on January 15, 2022)