Modeling and Simulation of a Multi-Hospital Intensive Care Network

by

Mina Moeini

B.Sc., Shahid Beheshti University, 2020

Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of
Master of Science

in the
Department of Mathematics
Faculty of Science

SIMON FRASER UNIVERSITY
Spring 2024

Declaration of Committee

Name: **Mina Moeini**

Degree: **Master of Science**

Thesis title: **Modeling and Simulation of a Multi-Hospital Intensive Care Network**

Committee: **Chair:** Tamon Stephen
Professor, Mathematics

JF Williams
Co-Supervisor
Associate Professor, Mathematics

Alexander (Sandy) Rutherford
Co-Supervisor
Adjunct Professor, Mathematics

Ailene MacPherson
Committee Member
Assistant Professor, Mathematics

Jessica Stockdale
Examiner
Assistant Professor, Mathematics

Abstract

The Intensive Care Unit (ICU) represents a specialized sector within a hospital setting, committed to offering advanced monitoring, precise therapeutic interventions, and highly specialized nursing care for patients facing critical medical conditions that demand rigorous medical supervision and support for essential physiological functions. Nonetheless, ICUs are often constrained by a finite availability of critical resources, such as beds, specialized nursing staff, and ventilatory equipment, among others. In this thesis, I construct a simulation model to analyze the operational dynamics of a network comprising eight major ICUs in British Columbia, Canada. The focus of this thesis is development and validation of a robust discrete-event simulation model designed to estimate patient flow through individual sections of the critical care system across multiple healthcare facilities. The model includes various strategies for admitting new patients when an ICU reaches full capacity, such as utilizing overflow beds, bumping patients, or transferring patients to other hospitals. The simulation model was calibrated using real world data from the British Columbia Critical Care Database and serves as an analytical tool for planning critical care capacity in the context of endemics and pandemics such as COVID-19. This work was done in collaboration with the Ministry of Health in British Columbia.

Keywords: Simulation Modelling ; Critical care model; AnyLogic Software; Intensive Care Unit(ICU)

Acknowledgements

This research was conducted on the ancestral and unceded territory of the Musqueam, Squamish, and Tsleil-Waututh Nations.

The research topic in this thesis was defined in collaboration with British Columbia Ministry of Health in their ongoing efforts to improve healthcare. I thank and acknowledge

- Dr. Donald Griesdale serves as the Associate Director of the Centre for Clinical Epidemiology and Evaluation and is an intensivist at Vancouver General Hospital (VGH).
- Dr. Peter Dodek, MD, MHSc, is an Intensivist and serves as the Chair of the Critical Care Working Group at the Center for Health Evaluation and Outcome Sciences at St. Paul's Hospital (SPH), Vancouver. Additionally, he holds the position of Professor of Medicine at the University of British Columbia (UBC).
- Steve Ahkioon, Expertise on the database structure was contributed by the Critical Care Regional Advisor at Vancouver Coastal Health.
- Mike Irvine, Insightful direction for the model's development, specifically within the scope of pandemic planning, was contributed by the Senior Scientist in Data and Analytics Services at the British Columbia Centre for Disease Control (BCCDC).
- Alexander Rutherford and Samantha Zimmerman, both members of the Centre for Operations Research and Decision Science (CORDS) team, undertook the valuable initiative of participating in shadowing ICU rounds at Vancouver General Hospital (VGH) to gain a comprehensive understanding of ICU protocols.
- The valuable feedback provided by the Critical Care Services Executive Committee in 2021 and 2023 greatly contributed to the significant improvement of the model.

Partial funding of this research was provided by Health System Improvement Grant (MoH) and Canadian Network for Modelling Infection Disease Grant (CANMOD).

I thank my supervisors and thesis committee for their feedback and supervision.

I thank my parents, sister, and partner for their personal support.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 The Essential Role of Intensive Care Units

The Intensive Care Unit (ICU) is a specialized section of the hospital designed for the treatment of patients facing severe or life-threatening conditions. This facility has expert staff and advanced technology to provide continuous 24/7 care focused on stabilizing patients for transition to less critical environments. The ICU plays a crucial role in the hospital for various reasons and serves a broad range of patients, from those with pre-existing conditions to those with unexpected injuries or illness, as well as those who need support before and after complex procedures. Nowadays with rapid growth of the aging population, ICU demand has increased at a rate 62% faster than the availability of beds [52]. Patients can be admitted to the ICU from different hospital wards, after surgery, directly from the emergency department, or from ICU of other hospitals due to medical reasons [7] which require a significant allocation of resources and staff. Their limited bed capacity means that when demand exceeds supply, it can quickly lead to bottlenecks. Therefore, ICUs can function as a bottleneck in hospital patient flow, where limited capacity can affect admissions and impact both upstream and downstream departments [109] (Figure 1.1).

ICUs are widely recognized as one of the most financially demanding healthcare resources within a hospital, typically utilizing over 13% of the hospital's total budget [49]. Providing medical care to patients in an ICU is more costly compared to treatment in a standard hospital ward. Data from the United States and Canada indicates that ICU-related expenses can be four to six times higher than those of a regular ward; ward beds cost as much as $\$1,000$ per day while critical care beds surpass $\$3,500$ per day [94, 45, 88, 49]. Therefore, the ICU plays a crucial role in hospital healthcare, both in terms of its critical importance and its substantial financial impact.

How did COVID-19 affect the ICUs?

The recent outbreak of SARS-CoV-2 (COVID-19) has once again highlighted the crucial role of ICUs in healthcare systems. High COVID-19 transmission rates, coupled with a significant number of patients requiring medical care, have put an extraordinary strain on healthcare providers [21]. This has led to a demand for resources, such as ICU beds and ventilators, that often exceeds available capacity, potentially resulting in increased patient mortality [117]. This limitation is a common issue in many hospitals across the UK, USA, and Europe [38, 96]. For example, Italy encountered severe pressure on its critical care system. Patients struggled to access ICU beds, and some tragically died in hospital corridors while waiting for beds to become available. During COVID-19 pandemic, 9% to 11% of COVID-19 hospitalizations necessitate ICU care in a year, impacting the availability of ICU beds, ventilators, and specialized nursing staff [96]. Also, a healthcare system's approach to managing COVID-19 has noteworthy effects on other categories of patients, such as those needing routine emergency or elective care. Many operations cancelled or delayed due to disruptions from COVID-19; therefore, the impact on these populations should not be disregarded [21, 36, **?**].

Beside COVID-19, seasonal influenza also constrains ICU resources. Influenza, characterized by high severity, extensive spread, and long duration typically causes self-limiting respiratory illnesses, with $5\% - 10\%$ of hospitalized cases needing ICU admission due to severe symptoms [9, 21]. Intensivists are vital for triage during pandemics when demand surpasses capacity, as they primarily care for severely ill patients who often arrive first in emergency departments [51, 61]. Additionally, trauma patients also can place considerable stress on ICU resources and staff due to the intensive care they require. Their conditions often demand immediate, complex interventions, and a multidisciplinary approach to manage life-threatening injuries. The high demand for specialized equipment, such as ventilators and advanced monitoring systems, alongside the necessity for a higher ratio of medical staff to patients, strains the operational capacity of the unit. Therefore trauma patients can similarly strain critical systems. Different patients exert varying levels of stress on ICUs, each presenting distinct challenges and requiring unique approaches for management and mitigation.

Governments are increasingly utilizing modeling to identify the most effective approaches to inform pandemic planning. These models fall into two different categories: epidemic models and operational models. For example, early in the COVID-19 pandemic special emphasis was placed on strategies to flatten the growth curve of this disease, in order to relieve stress on healthcare systems. There are significant numbers of epidemiological models which are ideal for predicting the number of new cases or for identifying the best measures to reduce transmission. Extensive epidemiological modeling has simulated pandemic spread, highlighting its potential in offering solutions to numerous challenges [31]. However, such models do not directly assist in managing ICU beds, ventilator use, medical staff allocation, or understanding the impact of individual behavior on healthcare system capacity [31, 69]. Operational models on the other hand, are valuable tools for decision-making at various levels, from national governments to local municipalities and individual hospitals to help in planning capacity and managing limited resources efficiently for the care of critically ill patients.

In scenarios where ICUs reach their capacity, critical care staff implement several strategies to accommodate new admissions. These include utilizing additional, or overflow beds located in other sections of the hospital [15, 70] or doctors might discharge current patients sooner when high bed occupancy risks the care quality for new patients. This involves the early discharge of patients who are nearing the completion of their ICU stay, relocating them to High Acuity Units (HAUs) or medical wards [70]. High acuity units deliver essential critical care to patients whose require more then can be accommodated by general medical wards but fall short of necessitating the comprehensive care provided in ICUs. Patients may also be transferred to ICUs at alternative hospitals may also be transferred. These strategies are crucial in managing the dynamic and high-demand environment of ICUs, ensuring optimal utilization of available resources and sustained provision of critical care [15, 117].

To effectively optimize ICU function, several important limitations of current simulations must be addressed. Existing models are designed to function for a single ICU and preventing analysis of all inter-ICU transfer [116, 117]. Hence there is a would be significant benefit in expanding these models to facilitate collaboration between multiple ICUs, potentially on both a provincial and national scale [102]. Particularly in the context of emerging diseases, fostering collaboration across all hospitals is pivotal and such collaboration is essential to mitigating impacts on the healthcare system, as exemplified by the coordinated response required during the pandemic. By developing models that incorporate multiple ICUs, hospitals can enhance their preparedness and response strategies for future healthcare challenges. A second key limitation of current simulation models is that existing studies have typically considered only one or two resource pools, such as mechanical ventilation or beds, in their ICU models [117, 116, 49].

This thesis addresses these limitations, broadening the scope of current simulations to include both multiple ICUs and multiple resource pools. I developed a simulation model designed to represent the network of eight major ICUs in British Columbia, Canada. The model is designed to support the management of critical systems during endemic situations and seasonal pandemics by coordinating the operations of various units and utilizing a comprehensive, detailed database relevant to British Columbia.The model is enriched with advanced features including, but not limited to, varying reasons for patient transfers, diverse patient types, and distinctive time distributions for each event occurring within the model. These features allow for a detailed examination of various patient admission strategies, enabling a thorough evaluation of their effects on the ICU network.

The simulation aims to offer insights into the dynamic interplay between different components of the critical care network, allowing for an improved understanding of optimal operational strategies and management practices, especially during periods of heightened demand such as pandemics. This work is motivated by the critical role that ICUs play in healthcare systems and seeks to contribute to the ongoing efforts to enhance their resilience and effectiveness in facing contemporary healthcare challenges. To ensure the privacy and confidentiality of the institutions involved, I am unable to reveal the names of the hospitals in this study. Instead, I have employed numerical labels, ranging from one to eight, to represent each of them.

1.2 Organizing and Overseeing Critical Care

In this section, I explore the structure of the BC critical care network in more detail and how the complexities of this network are captured in my simulations (Figure 1.1). The ability to provide critical care in British Columbia is mainly restricted by the number of healthcare resources available which includes ICU nurses and beds, and mechanical ventilation. During the pandemic, the number of critical care beds was increased in healthcare settings [34]. However, this expansion faced significant challenges in securing sufficient staffing for the added capacity. ICU management is challenging due to its unpredictable nature and numerous uncertainties. For instance, predicting patient arrival patterns is difficult as patients can be admitted either directly to the ICU, following a scheduled surgery, or via the emergency department, sometimes with an operating room stop [7]. Hence I begin this section by detailing, the admission process within the critical care system.

Patient pathways bridge the ICU with hospital units internally and externally, affecting decisions for upstream patients from the Emergency Department (ED) due to injuries or post-surgery (OR admission), and downstream for those transferred to the ICU from wards or the HAU due to severe conditions (ICU readmission), in Figure 1.1 you can see the arrival of patients [7]. Initial outbreaks and observations in affected countries revealed a marked rise in hospital and critical care needs for pandemic patients. Patients are divided into pandemic and non-pandemic groups for hospital admission [15].

In this study, the simulation model narrows its focus to four particular patient categories depending on their diagnosis:

- Viral Pneumonia (VP) or Acute Respiratory Distress Syndrome (ARDS)
- medical patients without VP or ARDS
- surgical patients without VP or ARDS
- trauma patients without VP or ARDS

Patient category has important implications for both the patient's length of stay (LOS) and the resources required, e.g., ventilation.

To initiate the ICU admission process, a patient needs a bed in the ICU and when the ICU reaches its capacity, critical care staff employ diverse strategies for managing patient admissions. These include utilizing overflow beds in other parts of the hospital, moving patients who are near the end of their ICU stay to HAUs or medical wards, or transferring patients to ICUs at different hospitals. This simulation model incorporates ICUs and HAUs to provide a more comprehensive understanding of healthcare dynamics under crisis conditions [15, 95, 107, 75]. These strategies play a key role in optimizing healthcare resources and patient care. In ICUs and HAUs, there are three pivotal initiating resources available qualified ICU nursing staff, beds, and mechanical ventilators. These resources are illustrated in the Figure 1.1.

The model considers three types of transfers. (Figure 1.1)

- When a patient arrives at the hospital needing an ICU bed, but none are available (ICU and overflow bed), the patient may be transferred to another hospital with an available ICU bed. This transfer is referred to as a “capacity reason transfer”.
- When critically ill patients can’t get necessary treatment at their original hospital, they need to be transferred to another ICU for proper care. This transfer is termed “medical reason transfers”.
- When patients transfer back to their home healthcare facility, usually after they have received specialized care or treatment in a different ICU and patients recovery or receive ongoing care in an ICU closer to home. This transfer is called repatriation transfer.

After discharge from the ICU, strategic relocation of patients occurs to various sections of the hospital to ensure smooth accommodation for new admissions.

Finally, I am going to discuss ways of measuring performance of the simulation model. The simulation model captures key performance indicators (KPIs) for resource utilization and patient outcomes. The main KPI for resource utilization is the expected number of occupied ICU and overflow beds at each ICU, also model captured the ICU’s high demand when occupancy surpasses 90% capacity, a level that adversely affects admissions and discharges, and is not ideal for the hospital. In general, higher occupancy of ICUs leads to higher mortality rates [47, 110]. In evaluating patient outcomes, the simulation model quantifies specific aspects within each ICU such as the proportion of patients unable to obtain a ventilator when necessary, the proportion of patients moved to the HAU due to the ICU reaching its maximum capacity, and the proportion of patients who are transferred to other facilities or units for different reasons. The model also addresses mortality resulting from different factors, such as when patients can not receive bed or mechanical ventilation (Figure 1.1). This simulation model tests various scenarios in critical care to identify the most effective response to high hospital arrival rates.

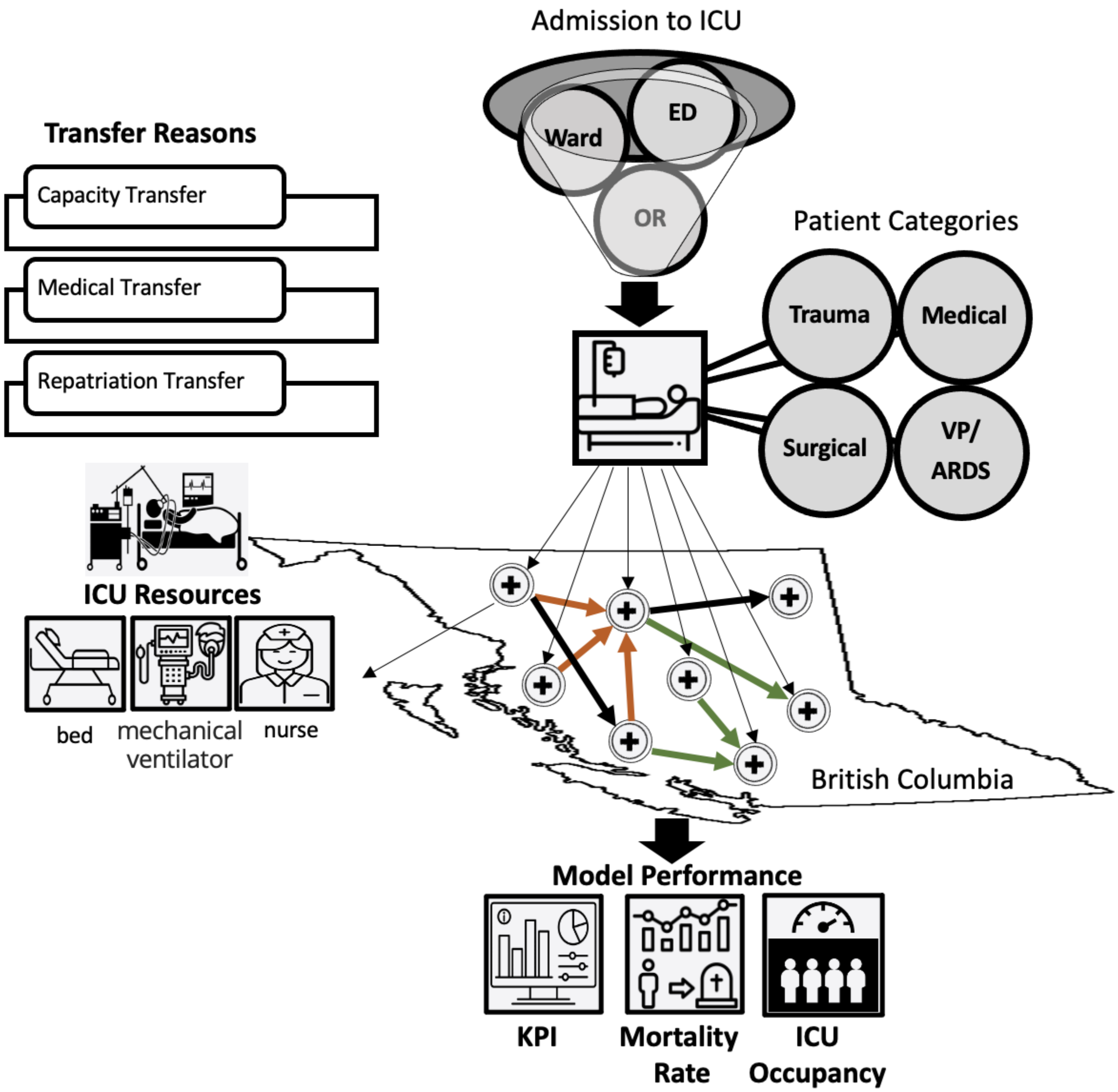


Figure 1.1: Functioning of individual ICUs are primarily determined by three key resources: beds, mechanical ventilators, and nurses. Patients may be admitted into the ICU from other departments such as operating room (OR), Emergency Department (ED), or Wards for a variety of reasons with different patient categories. Transfers of patients among ICUs within the hospital network as a result of capacity limitations, medical reasons, or repatriation. The functioning of the ICU can be assessed using various metrics, including key performance indicators (KPIs) capturing, for example, patient care and patient flow, mortality rates attributed to various causes, and ICU occupancy.

Abbreviation	Meaning
ICU	Intensive Care Unit
HAU	High Acuity Unit
HLC	High Level of Care
OR	Operating Room
KPI	Key Performance Indicator
ED	Emergency Department
DES	Discrete Event Simulation
VP	Viral Pneumonia
ARDS	Acute Respiratory Distress Syndrome

Table 1.1: List of abbreviation

1.3 Essential Concepts in Simulation Modeling and Discrete Event Simulation

The complexity of healthcare systems makes modeling them difficult, arising from the interplay of diverse participants (patients, providers, insurers, policymakers), unpredictable human behaviors, and a variety of health conditions. Coupled with the swift advancements in technology and medicine and the ever-changing healthcare policies, these factors contribute to a complex environment. This complexity necessitates sophisticated modeling approaches capable of adapting to and accurately capturing the nuanced aspects of healthcare delivery. [108, 13]. As healthcare systems evolve, the global challenge is to enhance care quality while minimizing costs. Consequently, strategic, tactical, and operational decisions are routinely made to assess and enhance the efficiency of healthcare services. To predict the impact of these decisions on system performance, healthcare providers require tools like simulation to effectively explore alternative scenarios.

Simulations replicate real-world systems over time, identifying key issues and bottlenecks, and enable the exploration of "what-if" scenarios without real-world limitations [63, 5, 35]. Simulations can evaluate healthcare interventions, incorporating behavioral factors and personal choices, to identify the best scenario based on specific criteria. A simulation study begins with developing a conceptual model—a theoretical or observed representation of a system issue, covering objectives, inputs, outputs, content, boundaries, assumptions, and simplifications. This model is then converted into computer software, helping healthcare professionals understand the relationship between input and output variables in the system [14, 92].

Simulations can estimate the performance of time-dependent queue models by sampling stochastic processes for various events (arrivals, service starts, and departures) using pseudorandom numbers. Discrete Event Simulation (DES) tracks system state changes after each event, creating a potential sequence of outcomes. In contrast, Discrete Time Simulation (DTS) updates the system state at fixed intervals. By aggregating results from multiple

Characteristics of discrete event simulation	
Scope:	Operational, tactical
Purpose:	Decision making like optimization, predictions, and comparison
Prospective:	Analytic, focus on detail complexity
Variability Importance:	High
Tracking individuals importance:	High
Number of Entities:	Large
Control:	Queues
Relative Timescale:	Short
Clarification of Model:	Distinct entities, attributes, decisions, and events
Data Sources:	Numeric data with some important elements
Model Results:	Prediction points, performance measurements, KPI
Software:	Simul8, FlexSim, Simio, AnyLogic, TreeAge

Table 1.2: Overview of Discrete-Event Simulation characteristics.

simulations, both methods can generate estimates for KPIs. DES generally provides more precise performance estimates than DTS, but DTS's efficiency can be enhanced by optimizing the time-step parameter. Both approaches effectively model complex queuing systems with their respective advantages [79, 93, 46], in Table 1.2 you can see the summary of the DES characteristics.

In healthcare, modeling and simulations can support clinical decisions, predict bed occupancy, help allocate resources like staffing, evaluate treatments, guide emergency room redesign organization, and test ICU information system usability. The simulation does not evaluate treatments, humans do that guided by the results of simulations. [3, 27, 29, 44].

1.4 Data Insights and User Perspectives

A Comprehensive Overview of Data Sources in This Thesis

The main source of data for this thesis comes from an anonymized extract of the British Columbia Critical Care Database, covering the years 2010 to 2020. This includes data on the patient treatment in the ICU, time of patient admission and discharge, patient transfers reasons, the hospital the patient was transferred to, start and stop times for all mechanical ventilation instances, and diagnostic information on the patient. This database was initiated in 1997 as the British Columbia ICU Database. To enhance the completeness of the Critical Care Database for some hospitals, I linked the Critical Care Database to the Discharge Abstract Database (DAD) for British Columbia. The DAD offers a comprehensive overview of all hospital admissions, it is notably less detailed when it comes to the specialized care provided in ICUs and HAUs. This study specifically targets general ICUs, and purposefully omits specialized ICUs such as those focused on cardiac surgery. The COVID-19 pandemic exerted significant strain on the critical care system, compromising data completeness and complicating the interpretation of data due to frequently fluctuating temporary capacities.

Applications and Impacts of Critical Care Modeling

The critical care system was already functioning at or above its designated capacity before the onset of COVID-19 pandemic, in the province of British Columbia. This pre-existing condition posed significant challenges to effectively managing and addressing sudden surges in demand for critical care services throughout the pandemic. Critical patients require immediate ICU admission and intensive monitoring due to the severe consequences of delayed treatment. Hospital capacities vary, leading to potential transfers if the necessary care is unavailable or if ICU beds are fully occupied. Through consultations with critical care experts in British Columbia, I identified the crucial need to integrate different transfer reasons between hospitals into the model. This adjustment was critical for improving the model's preparedness for future emergencies, such as COVID-19. Strategies include temporarily reallocating ICU nurses and medical staff from across the network to address critical care surges, and transferring stable patients from the ICU to other hospital sections to optimize patient admission and minimize rejections. This experience underscored the critical necessity of proactive operational planning to effectively anticipate and address surges in demand for critical care services. Utilizing the simulation model of the critical care network in British Columbia, I intend to provide support for planning across a spectrum of diverse scenarios currently being developed by the British Columbia Health ministry and they may employ this model to identify early warning indicators for a novel outbreak. These indicators will serve as triggers, prompting the initiation of preparations for the implementation of contingency measures.

This thesis is structured as follows. Chapter 2 gives background on operations research in ICU and application operation research during the COVID-19 pandemic, queueing models, and KPI estimation within queuing models. Chapter 3 outlines ICU resources, their allocation, and detailed patient arrival patterns, including the introduction of various patient transfer reasons and the model. Chapter 4 detail the analysis of ICU service time distributions and their application in modeling. Chapter 5 presents the simulation validation and results, which are discussed in more details in Chapter 6 along with limitations and future improvements broader conclusion.

Chapter 2

Literature Review

Modelling and simulation find widespread application in diverse scientific fields like ecology, social sciences, economics, healthcare, and engineering [67]. Modeling enables understanding of real system behavior, followed by simulation testing. Simulations explore multiple scenarios, explaining real system behavior and assess strategies for optimal system operation. Simulation results assess model quality and offer insights for enhancing accuracy. Models identify system deficiencies and forecast intervention impacts without disrupting complex system operations [27]. In recent decades, modeling and simulation have made notable advancements in healthcare operations research, addressing areas like emergency department efficiency, operating theatre scheduling, bed management, and waiting list optimization.

For example, various studies have demonstrated the utility of simulation and modeling techniques in improving healthcare delivery and operational efficiency. M'Hallah et al. [87] utilized a simulation model to optimize surgical scheduling in a hospital operating room (OR), suggesting the cancellation of surgeries post-OR hours to improve resource utilization. This approach directly addresses the inefficiencies in surgical scheduling, offering a practical solution to maximize the use of available resources. Similarly, Wang et al. [111] developed a model for emergency services aimed at enhancing doctor efficiency and introducing a fast-track system. This innovation effectively reduced waiting times for patients, demonstrating the potential of simulation models to streamline emergency department operations and improve patient flow. In the optimization of nursing, Legrain et al. [68] examined nurse scheduling at large hospitals. They introduced a multi-objective optimization model and local search techniques, which surpassed traditional scheduling methods without incurring extra costs. Their study highlights the significance of advanced scheduling techniques in managing nursing staff more efficiently, ultimately leading to improved patient care and staff satisfaction.

Furthermore, Harper et al. [50] explored various appointment scheduling strategies in an outpatient department. They identified key factors capable of reducing patient wait times and queue sizes without necessitating additional resources. This research underscores the impact of strategic scheduling on enhancing patient experience and operational efficiency in outpatient settings. These examples collectively underline the critical role of simulation and optimization models in healthcare settings. By employing these tools, healthcare facilities can significantly improve their service delivery, operational efficiency, and resource utilization, ultimately benefiting both patients and healthcare providers.

2.1 Operation Research in ICU

In Canada, the demand for ICUs is escalating more rapidly than that for general acute care hospitalizations. Based on information from the Canadian Institute for Health, the period of 2013–2014 saw over 230,800 adult ICU admissions, marking a 12% increase since 2007–2008, while adult hospital admissions rose by 7% during the same interval. Notably, 80% of ICU admissions were emergent, affecting patient outcomes, the use of resources, and the planning of capacity. Predominantly situated in large or teaching hospitals, it is a significant challenge to keep pace with the demand for ICU beds. These ICUs typically operate at approximately 90% capacity, experiencing overcapacity for up to 51 days annually in 2013–2014.

Additionally, there is a rising trend of ICU patients requiring specialized and resource-intensive care, such as invasive ventilation, which was necessary for 33% of patients in 2013-2014, an increase from 28% in 2007-2008. This upsurge, particularly in short-term invasive ventilation, suggests a potential for further pressure on ICU capacities. ICU capacity presents an ongoing and future challenge for Canada's health system, exacerbated by an aging population and the potential for more severe illnesses among hospital patients. The scarcity of ICU capacity and wrong decision-making can pose serious risks to patient safety. Issues such as ICU overcrowding have serious consequences, including higher morbidity and mortality rates, staff burnout, diminished revenue, and bottlenecks in hospital-wide patient flow. In response to these challenges, hospital administrators are diligently seeking strategies to deliver high-quality care within the constraints of ICU capacity. Bai et al. [7] performed an extensive review of how operational research supports ICU management, underscoring the ICU's central role in the hospital's patient flow. This review methodically sorted the existing literature, examining decision-making horizons, settings, and an array of modeling and solution strategies.

To aid in decision-making at policy and operational levels, detailed mathematical models offer valuable quantitative insights into evaluating critical care capacity strategies. Extending the focus beyond critical care, several researchers have delved into hospital bed allocation and capacity planning. Williams [112], for example, developed a model aiding a physician group in determining the optimal bed count for a hospital ICU expansion. The simulation showed that increasing beds from 7 to 15 led to more empty beds and fewer premature discharges, with changes in bed count affecting outcomes differently. Ultimately, the decision was made to maintain the ICU at 11 beds. This model facilitated decision-making by quantifying the impact of various options, resulting in an ICU that effectively met patient needs amid fluctuating service demand.

Ridge et al. [97] introduced a simulation model for bed capacity planning, revealing a non-linear connection between bed numbers, average occupancy, and patient transfers due to bed shortages. It highlighted a significant trade-off between occupancy rates and transfers, utilizing a basic deferral rule for elective patients when admissions are blocked by bed unavailability. This suggests the potential for improving elective patient admission scheduling systems. Mallor et al. [75] introduced a model for ICU bed occupancy analysis, incorporating generalized regression models to better capture patient stay variability. This analysis underscored the importance of including clinical staff decisions in model accuracy, leading to the creation of a mathematical model for these decisions. Furthermore, a combined optimization and simulation approach was developed for more precise parameter estimation. Zhu et al. [116] developed a discrete event simulation (DES) model to assist in determining optimal ICU bed capacity, aiming to balance high-quality service with cost-efficiency. This model tackled the challenges of ambulance diversion and surgery cancellations due to bed shortages, as well as the inefficiencies caused by surplus beds. Griffiths et al. [45] crafted a detailed simulation model for a large teaching hospital's ICU, which, despite being typically equipped with 14 beds, could accommodate additional beds and specialized equipment during peak demand periods. The model provided a precise reflection of patient admission times and durations, based on their sources of admission, and offered a strategic staffing plan to reduce nursing costs.

Various strategies are employed by critical care staff to manage capacity, such as using overflow beds, discharging nearly recovered patients to HAUs or medical wards, or transferring patients to other hospitals' ICUs. Litvak et al. [70] proposed a cooperative strategy focusing on regional bed, overflow bed, allocation to minimize patient refusal rates, where hospitals in a region reserve beds specifically for regional emergencies. This approach, inspired by telecommunication systems with overflow capabilities, utilized mathematical methods to determine the required number of regional beds to achieve desired acceptance rates, showing that inter-hospital cooperation can enhance service quality with fewer beds. Steins and Walther [107] developed a generic simulation model for ICU capacity planning, designed to accurately predict occupancy and immediate admission rates, incorporating data on overflow beds and patient transfers. This model, based on real admission data from four ICUs, proved useful in forecasting capacity needs. Reader et al. [95] explored the concept of "bumping" which is discharging patients nearing the end of their stay to make room for new admissions. Through interviews, they examined the decision-making processes of ICU physicians regarding admissions and bumping, revealing variations and implications for patient safety.

Dobson et al. [26] introduced a stochastic model focusing on patient bumping to aid in forecasting such events, utilizing a Markov chain model and an innovative algorithm to manage the model's complexity. Rodrigues et al. [99] created a DES model to forecast step-down bed requirements for a large university hospital. This model stands out by encompassing the hospital's entire inpatient flow, with a focus on the ICU's daily stochastic flows, utilizing nursing workload scoring metrics. Utilizing data from a significant academic hospital, the model demonstrates how step down beds enhance patient flow and reduce costs. Mitra et al. [86] studies the effects of creation a new HAU on patient outcomes in a major tertiary-care hospital in Canada, hypothesizing that the HAU would enable quicker access to critical care and facilitate convalescent care, leading to reduced in-hospital mortality and shorter ICU and hospital LOS. Their results confirmed that the HAU's introduction was linked to a lower risk of in-hospital mortality, shorter LOS in both the ICU and hospital.

2.2 Operation Research in Pandemic Responce

The COVID-19 pandemic has exerted immense pressure on healthcare systems worldwide, particularly impacting ICU occupancy and patient management. The surge in critically ill patients has strained ICU resources, leading to challenges in maintaining quality care and managing bed availability. Therefore, the rapid onset of the COVID-19 pandemic highlighted the value of integrating mathematical modeling with critical care system simulations. By utilizing mathematical models and simulations, healthcare professionals can predict ICU occupancy trends, assess resource needs, and strategize patient allocation more effectively. These models allow for scenario planning, helping hospitals anticipate surges in demand and optimize resource allocation accordingly. Moreover, simulations enable healthcare providers to test different intervention strategies, evaluate their efficacy, and refine protocols to enhance patient outcomes.

Currie et al. [21] highlighted the challenges posed by the COVID-19 pandemic and proposed simulation modeling as a tool for decision-makers to make informed early-stage decisions. The article serves as a mobilizing call for modelers and a guide for decision-makers on leveraging support from the simulation community. Meares and Jones [81] responded to the exponential increase in cases and ICU demand by developing a queueing theory model to estimate Australia's ICU bed needs during an ongoing pandemic, using the situation in late March 2020 and data from Lombardy for comparison. Bekker et al. [8] developed models to forecast hospital admissions and bed occupancy by COVID-19 patients in the Netherlands, aiding in both immediate patient transfer decisions and long-term policy formulation. They introduced a novel linear programming technique for admission forecasting and applied residual stay lengths and queueing theory to estimate bed occupancy. These models enhanced prediction accuracy and trust, facilitating effective pandemic management by optimizing bed use and preserving capacity for other care types.

Garcia-Vicuña et al. [32] developed a discrete event simulation model for enhancing ICU bed planning during outbreaks, notably the COVID-19 pandemic. This model, aimed at short-term forecasting, necessitates precise depiction of the current healthcare system and accurate simulation of its dynamics, primarily focusing on stochastic modeling of patient admissions and flows. Implemented daily, it provided essential forecasts to the regional healthcare logistics planning team, who then strategically allocated ward and ICU beds based on these predictions. Lu et al. [74] introduced a hybrid simulation approach combining a system dynamics model for COVID-19 case prediction and a discrete-event simulation for evaluating and allocating hospital beds. They suggested two policies—type-dependent admission control and early step-down, based on patient risk profiles—to decrease mortality among intensive care patients. The strategy focused on optimizing bed allocation for low-risk and high-risk patients to minimize death rates and maximize recovery outcomes. Baas et al. [6] developed a model featuring a network of two infinite server queues for multiple patient types, offering real-time forecasts of COVID-19 patient admissions to wards and ICUs. This model predicts patient inflow, length of stay in wards and ICUs, and inter-department transfers, utilizing data from the hospital's data warehouse. Tested against data from the Netherlands' first COVID-19 peak, the algorithm demonstrated high accuracy in its predictions.

During the COVID-19, mechanical ventilation emerged as a life-saving intervention for patients with severe respiratory complications, particularly those suffering from acute respiratory distress syndrome (ARDS). Modeling and simulation became instrumental in predicting the demand for mechanical ventilators. These predictive models, based on various epidemiological and healthcare data, helped healthcare systems and governments anticipate the needs for ventilators, optimize their allocation, and manage supply chains effectively. In pre-COVID-19 ventilator allocation studies, Zaza et al. [115] proposed a conceptual framework for public emergency ventilator allocation, while Meltzer et al. [85] estimated the mechanical ventilator demand during a US influenza pandemic. They project need for 35,000-60,500 additional ventilators in severe scenarios to prevent 178,000-308,000 deaths. Huang et al. [53] present a method for optimizing stockpiles of mechanical ventilators crucial for treating hospitalized influenza patients with respiratory failure. The studied examines mild, moderate, and severe pandemic scenarios in Texas, prioritizing local over central storage.

Huang et al. [53] do not address ventilator distribution over time but focused on minimizing expected ventilator shortfall and total stockpiling cost. With COVID-19, ventilator demand increased over time, posing challenges as only future demand forecasts are available. Mehrotra et al. [84] proposed a a stochastic optimization model for managing stochastic demand for life-saving resources, such as mechanical ventilators during crises like the COVID-19 pandemic. States exhibit varying peak demands, and each state's reluctance to share surplus resources due to risk aversion is captured by a safety threshold parameter. Simulations of their model, based on realistic forecasts and inventory availability, underscore the critical importance of ventilator allocation for COVID-19 patients. With over 40%

of existing inventory allocated to COVID-19 cases, national demand can be met in mild scenarios, but with less than 25%, shortages may occur, particularly in extreme scenarios. Zimmerman et al. [117] introduced a multi-class Erlang loss model to simulate ventilator demand from COVID-19 and non-COVID-19 patients, incorporating COVID-19 case projections influenced by public health measures and social distancing. They used the BC Intensive Care Unit Database for model calibration and validation. Through discrete event simulation, they forecasted ventilator capacity and patient access, comparing simulations to point-wise stationary approximation, modified offered load, and fixed point approximation methods to assess ventilator allocation efficiency.

2.3 Queueing Models

The literature on queueing models, which are extensively applied in healthcare [54] and ICU modeling, is reviewed in this section. Queueing models are stochastic processes that simulate entities—like customers or patients—as they request, wait for, and receive services [83]. In this thesis, model clients or modeled clients denote these simulated entities within queueing models, distinct from actual clients in the real world. The term modeled service similarly signifies the representation of services within these models. The service in a queueing model can vary widely, from healthcare delivery by doctors to interactions with call center staff or using a telephone line [43, 10, 30]. The key feature of this model is the service time (the time taken to complete a service), modeled as a non-negative continuous random variable. Model servers, representing staff or resources, are considered busy while providing services. Service requests are modeled as arrivals, using a stochastic process to track the cumulative arrivals over time [83]. If all servers are busy, incoming clients join a queue until a server is free. A client then receives service for a random duration before exiting the system. Figure 2.1 illustrates this basic queueing system, highlighting the processes of arrivals, queueing, service provision, and departures.

Background on Queuing Model

In a queueing model, the key elements are the arrival process, service time distribution, and the number of servers. A model is termed homogeneous when the arrival and service rates, along with the server count, remain constant. Models with service times and arrivals following exponential and Poisson distributions, respectively, exhibit the Markovian property, indicating future states depend solely on the current state, excluding historical data [82]. Non-exponential service times or non-Poisson arrivals are classified as general distributions. Phase-type distributions are defined by combining exponential distributions either sequentially or in parallel, offering a versatile framework for modeling [16].

For any set of queueing model parameters, time-dependent probability distributions describe key performance metrics such as client count in the system, queue length, and waiting

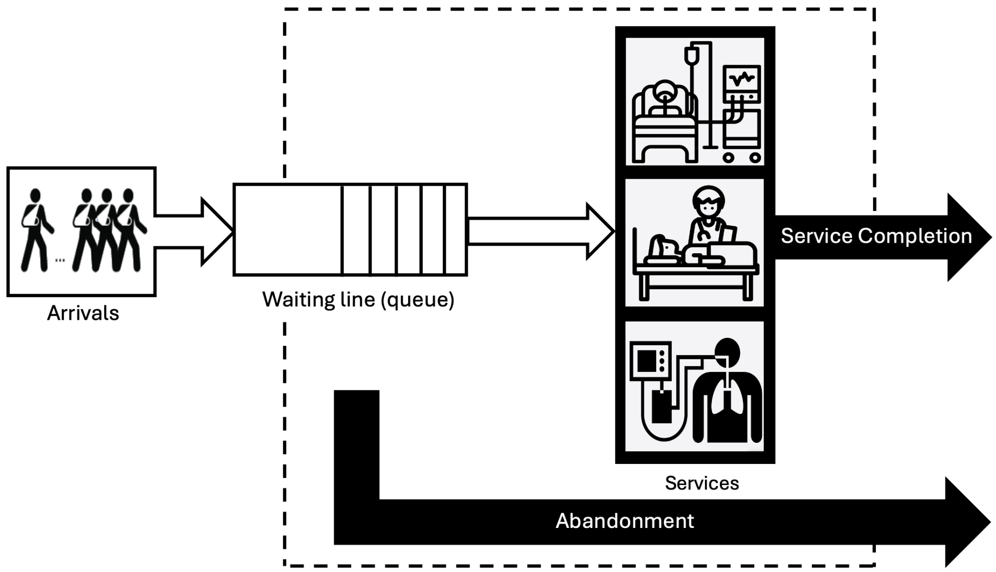


Figure 2.1: Queuing diagram illustrating a single-service queueing model where clients arrive for service and join a queue when all servers are busy. They receive service as soon as a server becomes available, and leave the system after their service is completed.

times. Under specific conditions, these systems can reach stochastic equilibrium, stabilizing these distributions over time [83]. Such stable distributions are termed steady-state or stationary, while behaviors outside this equilibrium are known as transient or time-dependent. While some queueing models allow for precise analytical solutions for steady-state distributions, more complex models necessitate approximation methods for these values.

Some queueing models account for client abandonment, where clients exit the queue without service if the wait is too long [77]. These models vary; in some, clients may rejoin the queue later, while in others, they exit the system permanently [77]. This phenomenon is modeled using a random variable for clients' waiting tolerance, known as patience, with a specific distribution [77]. Such stochastic outcomes, including abandonment and prolonged waits, serve as performance metrics to approximate real-world service quality and client accessibility [78].

The number of servers is a crucial parameter in queueing models. Infinite server models assume unlimited capacity, providing immediate service to all arriving clients [28]. Conversely, finite server models account for capacity constraints, leading to potential client waiting and abandonment. These models may also adjust the number of servers over time to reflect changes in staff availability or infrastructure, simulating real-world fluctuations. In scenarios of reduced staffing, servers might operate under an exhaustive service policy, staying active to complete services for current clients [23, 24, 55]. Alternatively, a non-exhaustive or preemptive policy may apply, where interrupted services lead clients to requeue for completion with an available server [56].

The queue discipline, another key feature of queueing models, dictates the service order for clients. The first-come-first-served (FCFS) principle services clients in their arrival order. Some models, however, employ a service order based on randomness or a last-come-first-served approach [82]. Models with priority classes serve clients based on their priority level, applying FCFS within each class. Outcome metrics for clients can vary by priority class in such models [106]. Additionally, in accumulating priority queues, clients' priority levels increase the longer they wait [17, 106].

Application of Queueing Models in Healthcare

Queueing models assess system performance through various stochastic metrics, including client wait times, service completion, system occupancy, utilization rates, and server idle periods [25]. These metrics, characterized by time-dependent distributions, are summarized through key performance indicators (KPIs) such as average values, probabilities, proportions, and rates. KPIs reflecting client experiences, like abandonment and extended waiting, serve as proxies for service access and quality. This section will explore methodologies for estimating or approximating these dynamic performance measures in queueing models.

To approximate time-dependent properties in queueing models, one strategy involves utilizing steady-state equations within discrete time intervals [39, 40, 41, 80]. This includes the pointwise stationary approximation, which integrates time-varying arrival rates into steady-state equations [39, 40, 41], and the modified offered load method, which estimates based on the expected number of occupied servers in an infinite server model [42, 59, 112]. These steady-state methodologies tend to be less precise with fluctuating arrival rates [58, 39, 55] and depend on the availability of precise steady-state equations [48].

Another set of techniques models time-dependent properties of queueing systems through differential equations [18, 19, 22, 42]. Specifically, for Markovian models, the Chapman-Kolmogorov equations offer a precise framework of ordinary differential equations (ODEs) for state transitions [42, 22, 57]. Yet, solving these ODEs numerically becomes less efficient as the model's complexity increases [55]. To manage this, approximation methods like state space truncation through randomization [55, 56], and closure approximation, which simplifies the system to initial client numbers using a smaller set of ODEs [18, 101], are employed.

Alternative methods in queueing models involve continuous frameworks, where fluid approximations model the net flow of arrivals and departures through a deterministic process, effectively capturing scenarios where changes in arrival rates significantly impact queue variability [76, 72]. Diffusion models, leveraging reflected Brownian motion, simulate stochastic variations in arrivals and departures, formulated through partial differential equations. These models are versatile, accommodating a variety of arrival, service, and abandonment scenarios, including client retrials, service networks, and different customer categories [71, 72, **?**]. The accuracy of fluid and diffusion approaches depends on correctly identifying the queueing system's load status—overloaded, underloaded, or critically balanced [65, 72, 71].

Performance in time-dependent queue models can be simulated by generating stochastic event processes (arrivals, service starts, and departures) through pseudorandom numbers [100]. Discrete event simulation (DES) captures system states post-event, offering a sequence of possible outcomes, while discrete time simulation (DTS) updates states at fixed intervals [11]. Both methods, through aggregation of multiple simulations, facilitate key performance indicator (KPI) estimation [11]. DES generally provides more precise estimates than DTS, but DTS efficiency can be optimized by adjusting the time-step parameter [11]. Both techniques adeptly model complex queueing scenarios.

In queueing models with a homogeneous Poisson process, Discrete Event Simulation (DES) uses pseudorandom sampling from an exponential distribution to determine customer inter-arrival times, thus scheduling their arrival in the event sequence. To simulate a non-homogeneous Poisson process, one method involves creating a homogeneous Poisson process at an increased rate and then using a time-dependent Bernoulli experiment to decide on the acceptance of arrivals, based on the ratio of the non-homogeneous to the inflated rate [100]. This method, known as Poisson thinning, utilizes the decomposition property of Poisson processes, where any randomly selected subset also forms a Poisson process [82].

Chapter 3

Details of the Critical Care Modeling

In this chapter of my thesis, I lay the groundwork for creating a model and simulation that will help improve how ICUs operate by first looking closely at what resources are available. This includes checking how many beds there are, how many nurses are working, and how much ventilation equipment is on hand. It is important to understand not just what is available, but also what each patient needs during their stay in the ICU, which can vary a lot depending on their condition. I also take a close look at patient admission pattern to the hospital and use this information to explain how I built a model of the ICU system.

3.1 Critical Care Resource Pool

In a hospital's ICU, where a wrong decision can endanger a patient's life, crucial decisions affecting patients are made both inside and outside the unit, and not just after a patient has been admitted, therefore, decisions need to be made regarding which patients are admitted to the unit and the timing of each patient's admission.

In an ideal scenario with unlimited resources, ICU beds and staff would be readily available for every patient. However, in reality, hospital administrators must balance costs and benefits due to limited resources before deciding on resource allocation, therefore, the outcome is ICU units functioning with diverse capacity limitations. Among these constraints, the most critical typically arises from the fixed number of beds, nurses, and ventilators to the unit. Insufficient resources, resulting in the referral of patients to ICUs in other hospitals, early discharge of existing patients, and cancellation of elective surgical procedures, have severe consequences for the recovery of the patient. There are several main resources that need to be considered: namely, beds, specialist equipment and nursing staff [62, 60, 45].

Patients require immediate ICU admission once requests are made. A slight delay may affect patient safety severely and lead to irreversible consequences. Therefore, we assume that patients are unable to wait for a bed. When ICU beds are fully occupied, emergency arrivals are directed to available beds in other hospital departments which are referred to as overflow beds which initially not staffed, is activated to accommodate an additional patient, burdening the staff and leading to a reduced quality of care in the ICU. Overflow beds, used when specific units like the ICU are at capacity, ensure patients receive the same treatment as they would in a regular ICU bed [116, 7, 70], and they can move to the ICU as soon as a bed becomes available. Length of stay for patients in overflow beds remains consistent with ICU length of stay.

The hospital aims to improve patient flow, provide adequate care, and cut costs by establishing a step-down or HAU between the ICU and downstream wards [33, 90]. HAUs offer more acute care and closer monitoring than general wards, with fewer resources than an ICU. HAUs can admit stable, sub-acutely ill patients, thereby relieving ICU pressure and freeing beds for more severely ill patients. While these wards typically do not support mechanical ventilation, they do offer some organ support. HAUs offer a higher level of care compared to general hospital wards, with better nurse-to-patient ratios, greater access to respiratory therapists and perfusionists, and advanced equipment. HAUs are more economical in terms of technology and staffing, with a typical ratio of two patients per nurse, compared to the one-to-one ratio in ICUs. They require fewer resources than ICUs but are more expensive than standard hospital wards [99, 104, 66].

When a patient requires an ICU bed but none is available, and overflow beds are also full, then patients nearing the end of their ICU stay may be discharged early to HAU to free up beds for new critical care admissions, this process is known as bumping [98]. Releasing a patient 48 hours early can raise post-discharge mortality risk by up to 39%, especially for night discharges [37]. Patient vulnerability often emerges after a reduction in care level, and care continuity is disrupted during the handover. On the other hand, bumping decisions may affect patient safety; denying ICU admission to a critically ill patient reduces their survival chances, as they are placed in an area with less intensive medical and nursing support [37]. If an ICU admission is granted, the displaced patient faces risks if their recovery assessment is inaccurate. Full ICUs see more bumping decisions, often negatively impacting patients who are improperly moved [105].

Considering the high demand for ICU admissions and the outcomes of bumping decisions, the critical care model defines bumping as transferring an ICU patient who no longer needs ventilation, has completed 85% of their treatment, and is not critically ill, to another care setting. This frees up ICU beds for more critically ill patients. Rutherford et al. [102] only considered the bumping from ICU, But I also assume that, like in the ICU, patients in overflow can be bumped to a HAU to optimize bed and nurse availability.

Some patients, including those bumped from ICU and overflow or post-ICU stay, may require a stay in HAU. The model assumes that if a patient is bumped from the ICU before completion and needs to monitor more closely after ICU stay in HAU, their remaining ICU time adds to their projected HAU length of stay. After completing HAU treatment, patients might transfer to a ward for further care or may transfer home. This model assumes immediate ward bed availability without any wait after HAU stay [102]. Table 3.1 details different bed resources in this model.

In the simulation, another key resource is the number of available nurses in the critical care system. I will discuss the crucial roles of nurses in hospitals, particularly in ICUs, and elaborate further on this in nurses are vital in hospitals and ICU, undertaking numerous essential duties:

- Patient Care: Nurses provide direct care to patients, administering medication, monitoring vital signs, and assisting with daily activities. In the ICU, this involves more complex and critical care, often for patients with life-threatening conditions.

- Assessment and Monitoring: Nurses continuously evaluate and monitor patient health, ensuring prompt intervention. This is especially crucial in the ICU for patients on life support or with unstable conditions.

- Emergency Response: Nurses commonly serve as initial responders in emergencies, administering crucial care before doctors' arrival. In the ICU, their expertise lies in swiftly assessing and responding to life-threatening situations, and ICU nurses are required to have highly specialized training.

Therefore, nurse resource pool is one of the most important resources in the ICU. Shortages of experienced nursing staff and rising demand largely contribute to ICU bed closures, admission refusals, and excessive patient transfers to other hospitals [2]. I utilized a shared resource pool for both HAU and ICU nurses, driven by the efficiency of pooled resources over dedicated ones. For instance, a single queue for a group of servers (like cashiers, bank tellers, ICU beds) leads to reduced waiting times compared to individual queues for each server. Nurse-to-patient ratios are set using patient dependency scales, based on the idea that critically ill patients require more nursing care. In the ICU and overflow, this ratio is one nurse per patient, ensuring individualized care. In the HAU, it shifts to one nurse for every two patients [45, 102, 64]. The model operates under the assumption that these specified nurse-to-patient ratios remain consistent for all patients during their entire stay within each respective unit. Currently, the model operates on the assumption that there exists a unified pool of nurses shared between the ICU and HAU units, without distinguishing separate staffing for each unit. In the Table 3.1 the ratio of patient to nurse in each section of this critical care model is readily available

In addition to nurses and beds, the model also considers the availability of mechanical ventilators for ICU stays. A mechanical ventilator is a medical device that supports or takes over breathing for patients who are unable to breathe adequately on their own, typically due to respiratory failure, neurological conditions, or post-surgery. It can either completely control or assist with breathing and is used either invasively through a tracheal tube or

Level of care	Bed characteristics	Patient/Nurse ratio
1	Standard Ward bed No organ support, no ventilation	3 or more to 1
2	HAU bed Support single failed organ system, no ventilation	2 to 1
3	ICU bed Invasive ventilation and multiple organ support	1 to 1

Table 3.1: Levels of care characteristics at BC Hospital [1]. This information is current to Aug 2019.

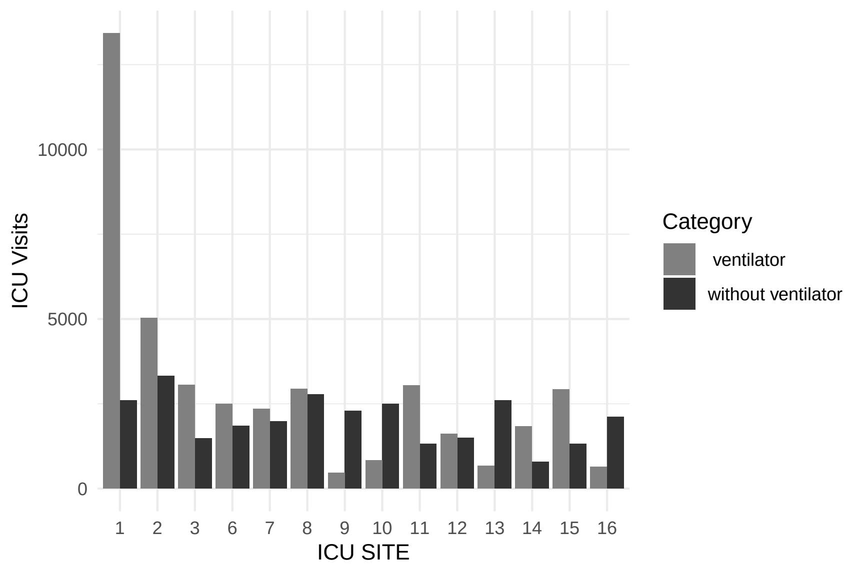


Figure 3.1: Number of ICU visits involving ventilation usage at different ICU sites based on critical care database 2019.

non-invasively with masks. Mechanical ventilation is a leading reason for ICU admissions. Recent studies estimate that around 70% of occupied ICU beds are used by patients needing mechanical ventilation at any given time [12, 114]. Figure 3.1 illustrates that most patients require ventilation during their ICU stays, and Figure 3.2 indicates that patients needing ventilation tend to have longer ICU stays. Mechanical ventilation is not available in the HAU; it's only accessible to patients in ICU or Overflow beds during their stay. The model includes a provision for a maximum of five mechanical ventilators per ICU visit, since occurrences beyond this number are rare Figure 3.3 shows comparison of number of ICU visits for patients with ventilation v.s. without ventilation usage. The simulation model captures a 2-hour cleaning time for ventilators after each use, following expert recommendations. Also, the mean ICU stay for patient with ventilation v.s without ventilation usage in ICU is capture in Figure 3.2.

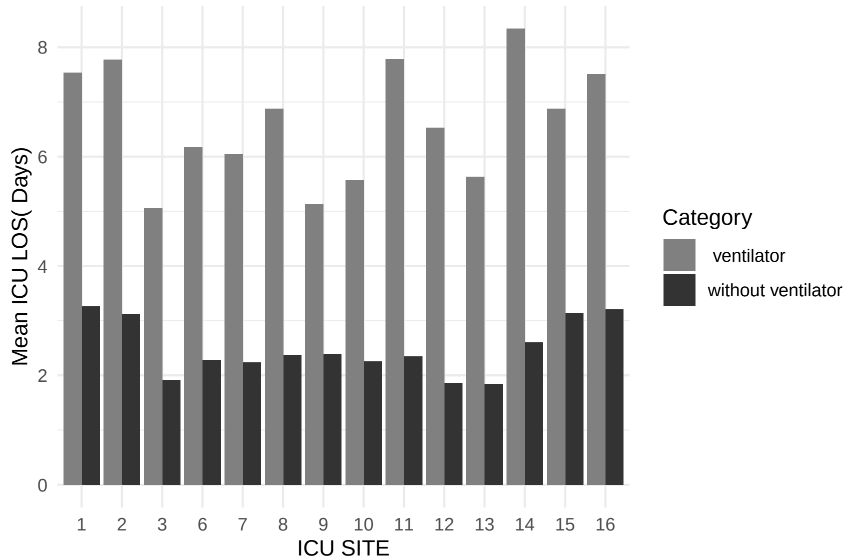


Figure 3.2: Average ICU stay length of stay (Days) comparing ventilated and non-ventilated patients across different ICU sites based on critical care database 2019.

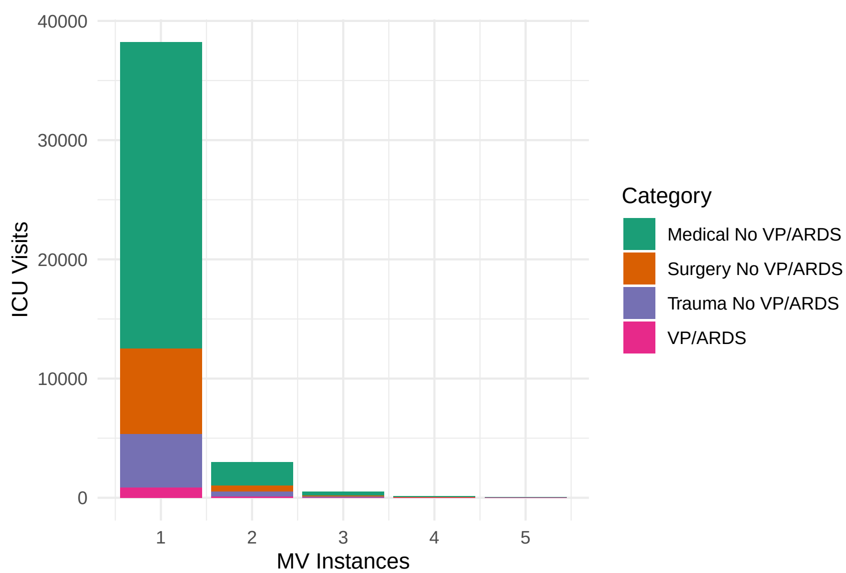


Figure 3.3: Distribution of patients across different categories for different number of mechanical ventilator instance during their ICU stay based on critical care database.

3.2 Patient Arrival Rates

Many studies [103, 70, 89, 97] use a Poisson process as a statistical model for patient arrival at a health care center. For the critical care model, does the rate of arrival vary over time? The arrival rates for each of the ICUs in the simulation model were calibrated from the patient admission data in the Critical Care Database. Arrival rates were analyzed by time of day and day of week and month of the year, and the results are shown in Figure 3.4. Although the simulation model contains only the eight major ICUs in British Columbia, arrival rates were calculated for 16 hospitals in the Critical Care Database. I found only slight variation in the admission rates by days of week and months of the year as you can see in the Figure 3.4 part (b) and (c), but there is a significant variation by time of day in Figure 3.4 part(a). Therefore, the non-homogeneous Poisson processes that simulate patient arrivals in the model were calibrated to use a different arrival by hour of day, but day of week and moth of the year dependence were not included in the model. The following characterization of a non-homogeneous Poisson process illustrates the model's suitability for representing the patient arrival pattern.

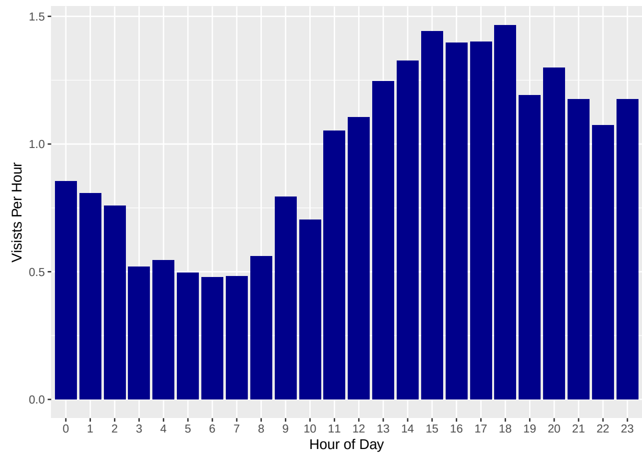
Let us denote by N_t the number of patients that arrive up to time t. The stochastic arrival process of patients $\{N_t, t \geq 0\}$ is a Poisson process if:

1. Patients arrive one at a time.
2. The number of arrivals in the time interval $(t, t+s]$, $N_{t+s} - N_t$, is independent of the number and times of arrivals from 0 to time t. That is, it is independent of the variable set $\{N_u, 0 \leq u \leq t\}$.
3. The distribution of $N_{t+s} - N_t$ is dependent of t for all $t, s \geq 0$. Here t is defined as the time of the day.

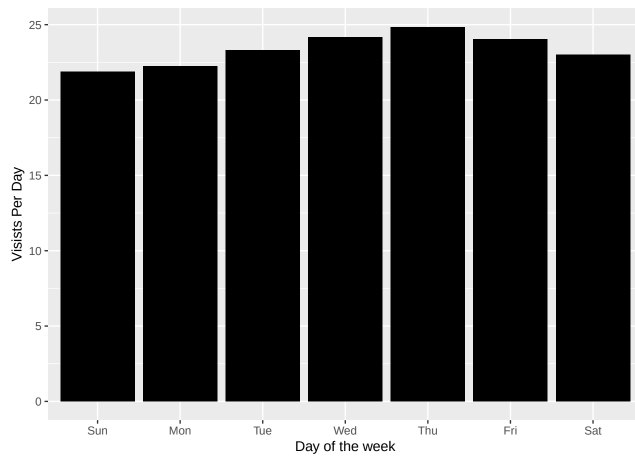
Patients arrive at the ICU on an individual basis, their arrival times are not influenced by prior patient arrivals and they are not coordinated according to any pre-established plan which are address the first two conditions, and Condition 3 sets the non-homogeneity of the process through time.

Patient Arrival Capacity Check Protocol

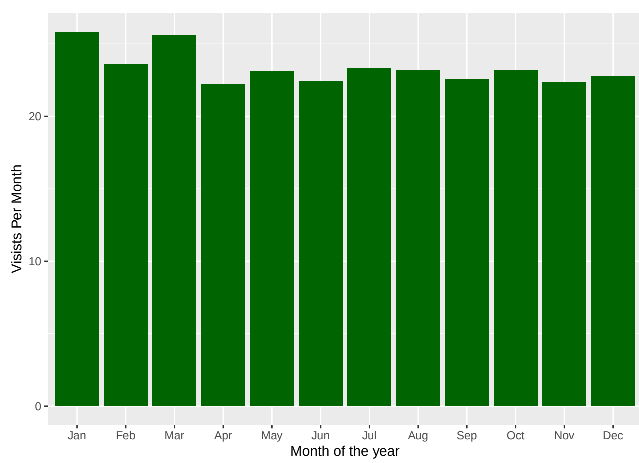
The detailed algorithm for determining if there is sufficient capacity to admit a new patient is as follows patient admission to the ICU depends on the availability of resources such as beds and nurses. A new patient can be admitted without restrictions if an ICU bed and a nurse are available. However, admissions become more challenging when the ICU has limited resources. When there is only one available ICU nurse, Algorithm 1 is used to admit patients. In situations where no ICU nurses are completely available and a HAU nurse is caring for only one patient, Algorithm 2 is applied for admitting the patient. If there are



(a) Arrival rate by hours of day.



(b) Arrival rate by day of week.



(c) Arrival rate by month of year.

Figure 3.4: Arrival rate by (a) hours of day and (b) day of week and (c) week of month for ICUs at 16 hospitals in British Columbia.

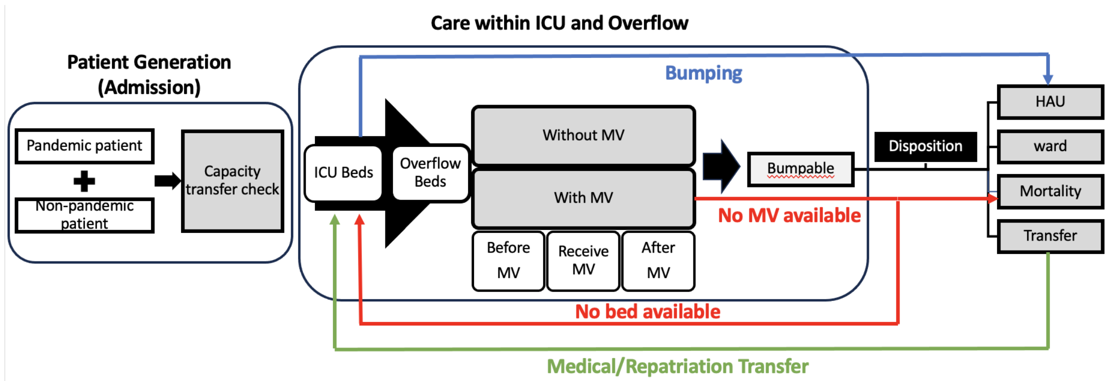


Figure 3.5: ICU Admission process for pandemic and non-pandemic patients based on bed availability. The red line shows patients who they can not admit to the ICU because of no avaiable bed then they can transfer to new ICU. The model includes bed assignments (ICU or Overflow), ventilation needs, ICU stay phases (between, during, after-Ventilation), and patient transfers to other ICUs due to repatriation and medical transfer reason. After completing their treatment in the ICU, patients may be transferred to a ward, HAU, another ICU, or in some cases, may pass away. Patients also will die if they couldn't recieve ventilation in time.

no available beds in either the overflow or ICU, and no patients in either area are suitable for bumping, then Algorithm 3 should be utilized. Figure 3.5 presents a simplified overview of the critical care system and its dynamic nature, offering a visual representation of the complexities involved.

3.3 Simulation of the Critical Care Queueing Model

I developed a Discrete Event Simulation (DES) model to accurately estimate performance indicators within the queueing framework, effectively capturing the dynamics of the Critical Care network in major hospitals. This simulation modelling can demonstrate whether the ICU supply is sufficient and estimate when capacity is reached. Additionally, comparing different scenario plannings with the model can show the effect of different public health policies on critical care system. The DES model, created using AnyLogic modeling software, features a visual interface representing each element of the patient flow, as illustrated in Figure 3.6. The simulation was presented to the Critical Care Services Executive Committee in Fall 2023, and their feedback was incorporated into model development.

AnyLogic is a comprehensive simulation software used across various industries for modeling complex systems and processes. Its multi-method modeling capability enables users to integrate different modeling paradigms, such as agent-based, discrete event, and system dynamics, within a single simulation model. With a user-friendly graphical interface and extensive libraries of pre-built objects, AnyLogic facilitates the development of sophisticated models with minimal coding effort. It also supports advanced visualization tools,

Algorithm 1 Algorithm for assigning ICU beds to new patients with a single available nurse in ICU

a new patient requires an ICU bed

while a nurse is available **do**

    if an ICU bed is available **then**

        Patient is assigned to an ICU bed

    else if ICU beds are unavailable, but an overflow bed is available **then**

        Patient is admitted to overflow and will be transferred to an ICU bed as soon as it becomes available, thereby freeing up the overflow space.

    else if ICU and overflow beds are unavailable **then**

        if bumping is possible for an ICU patient and a HAU bed with nursing staff is available **then**

            ICU patient will be transferred to the HAU. Consequently, the patient in the overflow can move to the ICU (First in, First out), making room in the overflow for a new patient.

        else if bumping from the ICU is not feasible and an overflow patient ready to bump and a HAU bed with nursing staff is available **then**

            Overflow patient will be transferred to the HAU, thereby making an overflow bed available for a new patient.

        end if

    end if

end while

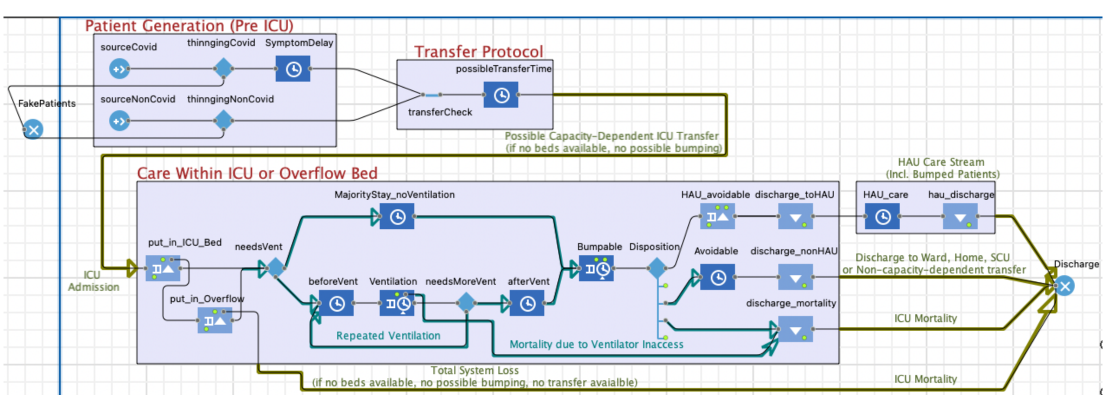


Figure 3.6: Screenshot of the DES implementation of critical care model without different transfer reasons in the Anylogic software

Algorithm 2 Algorithm for allocating ICU Beds to Arriving Patients with One Available Nurse in HAU

A new patient requires an ICU bed

while no nurses are entirely free, and there is a nurse in the HAU who is responsible for only one patient **do**

  if ICU patient is ready to bump, ICU patient transfers to HAU to free up a bed at ICU. **then**

    if an overflow patient exists **then**

      Overflow patient will be moved to the newly available ICU bed (First in, First out), thus freeing up an overflow bed for a new patient.

    else if there are no patients in overflow **then**

      A newly available ICU bed is ready for a new patient.

    end if

  else if an ICU patient is not ready to bump and ICU is full **then**

    if Overflow patient is ready to bump from overflow **then**

      A patient moves from overflow to HAU, making the overflow bed available for a new ICU patient.

    end if

  end if

end while

Algorithm 3 Algorithm for Assigning ICU Beds to Incoming Patients with No Available ICU Beds or Nurses

A new patient requires an ICU bed

while Neither overflow nor ICU beds are available, and there are no patients ready to bump **do**

  if transferring a patient is feasible **then**

    the transfer will occur based on the capacity and preferences specified in the transfer matrix.

  else if no transfer is possible **then**

    Regrettably, the patient is lost to the system.

  end if

end while

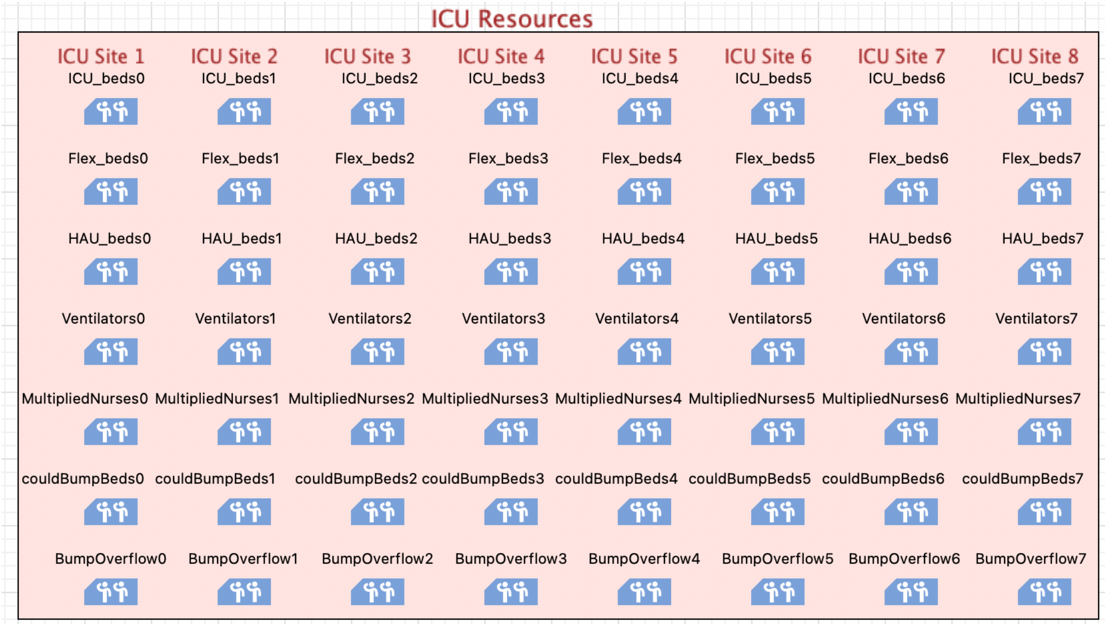


Figure 3.7: Screenshot of the ICU resources in the critical care model in the AnyLogic software

allowing for the creation of interactive 2D and 3D animations to visualize system behavior. Overall, AnyLogic empowers users to analyze and optimize the performance of complex systems through simulation modeling, making it a valuable tool for researchers, engineers, and decision-makers alike.

The model includes five resource pools: ICU beds, overflow beds, HAU beds, ventilators, and nurses. These represent units that patients can seize and release during their ICU stay and Figure 3.7 displays the ICU resources in AnyLogic model. The specific resource pool quantities are determined based on data from each ICU site and can be easily updated as needed.

On the far left of Figure 3.6 the DES model uses two source nodes based on pandemic status, each of which generates patients according to a nonhomogeneous Poisson process obtained by thinning a homogeneous Poisson process with a time-dependent rate. Pandemic patient arrival rates are based on epidemic rate and a symptom delay service block translates this into ICU admission demand. Epidemiological case projections are translated into ICU bed demand, using an $M_t/G/\infty$ queue model for the time from symptom onset to receiving an ICU bed [102], $M_t/G/\infty$ queue model represents a stochastic queuing system where arrivals follow a Markovian process (M_t), service times follow a general probability distribution (G), and there are an infinite number of servers (∞). This infinite server system simply implements a generally distributed stochastic delay and does not correspond to utilization of any physical resource. The arrival process for this model is non-homogeneous Poisson and can be based on localized pandemic epidemic projections, scaled by the proportion of critical patients. Each patient who is admitted to the hospital will be in the one of the patient categories: trauma, medical, surgical and VP/ARDS. Each of these categories is assigned a specific service time. The ICU bed used model is a single service block for different categories of patients, which sets the service time distribution based on their patien category. If a bed is unavailable, the patient is transferred to an alternate ICU locations.

ICU and overflow beds, maintained as constant-size resource pools within each ICU, are accessed by simulated patients through a seize block upon ICU admission. The seize block is essential for entities to acquire necessary resources from a pool before proceeding with their activities. Parameters within the block specify resource type, quantity, and acquisition conditions. Once seized, entities utilize resources until release. This block is crucial for modeling resource allocation and utilization in various industries. Patients will compete for the bed and nurse at the same time if one of them it is not available then patient could not admit to the ICU. If an arriving patient finds that all beds are occupied, they are sent to a queue. If ICU bed available, patients occupy these beds throughout their ICU stay, and patients my requiring ventilation or not.

The DES model simulates patient flow in the critical care network using two service blocks representing two care components (ventilator usage and bumpable). The service block is vital for simulating the processing of entities through specific tasks or services. Entities enter the block when they require service and remain until their service time is completed. Parameters within the block, such as service time distributions and resource requirements, define the service process. The Service block is essential for modeling resource utilization and queueing behavior in various industries within DES models. Each block models a patient waiting until a resource (mechanical ventilator or HAU bed) becomes available. Once available, both the patient and the resource experience a pseudo-randomly generated service time. Simulated patients have a wait limit, and if exceeded, they timeout and exit the model. Ventilation, nursing, and bed resources are each modeled with separate resource pools, their levels varying within defined usage probabilities for both resources and patients. When a simulated ICU stay ends or if a patient needs to transfer to HAU, or if they cannot receive ventilation when needed, a release block frees up the bed resource for other simulated patients. During their ICU stay, patients may need to use ventilation. I present a queuing system, the core of which is a generalized Erlang loss model [117] for ventilator use by non-pandemic and pandemic ICU patients.

The ventilator use model has the form of a generalized ./G/c/c Erlang loss model with a limited supply of c ventilators, where ventilation time has different distributions for different patients categories. If all c ventilators are in use, then ICU patients needing mechanical ventilation are lost to the system, motivated by the life-threatening nature of respiratory failure. ICU patients require a ventilator based on a non-homogeneous Poisson process, which can capture time-dependence in ICU admission rates. Patients may discharge from the ICU to HAU or experience a mid-stay bump, but there could be delays in receiving an HAU bed. This may result in patients staying in the ICU for a longer duration until an HAU bed becomes available. HAU beds, represented as a constant-size resource pool, are accessed by simulated patients through a seize block when they require an HAU bed. Patients occupy these beds throughout their HAU stay for close monitoring. When a simulated HAU stay concludes, a release block makes the HAU bed resource available for other simulated patients. I utilized five delay blocks (without ventilation, before ventilation, after ventilation, HAU stay, and ward stay) to determine patient ICU LOS. These blocks delay patients based on distribution parameters specific to each block's data.

3.4 Different Transfer Reasons

There are several reasons why patients may choose to transfer to a new hospital, including factors such as capacity constraints, repatriation, and, most notably, a higher level of care. This model considers various reasons for transfers. A detailed summary of these reasons can be found in Table 3.2.

Capacity Related Transfer

Patients who can not receive a bed in the ICU and overflow and there is no bumping possibility from ICU and overflow, therefore the home ICU site need to send this patient to a new ICU site which has a available bed. When a transfer is requested, a transfer preference/priority matrix is employed to evaluate and prioritize alternative ICU options, assisting in the selection of the preferred destination ICU for capacity-dependent transfer patients. Overflow beds are excluded from consideration when assessing availability in other ICUs, ensuring that patients are not transferred based on bumping others in ICU or overflow. The readiness of an ICU to accept transfer patients is determined by the availability of both an ICU bed and nurse. If no ICU can accommodate the transfer due to lack of available space, the patient is die.

The model integrates a database call to access transfer time data, currently set to zero pending data analysis. Presently, the database call assume deterministic transfer times, but introducing a stochastic element could better reflect real-world variations. However, introducing nonzero transfer times raises concerns about managing situations where a patient begins transferring to an empty bed, but another patient arrives concurrently. Presently, the model does not address this scenario, highlighting the need for further consideration and potential implementation of protocols or mechanisms to manage conflicts during transfers.

Presently, the model operates under the assumption that a patient's LOS is exclusively determined by the host ICU and remains unaffected by transfers. However, future enhancements could contemplate integrating additional factors such as potentially extending LOS by one day to accommodate reduced stability stemming from transfers. Implementing this adjustment may pose increased complexity due to the existence of varying treatment components within the model.

Medical Reason Transfer

Some patients, despite being critically ill, may be unable to receive the required treatment at their original hospitals. In such cases, transferring them to the ICU becomes essential to ensure they receive the necessary care. These transfers are classified as Medical Reason Transfers or patients who need to receive higher level of care (HLC Transfer). The analysis reveals that while the number of patients requiring transfer to a higher level of care may be relatively low, these individuals tend to have significantly longer durations of stay in the transferred hospital compared to the overall ICU patient population, as illustrated in Figure 3.8. As this group of transferred patients tends to occupy ICU beds for an extended period, it can significantly impact the capacity of critical care units, resulting in a reduction in available ICU beds. From Figure 3.9, the majority of transferred patients will transfer to larger hospitals like ICU sites 1, 2, and 3, primarily due to the presence of specialized units for specific diseases or conditions.

Upon arrival at the receiving hospital, transferred patients receive expedited admission to the ICU without the need to wait in a queue, ensuring prompt access to critical care. If no ICU bed is available, the patient may be accommodated in the overflow unit. In the event that neither the ICU nor the overflow unit has an available bed, the model checks the possibility of bumping to free up a bed at either location. Patients may also be transferred from the overflow unit to the ICU if a bed becomes available. If no such options are feasible, the patient is accommodated at ICU site 1, as this particular site does not have a refusal policy, thereby ensuring immediate care needs are met. This observation is supported by Figure 3.10, which illustrates that ICU site 1 experiences higher ICU visitation rates compared to other sites.

Additionally, patients within ICU site 1 who are unable to secure a bed in the ICU or overflow units will be accommodated at the HLC Overflow of ICU site 1, which offers an unlimited number of beds and nursing support. For patients outside of ICU site 1, if an ICU bed is unavailable upon transfer, they will be redirected to ICU site 1 to ensure they receive the necessary care. Furthermore, patients transferred from HLC have the highest priority for securing a bed in the ICU compared to other patient types. Notably, HLC patients are unable to bump in their original ICUs. Most patients transferred for medical reasons fall within the Medical no VP/ARDS category, as shown in Figure 3.10, which also highlights the mean ICU length of stay for medical reason transfer patients across different patient types. From the database, it is observed that 2.247% of patients are transferred due to the need for a higher level of care.

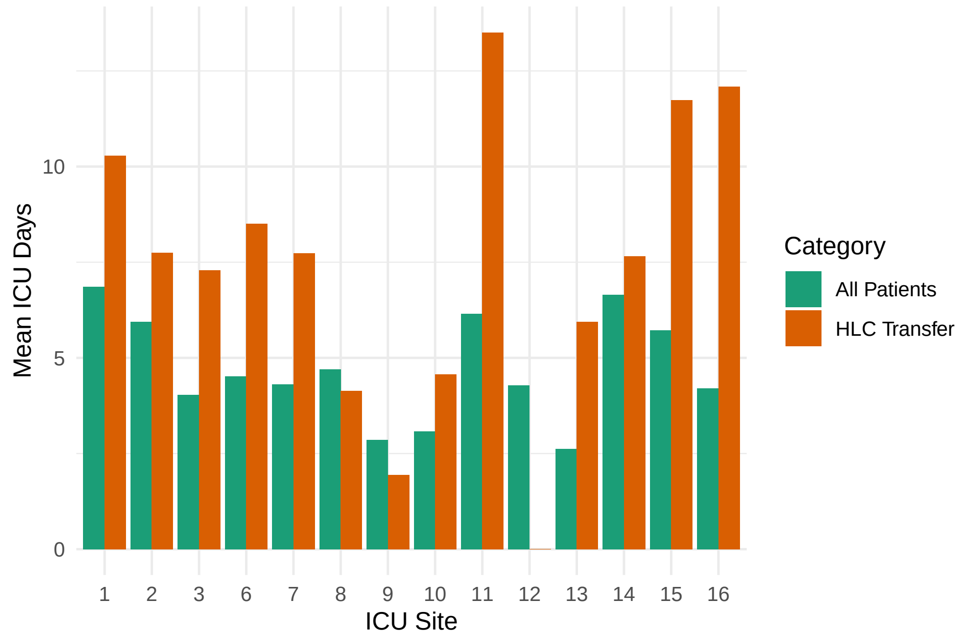


Figure 3.8: Mean ICU LOS for all ICU patients versus medical reasons transfer patients based on critical care database.

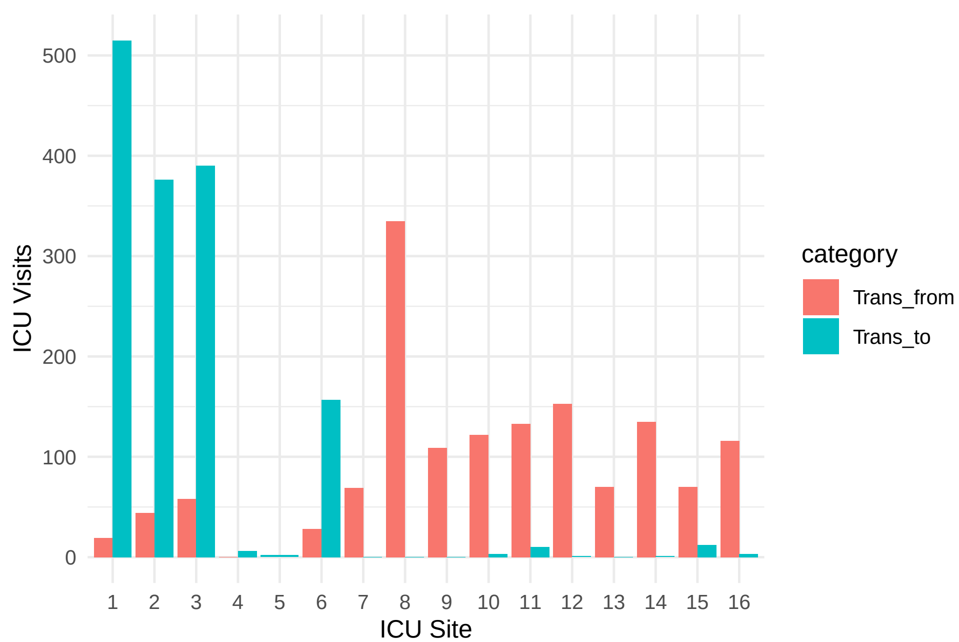


Figure 3.9: Comparing ICU visits for medical transfers between ICU sites

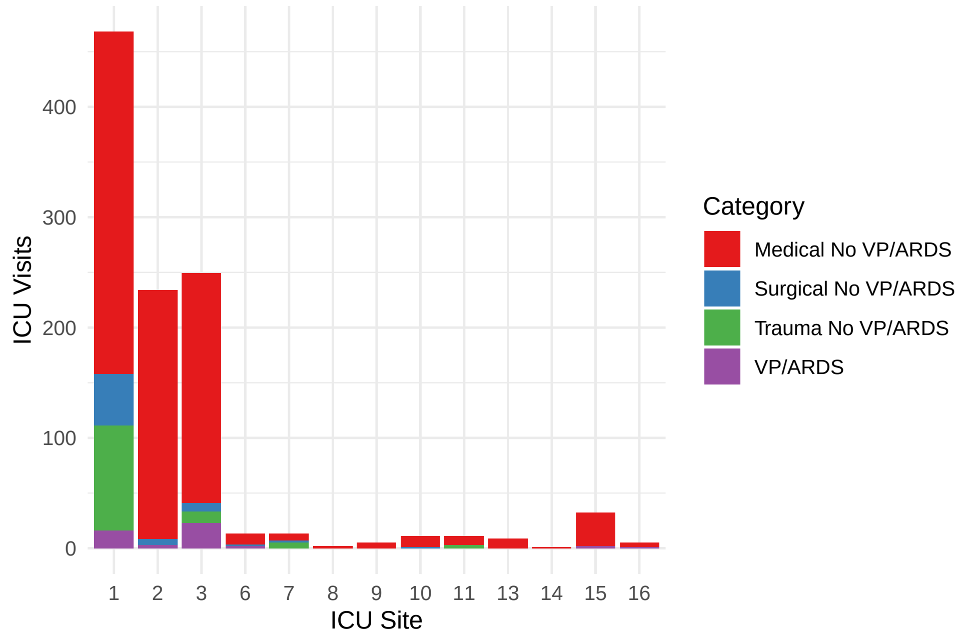


Figure 3.10: ICU visits for medical transfer patients by ICU Site and patient type

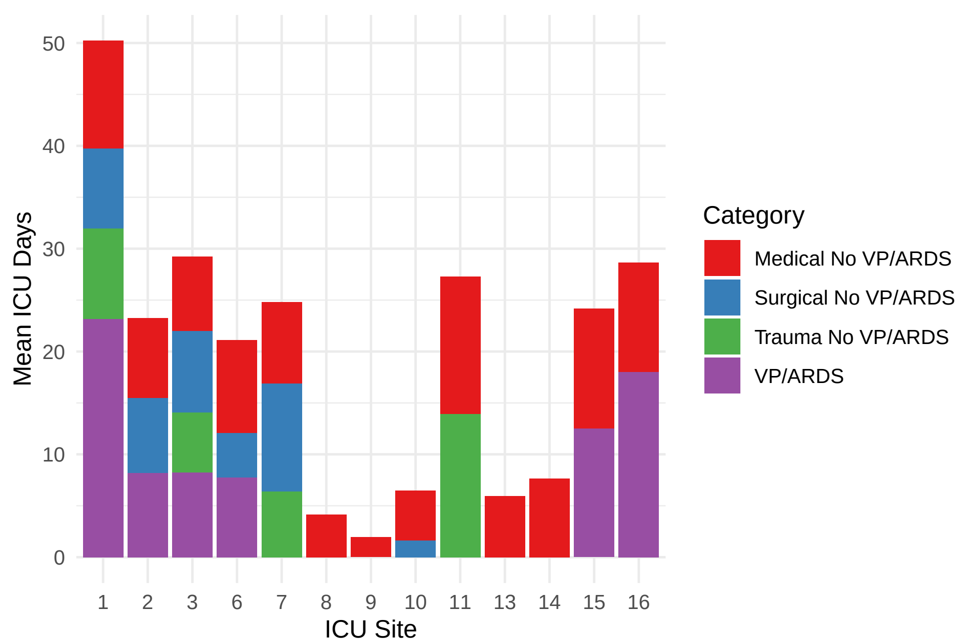


Figure 3.11: Mean ICU LOS for medical transfer patients by ICU site and patient type

Repatriation Transfer

Repatriation transfer can refer to the transfer of a patient back to their home hospital or healthcare facility after receiving treatment or care at another facility, typically a higher-level or specialized medical center. This transfer is often necessary when a patient's condition has stabilized or improved to the point where they no longer require the specialized services or resources available at the treating facility, and it is more appropriate and convenient for them to continue their care closer to home. In the context of ICU capacity management, repatriation of patients becomes a practical solution when the admitting ICU is at full capacity and needs to accommodate new critically ill patients. While these patients could potentially be transferred to either an ICU or a HAU bed at their original location, this model is focused solely on their return to the ICU, excluding HAUs. Notably, the volume of repatriation patients is less than that of medical transfer patients, and their ICU stays are typically shorter. This process highlights the intricate balance between ensuring ongoing care for stabilized patients and freeing up vital resources for those in urgent need.

Transfer Type	Transfer Reason	Priority
Capacity Related Transfer	Lack of available beds	Low
Medical Transfer (HLC)	Access to Advanced Care	High
Repatriation Transfer	No longer requires advanced monitoring	Low

Table 3.2: Patient transfer reasons in the critical care model

Critical Care Model with Different Transfer Reasons

A key aspect that I now incorporate into the model is the process of medical transfers between hospitals. This scenario typically arises when patients, already admitted to an ICU, require more specialized care that is not available at their current location, necessitating their transfer to larger facilities. I assume that patient will not bump in the current hospital location as they are critically ill and we need to transfer them between ICU sites. Depending on the nature of their condition, these patients are admitted through either non-pandemic or pandemic streams.

High Level Care (HLC) patients represent a critical category requiring immediate ICU admission upon request, given their severe health conditions. Any delay, even slight, in providing the necessary care could have dire consequences, including irreversible harm or endangering patient safety. Consequently, HLC patients are granted the highest priority, without any waiting time for ICU beds. If ICU beds are fully occupied, these patients are either overflowed to available beds in other departments or diverted to other hospitals. When HLC patient receive an overflow bed, they remain in the queue for an ICU bed and are transferred to the ICU as soon as a bed becomes available. In extreme cases where neither ICU nor overflow beds are available, and patient bumping is not an option, patients are directed to a designated HLC overflow facility, specifically ICU site 1, Figure 3.12. This site is unique in its refusal policy, ensuring HLC patients are guaranteed a bed; it acts as a last resort when no other options are available.

This model implements a first-in, first-out queue system for managing ICU admissions, emphasizing the elimination of waiting times for HLC patients. However, this placement is not the end goal for their care pathway. While in the HLC overflow, patients are essentially on a priority list, awaiting transfer to a more appropriate care setting that matches the intensity of care they require. This could be either a standard ICU bed as it becomes available or a bed in the overflow area that offers a closer approximation to the specialized services found in an ICU.

Mechanical ventilation resources are not bound to a fixed location but are shared across different care settings ICU, overflow, and HLC overflow. This flexible sharing mechanism ensures that mechanical ventilation can be administered to patients irrespective of their physical location within the hospital, reflecting a realistic and responsive approach to critical care. A key principle underpining the model is the prioritization of HLC patients for mechanical ventilation access. Given the severity of their conditions, these patients are considered the most vulnerable and hence are given precedence in the allocation of ventilators. Moreover, the model reflects a crucial operational detail from real-life healthcare settings: the avoidance of unnecessary cleaning or sanitization tasks for the ventilation equipment when it remains with the same patient during transfers between care settings. In actual hospital environments, comprehensive cleaning and sanitization protocols for mechanical ventilators are typically reserved for instances when equipment is transferred between different patients to prevent cross-infection and ensure patient safety.

An important aspect of this journey involves the stabilization of an HLC patient's condition and the subsequent decisions regarding their care trajectory, including the potential for them to be 'bumped' if necessary or for them to undergo a process known as repatriation transfer. Bumping refers to the practice of reassigning beds based on patient acuity and resource availability, ensuring that those in most critical need have access to the appropriate level of care. When an HLC patient's condition stabilizes, they may be considered for bumping to accommodate another patient requiring more urgent care. This transfer marks a significant phase in the patient's care journey, as it often indicates improvement to a point where the specialized resources of the HLC overflow are no longer necessary. In the model, these patients are considered "normal" patients upon their return to the initial ICU site, meaning they no longer have the heightened priority status that was accorded to them during their critical phase.

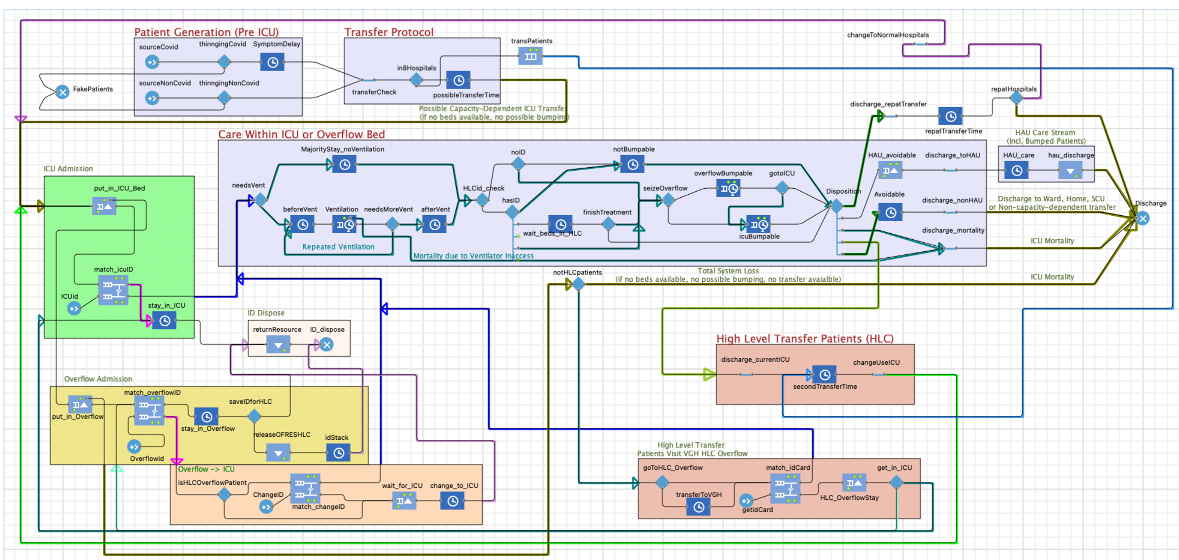


Figure 3.12: Screenshot of the DES implementation of critical care model with different transfer reasons in the AnyLogic software

Chapter 4

Analyzing Critical Care Data

In this chapter, I will analyze detailed data from the British Columbia Critical Care Database covering the years 2010 to 2020. My focus is to use statistical analysis to figure out the best lengths of stay in the ICU for different groups of patients, considering their diagnosis, whether they need to use mechanical ventilator, and other specific needs. I will also look into the different Transfers between hospitals to better understand how the care process works. The goal is to build a model that shows how the ICU really operates.

4.1 Different patient categories

Based on expert opinion, I categorized patients into four patient types:

Patients with either VP or ARDS

Patients with Viral Pneumonia (VP) or Acute Respiratory Distress Syndrome (ARDS) present notable considerations for ICU management. Viral Pneumonia is characterized by lung inflammation resulting from a viral infection, manifesting symptoms such as cough, fever, and shortness of breath, with common viruses including influenza and SARS-CoV-2. On the other hand, ARDS is a critical lung condition caused by the accumulation of fluid in the air sacs, leading to reduced oxygen flow to the bloodstream. This condition can arise from various sources, including pneumonia, and typically necessitates intensive care and mechanical ventilation. The diagnosis of VP or ARDS significantly affects the LOS in the ICU. However, due to their relatively low incidence rates in ICU populations, patients within the VP or ARDS categories are not further subdivided by admission criteria.

Medical patients with neither VP or ARDS

Medical patients who do not have VP or ARDS encompass a diverse group with a wide range of medical needs. These individuals present with various conditions that span from minor infections to severe disorders, including heart disease, cancer, and respiratory problems. The nature of these conditions varies; they can be either acute, requiring immediate and short-term care, or chronic, necessitating long-term management and treatment. This diversity in patient conditions necessitates a flexible and responsive approach to care within medical facilities.

Surgical patients with neither VP or ARDS

Surgical patients without VP or ARDS consist of individuals requiring intervention through surgical procedures. The conditions prompting these surgeries are diverse, covering a spectrum that necessitates either elective surgeries, planned in advance for non-life-threatening conditions, or emergency surgeries, required immediately to address acute, life-threatening issues. This variety underscores the need for a broad range of surgical expertise and preparedness within healthcare settings to effectively address the unique requirements of each surgical patient.

Trauma patients with neither VP or ARDS

Trauma patients without VP or ARDS include individuals suffering from physical or psychological injuries caused by external forces. The severity of these injuries varies widely, from minor cuts and bruises to severe conditions such as head injuries and organ damage. Such patients often require immediate and specialized care to address the complexities of their injuries, underscoring the critical need for prompt, effective trauma management within healthcare settings to mitigate the impact of these injuries and facilitate recovery. Figure 4.1 presents the distribution of ICU visits by patient categories across non-Fraser Health Authority sites, while hospitals 4 and 5 are within the FHA so there is no data available for these two hospitals. The figure indicates that medical patients without VP/ARDS have the most ICU visits compared to other categories in each ICU site.

Figure 4.2 depicts the average ICU LOS days by patient categories across various ICU sites. The figure indicates that VP/ARDS patients have the higher mean ICU LOS compared to other categories in each ICU site.
Figure 4.3 shows the number of ICU visits across different patient categories with v.s without ventilation usage, also the number of ICU visit across different patient categories for each number of ventilation instances usage are shown in the Figure 3.3. Note that only instances of ventilation usage up to five times are considered due to their limited occurrence.

For my critical care queuing model of multiple ICUs, I used data from the British Columbia Critical Care Database (2010–20) to inform parameter choices. Where only partial data was available, I relied on expert opinions from clinical experts. This database began in 1997, originally known as the British Columbia ICU Database, and it provides comprehensive details on ICU and HAU admissions in most British Columbia hospitals, including patient treatment, admission and discharge times, transfers, mechanical ventilation durations, and diagnostic details. It was linked with the DAD to supplement missing

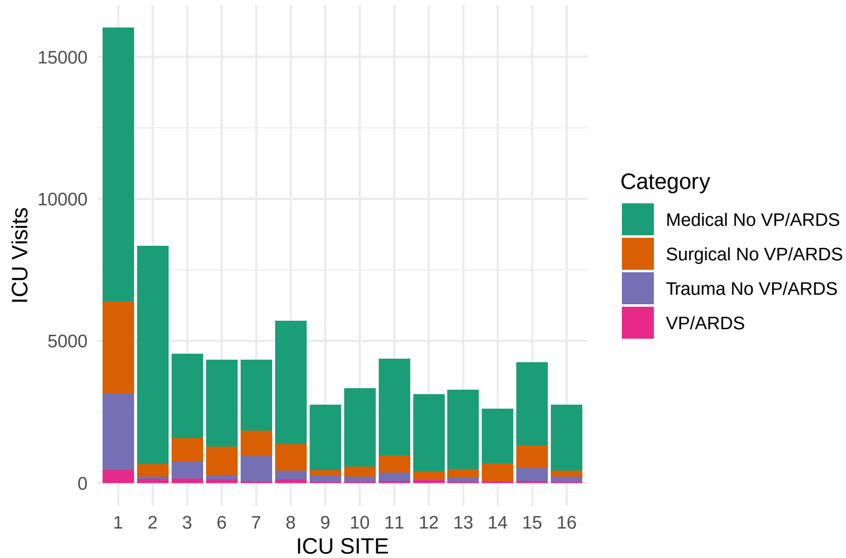


Figure 4.1: Number of patient categories in each ICU sites

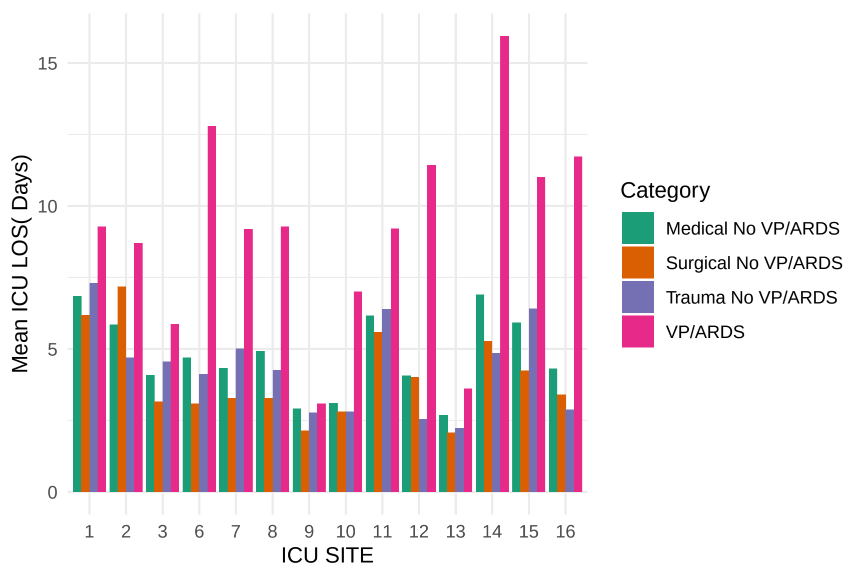


Figure 4.2: Average LOS time for each patient categories for each ICU sites

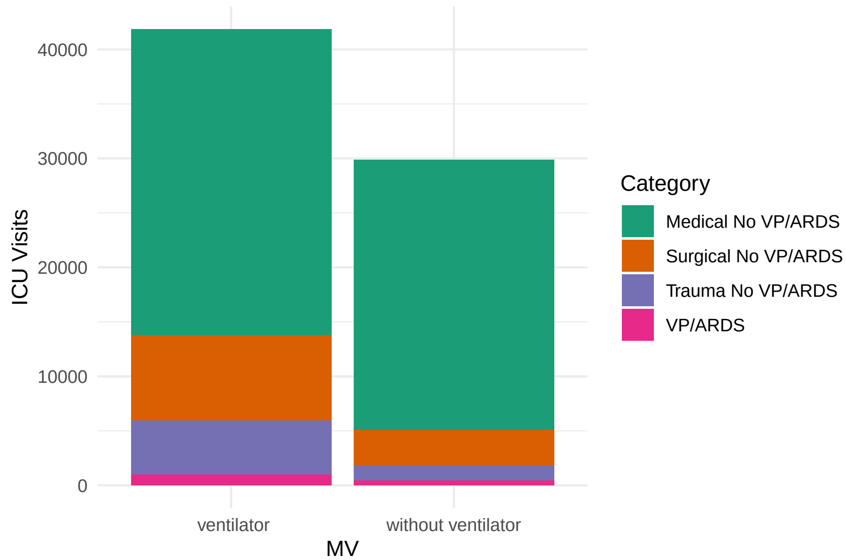


Figure 4.3: Number of patients across different categories with v.s. without ventilation usage

data. While DAD offers complete hospital admission data, it has less detail about ICUs and HAUs. The model's calibration and validation primarily relied on 2019 data, as it's the most recent full year before the COVID-19 pandemic. The pandemic's impact on critical care and the resulting data inconsistencies and capacity fluctuations led to the decision not to use 2020 data for these processes.

4.2 Details of ICU Modeling and Analysis of Service Time

The decision-making process for patient flow in the ICU involves evaluating both the individual characteristics of the patients—such as their type, the need for ventilation, LOS in the ICU, and their expected outcomes—and the current state of the ICU itself, including the availability of beds and ventilators as well as the potential for early discharge to make room for incoming patients. The duration of a patient's stay in the ICU can be broken down into five distinct phases related to treatment, specifically for those requiring mechanical ventilation: the period before they receive ventilation, the duration of ventilation, intervals between ventilation sessions, and the period following ventilation. Additionally, there is consideration for patients who do not require mechanical ventilation during their stay, focusing on their overall time spent in the ICU.

Various distributions have been used to model the LOS distribution of hospital patients, including lognormal [44, 70, 73], hyperexponential [44], weibull [91]. In this thesis, distributions for each temporal component associated with ICU stays were independently modeled at each site. For the intervals preceding ventilation, during ventilation, between ventilation sessions, and for the ICU LOS absent of ventilation, mixed gamma distributions were applied, stratifying the data according to four distinct patient categories. It is important to note that, for the distribution modeling of ventilation instances and the periods subsequent to ventilation, differentiation among patient categories was not employed due to the limited size of datasets for each category; instead, an aggregated approach was taken, combining all patient categories into a single analysis. Furthermore, each entry in the ICU database was treated as a discrete visit, acknowledging that multiple admissions to the ICU during a single hospitalization episode might not exhibit statistical independence. Subsequent sections of this document will elaborate on the precise definitions, methodologies for measurement, and the specific techniques employed for fitting distributions to these defined temporal periods.

ICU Service Time Analysis: Time Before Ventilation

The time preceding ventilation is calculated by subtracting the ICU admission timestamp from the first ventilation instance during each ICU visit. This requires merging two tables from the ICU Database: the ICU admissions (MasterADM table) and the ventilation times (Procedures table). This merger provides a complete overview of the patient's experience from admission to the start of ventilation. In scenarios where patients need immediate ventilation upon ICU admission, some cases may show no delay in starting ventilation. Considering that some patients are ventilated immediately upon ICU arrival, I calculated the notable probability of zero time before ventilation, based on the proportion of data entries with immediate ventilation. For each ICU location and patient group, we applied a hybrid gamma distribution for the duration before ventilation, combining a Bernoulli distribution for instances with zero time and a gamma distribution for non-zero times. The shape and scale parameters of the gamma distribution were estimated using the *fitdistr* function in the *MASS* R package, employing maximum likelihood estimation (MLE). Figure 4.4 displays plot of the gamma distribution fits for before ventilation usage.

ICU Service Time Analysis: Ventilation Time

Ventilation duration is determined from the start and stop times of mechanical ventilation in the ICU DB's Procedures table, focusing on procedures with the ventilation usage code. I excluded entries lacking a start or stop time but did not filter for unusually long ventilation duration as it is possible to have a long ventilation usage in ICU. My data analysis and model treat each instance of mechanical ventilation as independent and and uniformly distributed, even for multiple instances for the same patient during a single ICU visit. For each ICU location and patient category, I fitted a distinct gamma distribution to the duration of mechanical ventilation, using *fitdistr* for parameter estimation. Figure 4.5 displays plot of the gamma distribution fits for before ventilation usage.

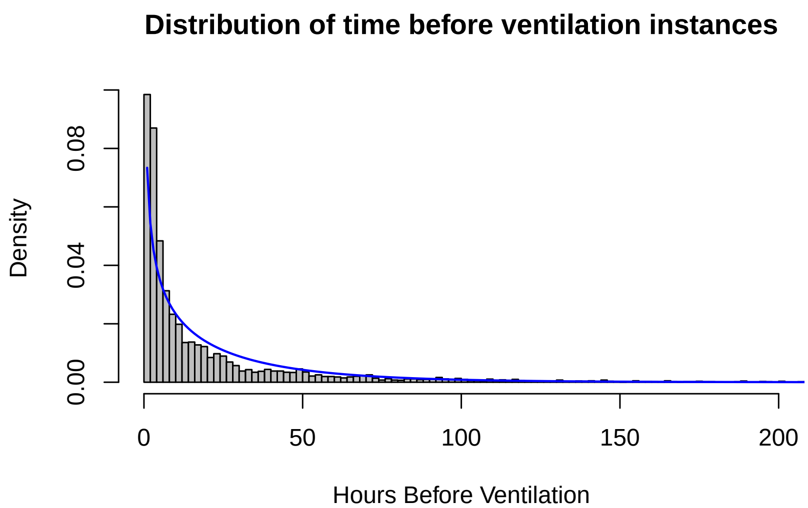


Figure 4.4: Distribution of time before ventilation instances, ICU site 1

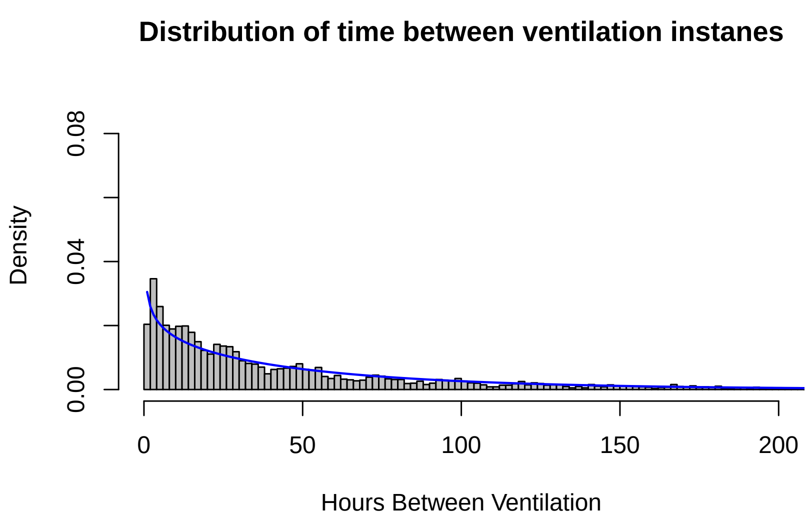


Figure 4.5: Distribution of time between ventilation for all patient groups

ICU Service Time Analysis: Time After Ventilation

In this thesis, I define the time after ventilation as the period from the last recorded ventilation instance in an ICU visit to the recorded "ready time" for patient discharge. If ready time is not documented, the ICU discharge time is used instead. In cases where ready time exceeds the actual ICU discharge time, the latter is considered. Our analysis excludes any after ventilation times that are negative or exceed 30 days and 12 hours.

The plots for time after ventilation showed distinct daily peaks at $0, 24, 48, 72$-hour marks, etc. Consequently, I used a mixed distribution model with discrete probabilities for specific day bins, segmented at $12, 36, 60$ and 74-hour intervals, and so on. For the initial interval of 0 to 12 hours, we fitted a hybrid truncated gamma distribution, accounting for a notable probability of zero after ventilation time and using a truncated gamma distribution for non-zero outcomes. I encountered convergence problems with MLE for a directly truncated gamma distribution. To address this, I estimated the shape and scale parameters for a gamma distribution with an infinite right tail, and then applied these parameters to a gamma distribution right-truncated at 12 hours. I faced convergence challenges when fitting truncated normal distributions for subsequent bins. To resolve this, I aligned all after ventilation time instances exceeding 12 hours into the (12,36) hour interval, then estimated the mean and standard deviation for this shifted data. These estimates were used as the adjusted mean and standard deviation for each daily group, with the mean subsequently re-shifted to its original range for each group. I fitted the time after ventilation distribution individually for each ICU site without further subdivision based on patient type.

Since I use observed ready and discharge times from a process with bumping as input for a model that also includes bumping, we may underestimate the true after-ventilation time distribution.

ICU Service Time Analysis: Number of Ventilation Instances

To determine the number of mechanical ventilation instances in each ICU visit, it's essential to use the "Procedures" table in the dataset (detailing ventilation times). I used the ventilation counts in visits to estimate the probability of ventilator use in a visit. I arbitrarily excluded visits with over 5 ventilation instances as outliers because more than fine times ventilation is rare, but this can be reviewed. I modeled distributions for each ICU site without differentiating by patient category.

The ventilator count relied heavily on the procedures table, so any ventilation episodes not in that table were recorded, leading to an underestimation in our distribution. Another source of underestimation is counting each ICU database entry as a single visit. For instance, if someone used a ventilator at one of the ICU site for a few days, then transferred to another ICU site and used a ventilator there, I counted this as two separate ICU visits, each with

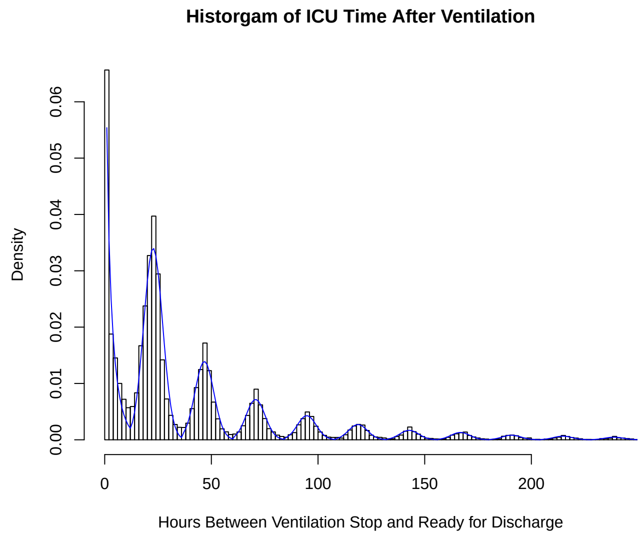


Figure 4.6: Distribution of time after ventilation for all patient groups

one instance of ventilation. However, an alternative approach might view it as one ICU "episode" with two ventilation instances.

ICU Service Time Analysis: ICU Length of Stay Without Ventilation

ICU LOS without ventilation is the time from ICU admission to the "ready time to discharge" for patients not designated as receiving mechanical ventilation in the procedures table. If there's no Ready Time to discharge recorded, I use the ICU discharge time. Also, if the Ready Time to discharge is longer than the ICU discharge time, I still use the ICU discharge time. ICU LOS without ventilation may be biased due to under-reporting of ventilation in the procedures table, potentially leading to an overestimate of the number of patients in this category.

I excluded stays with zero duration and those exceeding 45 days, then applied a gamma distribution to this data for each ICU site and patient type. In this analysis, I treat each entry in the ICU database as a separate visit, assuming they are independent and identically distributed. However, this assumption may not hold for multiple ICU visits during the same hospitalization. I do take into account multiple ICU visits for different transfer reasons, such as repatriation and medical transfers between ICU sites.

ICU Service Time Analysis: Avoidable Time

Avoidable time is the period between the recorded "ready time to discharge" when a patient could potentially leave the ICU and the recorded ICU discharge time. If there's no Ready Time to discharge recorded, then avoidable time is considered as zero. Avoidable time has

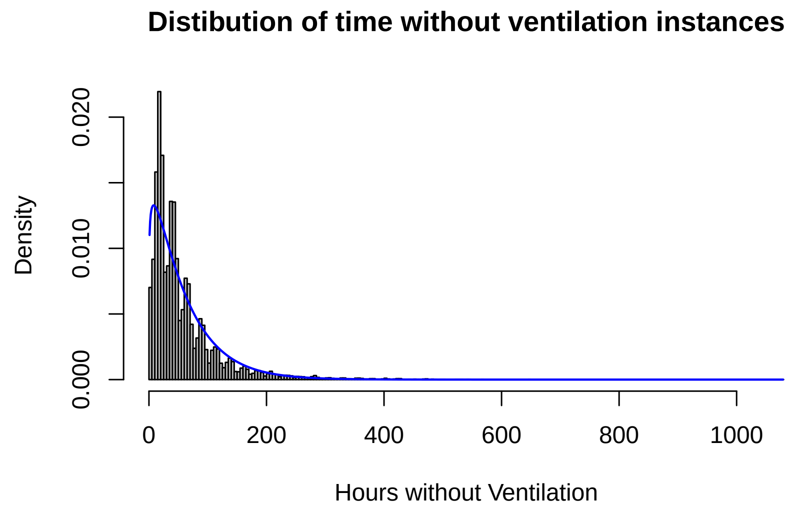


Figure 4.7: Distribution of time without ventilation for all patient groups

many sources and it is influenced by external factors beyond the ICU, such as waiting for an available HAU, ICU or ward bed, or transfer services.

I categorized avoidable time into three groups based on discharge diagnosis:

- In-ICU mortality when discharge disposition is labeled as died in the unit.
- HAU discharges identified HAU discharges by examining the discharge to special care field, which included labels like high acuity unit, ICU step-down (same site), HAU – Surgical, and cardiology stepdown.
- Other discharge dispositions For other discharge dispositions, which encompass direct transfer from ICU to an outside facility, discharged home, transfer to hospital ward (same facility), transfer to ICU (within same facility) – FHA only, and transfer to special care unit (same facility).

Patients requiring transfers for medical reasons or repatriation may experience delays due to factors like ambulance availability or securing an ICU bed at the new site. However, this model does not account for avoidable time associated with various transfer reasons.

The model assumes zero avoidable time for ICU mortality, despite some non-zero entries. I applied a single distribution for avoidable time to all non-mortality, non-HAU discharges, separately fitting it for each ICU site without considering patient types. I applied a mixed distribution to these avoidable times, following a similar approach as with time after ventilation. The discrete time intervals were set as $[0, 18], (18, 42], (42, 66]$, and so on in hours. For bins over 18 hours, we fitted individual truncated normal distributions using the mean within each time bin.

ICU Discharge Diagnosis Frequency

I determined the probability of our five discharge diagnosis groups by analyzing their occurrence rates within the data for each site and patient type. The discharge diagnosis groups are:

- HAU
- Mortality
- Repatriation transfer
- Medical Reason Transfer
- other discharge disposition

For the calculation of these probabilities, I employed data encompassing all available years, operating under the assumption that these probabilities remain invariant over time. Within the ICU database, entries indicative of mortality were delineated by an ICU discharge disposition status designated as died in the unit. Conversely, discharges directed towards the HAU were discerned through identifiers such as high acuity unit - surgical, cardiology step-down and high acuity unit or ICU step-down (same site), thereby facilitating a systematic categorization of patient outcomes post-ICU stay.

Additionally, patients were classified as repatriation transfers if their ICU Discharge Disposition matched repatriation within health authority or repatriation outside health authority. They were considered medical reason transfers if ICU discharge disposition corresponded to higher level of care within health authority or higher level of care outside health authority.

I utilized the proportion of patients discharged to the HAU as the probability of a patient requiring HAU care, irrespective of whether they were bumped from the ICU to an overflow bed. It's important to note that the data I have on HAU discharges already accounts for bumping, which means our model could potentially overestimate ICU discharges.

Chapter 5

Results

In this chapter, I explore the validation of the simulation model against actual data from the Critical Care Database, covering eight hospitals. The analysis focuses on comparing the model's predictions to real-world data from 2019, considering both standard and overflow bed utilization within ICU sites. Each site's unique characteristics, including their specializations and capacity, are taken into account. The validation process involves detailed examination of admissions and discharges to determine ICU occupancy trends, essential for assessing the model's accuracy. I considered the model's warm-up period, a critical phase for ensuring its predictions reflect realistic operational conditions. I captured the comprehensive analysis of ICU resource utilization, taking into account the mortality associated with various limitations and the diverse reasons for patient transfers.

5.1 Simulation Model Calibration and Validation

The model calibrated with the 2019 ICU database from the Critical Care Database across eight hospitals and was validated using 2019 occupancy data, a complete year of records before the COVID-19 pandemic. The pandemic-induced strain on the critical care system compromised data integrity and interpretation due to variable temporary capacities. Thus, I excluded 2020 data from the model's calibration and validation process. ICU occupancy in this model comes from the summation of ICU beds and overflow beds in each ICU site. ICU occupancy was determined by monitoring admissions and discharges in the Critical Care Database over several years.

For the simulation model, reaching a steady state (warm-up time) is crucial. Initially, the simulation model was adapted to align with queuing model principles, facilitating the identification of an optimal 'warm-up' period to eliminate initial transient states and determine an appropriate 'run time' for the simulation outputs to converge with the limiting values forecasted by queuing theory. I initially export hourly data points, calculate mean values for ICU occupancy and HAU beds, and identify the earliest point these averages are achieved, which we designate as the model's warm-up time. The warm-up time is unique for each simulation model, as follows: in the construction of the base model, which does not account for transfer factors, an initial warm-up period of 8 weeks, or 1,344 hours, was determined to be adequate. However, the revised model, which includes considerations for transfer reasons and allows for patient transfers from other hospitals to the 8 major hospitals, necessitated a longer duration to achieve a steady state. Consequently, the warm-up period for this model was extended to 21 weeks, equivalent to 3,528 hours, to accommodate the additional complexity introduced by these factors.

Upon reaching steady state in the simulation model, the selection process for acquiring additional data points is as follows: to secure robust data, the simulation model documented the system's state every 169 hours—mirroring a week plus an additional hour—to enable nearly independent evaluations and the incremental acquisition of weekly data. Following a critical warm-up phase to standardize initial conditions, the model initiated a comprehensive data collection regimen, accumulating 100 data points across each of the 24 hours in a day, resulting in a daily total of 2,400 points. This operation persisted for a duration of $409,128$ hours, calculated as $(100 \times 24 \times 169 + 3,528 = 409,128$ hours), ensuring the collection of detailed hourly data. This methodology significantly bolstered the data's reliability and analytical depth. Table 5.1 provides a summary of the critical care model's running time.

Model version	**Warm up time**	**Time interval**	**Total running time**
Without transfer reasons	1,344 hours (8 Weeks)	169 hours	406,944 hours
With transfer reasons	3,528 hours (21 Weeks)	169 hours	409,128 hours

Table 5.1: Summary of running time details for the simulation model

5.2 Validation Results and KPI Measurement

ICU Occupancy For Each Hospital

Table 5.2 compares simulation outcomes with mean occupancy from the Critical Care Database. To maintain confidentiality, hospitals are anonymously numbered from one to eight. Hospitals 1, 4, and 5 have large ICUs that provide specialized care. Although hospital 8 does not have a large ICU, it is a major regional centre for critical care. In line with my expectations, the model accurately predicted the occupancy for ICU 1 upon incorporating medical transfer data, acknowledging that patients in this category often have prolonged stays, significantly influencing occupancy rates. Specifically, as the primary care center, hospital 1's ICU continuously accommodates patients needing specialized care, irrespective of capacity constraints Conversely, the predictions for Hospitals 2 and 3 were overestimated, a discrepancy attributable to missing of repatriation transfer data for many patients. Typ-

ICU Occupancy Comparison		
ICU Site	Mean ICU 2019 Data	Updated Simulation
1	31.8	31.5
2	16.8	17.3
3	8.9	9.5
4	15.2	14.4
5	18.5	18.4
6	9.1	9.1
7	7.8	8.2
8	9.5	8.3

Table 5.2: Mean ICU occupancy (ICU and Overflow beds) and simulation results comparison with 2019 ICU occupancy from 8 major hospitals in British Columbia.

ically, patients transferred for medical reasons are returned to their original home ICU following the necessary treatment. However, the critical care database lacked substantial information on these repatriation transfers, leading to the observed overestimation in the model's predictions for these hospitals.

Another contributing factor to the overestimation was the uniform application of the transfer matrix across all types of medical patients. This approach overlooks the variability in medical transfer reasons, as each hospital is equipped with specialized facilities catering to specific medical conditions, such as cardiac issues. Table 5.2 shows that including transfer reasons model ICU occupancy are close to mean ICU occupancy 2019 data. This is due to most medically transferred patients moving to these major ICUs and their prolonged stays impacting occupancy levels substantially. Table 5.3 summarizes the ICU simulation validation, and ICU site 1 shows great accuracy, as it consistently accepts patients requiring specialized care, regardless of capacity limits and it is the biggest hospital in the BC. The validation test calculates the percentage difference between actual and simulated ICU occupancy across various sites, quantifying the model's prediction accuracy in relative error terms. This analysis, summarized in a validation table, highlights how the simulation either underestimates or overestimates actual occupancy, using a straightforward statistical approach known as percentage error calculation. This method, devoid of complex statistical techniques, evaluates the deviation of model predictions from real data, where positive values signal underestimation and negative values denote overestimation. This approach provides a clear, interpretable measure of the model's accuracy through percentage differences, effectively assessing its performance.

Details of ICU and Overflow Bed Usage

Due to its status as one of BC's largest and its high ICU bed demand, ICU site 1 is under closer observation to analyze its operational dynamics. Figure 5.1 illustrates the use of ICU and overflow beds, capturing 10 data points each hour (240 in total). The red line

Simulation Model Validation	
ICU Site	Validation
1	0.94%
2	-2.97%
3	-6.74%
4	5.26%
5	0.54%
6	0%
7	-5.12%
8	12.90%

Table 5.3: Results of simulation validation for eight major hospitals based on deviation of model predictions from real data, where positive values denote underestimation and negative values denote overestimation.

indicates the ICU's bed capacity at site 1. Data points exceeding this line signify the need for overflow beds or, if unavailable, patient transfer to HAU to accommodate new admissions. Figure 5.2 separately depicts the counts of ICU and overflow beds. Denying a patient ICU admission inappropriately lowers their survival and recovery chances due to reduced nursing and medical support, increasing the risk of poor recovery and mortality.

Figure 5.1 illustrates that when patient numbers exceed ICU bed capacity, overflow beds are utilized to accommodate additional patients, preventing deaths and the rejection of new admissions, and hospitals in the real world favor retaining patients in ICU beds over utilizing overflow beds to avoid added pressure on nurses and ICU resources, potentially impacting ICU efficiency. Figure 5.2 shows the comparison of the ICU and overflow bed for each data point. To provide a comprehensive analysis of ICU occupancy across different hospitals, I have documented the ICU occupancy rates using 10 data points for every hour of the day. This detailed representation can be found in Figure 5.3. This approach ensures a understanding of occupancy trends and variations throughout each hospital site.

In healthcare systems, resorting to overflow beds or moving patients from the ICU to the HAU is not the preferred strategy, as these measures can prolong treatment times and potentially diminish the standard of care patients receive, reflecting real-world consequences. As demonstrated in Figure 5.4, a more effective approach involves managing ICU demand spikes by temporarily accommodating patients in overflow beds instead of the ICU (as seen at 15:00 p.m.). This strategy prevents the immediate need to bump patients, preserving the continuity of their care. Once a patient completes their treatment, and an ICU bed is free, patients in overflow can be transitioned back to the proper ICU setting (evident at 16:00 p.m.), thereby optimizing bed usage and ensuring patients receive the appropriate level of care. This cyclical adjustment not only maintains a balanced flow of patient care but also minimizes the strain on critical healthcare resources.

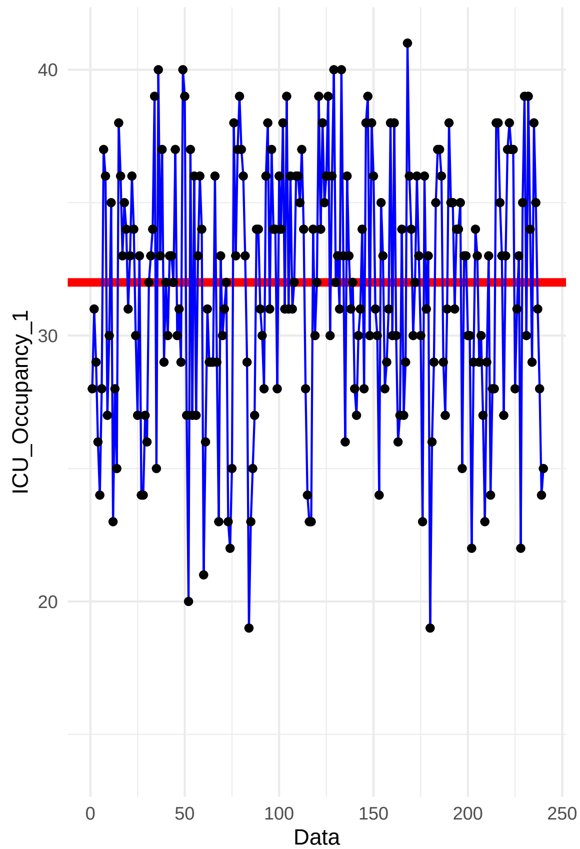
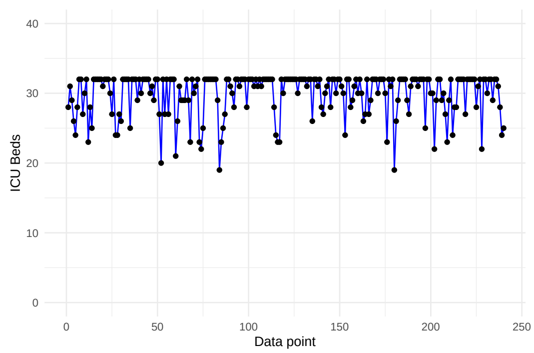
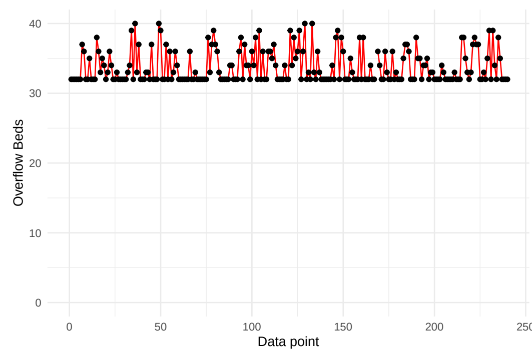


Figure 5.1: Number of ICU and overflow bed usage hourly for 10 data points for each hours of the day, for ICU site 1. Red line shows the fixed number of ICU beds in ICU site 1.



(a) ICU bed usage



(b) Overflow bed usage

Figure 5.2: Number of ICU and overflow beds used over time. Data is shown for site 1 with 10 data points per hour. Overflow bed is used when there is no available ICU bed, therefore if numbers of ICU bed exceed the fixed ICU bed we will have overflow.

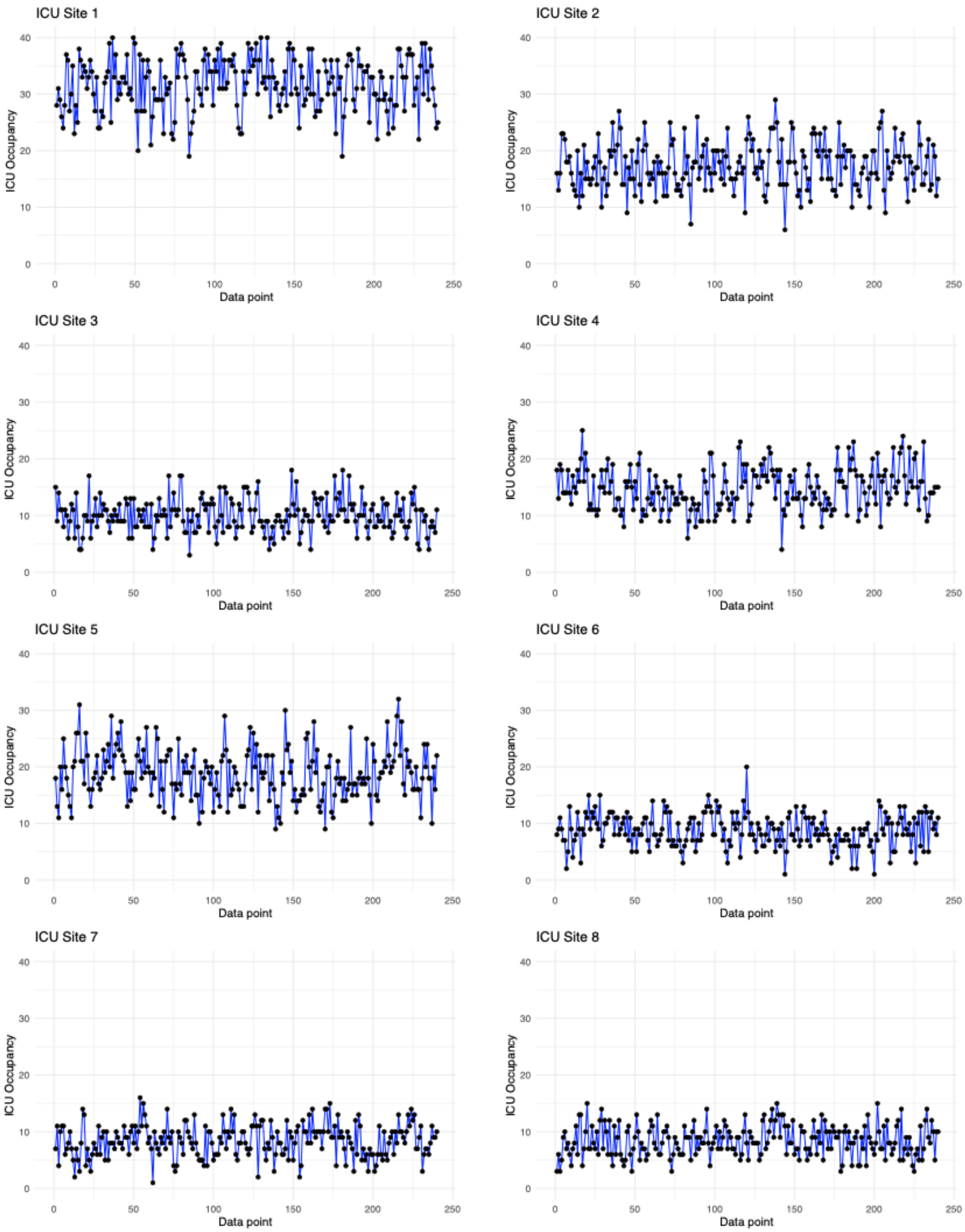


Figure 5.3: ICU occupancy is recorded through 10 data points hourly, for each ICU site

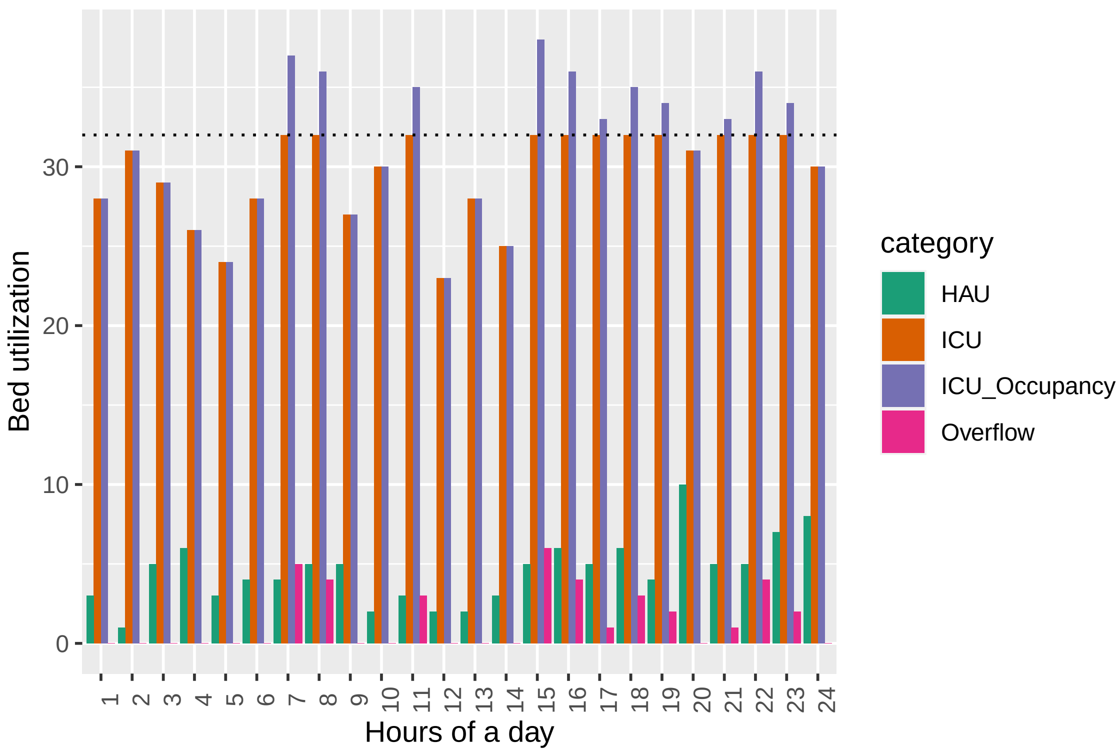


Figure 5.4: Hourly demand of different bed types in the critical care model. Dotted line shows that when ICU occupancy is exceed from the fixed ICU beds and they need to use overflow bed for new patients

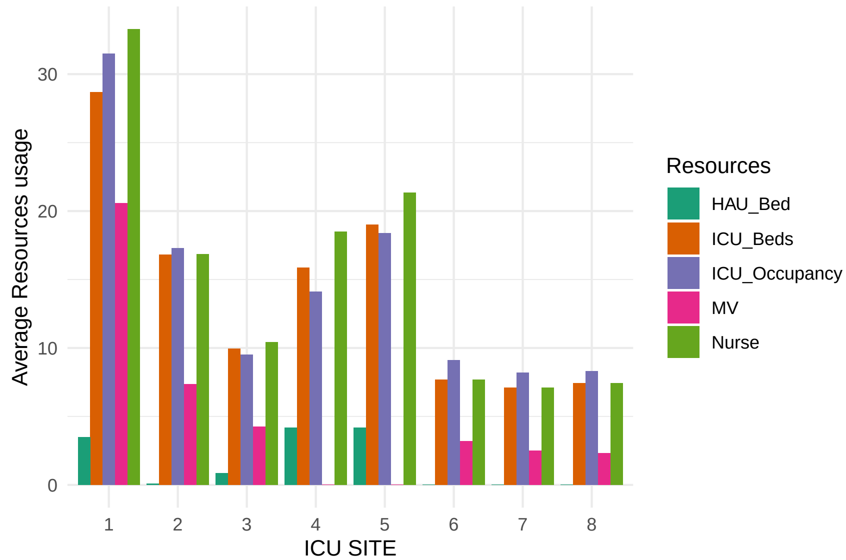


Figure 5.5: Average different resource utilization in each ICU site

ICU Resources Usage

In an effort to detail resource allocation within ICUs, each category, including ICU beds, overflow and HAU beds, mechanical ventilation, and nursing staff, has been distinctly analyzed. Despite a uniform pool of nurses servicing both HAU and ICU across all sites, patient-to-nurse ratios vary significantly. As depicted in Figure 5.5, the available nursing staff effectively meets the requirements posed by the total ICU occupancy (encompassing ICU and overflow beds) and HAU beds. It is important to note that for ICU sites 6, 7, and 8, there is a lack of transfers from ICU to HAU during patient stays, evidenced by the zero usage of HAU beds. This figure does not account for patients discharged from the ICU directly to the HAU.

Over Occupancy in ICU

The 2016 Canadian Institute for Health Information report revealed that ICU occupancy rates averaged 86% across Canada, with figures reaching 90% in major urban and teaching hospitals. Such data highlight instances where occupancy nears full capacity, especially during peak times like the winter influenza season or the trauma-heavy spring and summer months. Studies have linked high ICU occupancies, specifically those exceeding 80%, with increased rates of ICU and hospital mortality, as well as higher chances of ICU readmission within a week post-discharge. Consequently, recommendations suggest maintaining ICU occupancy rates below 80% to adequately accommodate demand surges. I conducted a detailed analysis of ICU occupancy rates, focusing specifically on instances where occupancy exceeded 90% across various hospitals in the Table 5.4. ICU site 1 plays a crucial role in BC healthcare system, and most transfer patients will transfer to this ICU site. Due to the

ICU Site	2019 Percentage of time over 90% capacity	Simulation Percentage of time over 90% capacity – No bumping or transfer	Simulation Percentage of time over 90% capacity – bumping or transfer
1	92.40	70.8	71.93
2	0	0.08	0.08
3	0	0.08	0.04
4	0	0.2	0.08
5	0	0	0.04
6	0	0	0
7	0	6.2	5.91
8	0.24	8.7	6.4

Table 5.4: ICU Occupancy over 90% capacity and simulation results comparison with 2019 ICU data for 8 major hospitals in British Columbia

Total ICU mortality rate in each ICU site	
ICU Site	Mortality rate
1	17.65%
2	13.94%
3	17.86%
4	20.22%
5	19.76%
6	19.6%
7	12.79%
8	15.55%

Table 5.5: Simulation-based mortality rates across hospitals

elevated rate of patient arrivals and the absence of a refusal policy at this ICU facility, the ICU operates at over 90% of its capacity in 71.93% of time. which means a lot of patients need to receive a bed at ICU.

5.3 Mortality Rates

The model recognizes multiple critical care mortality factors such as bed unavailability, insufficient ventilators, and critical health deterioration, detailed in Table 5.5, which lists total mortality rates per ICU site. Total mortality for all ICU sites is 17.37%. At ICU site 1, I analyzed mortality reasons in greater detail, as shown in Table 5.6. Medical transfers (HLC) are crucial in this model as they involve critically ill patients, and the model aims to prevent their mortality when they are transferred to a new ICU. Given that most medical transfers go to ICU site 1, I focus on examining these cases in greater detail. Mortality due to bed unavailability for HLC patients is zero, assuming the ideal scenario where all HLC patients receive a bed, thus preventing any deaths due to bed shortage.When patients can't receive necessary ventilation, it can lead to mortality. Refer to Table 5.6 for the mortality rate due to ventilation unavailability for HLC patients.

In the model, bed availability is prioritized over ventilation, ensuring that most patients who receive a bed and need ventilation will indeed receive it. Consequently, the primary cause of patient mortality is attributed to bed unavailability rather than a lack of ventilation.

Each mortality reasons with rates	
Mortality reason	Rate
Total HLC patient mortality (No Vent)	0.003
Total HLC patient mortality (No Beds)	0.000
Total Mortality Non HLC patients (No Beds)	0.007
Total Mortality Non HLC (No Vent)	0.002
Total Mortality Rate for all ICU sites	0.176

Table 5.6: Different mortality rates in ICU site 1

5.4 Transfer Reasons Comparison

Most of the medical reason transfers occur from smaller or rural hospitals to bigger and specialist hospitals, and from Table 5.7 you can see the results match with this assumption as you can see most of the Medical reasons transfer to ICU site 1,2,3. Repatriation transfer happens when patients receive the care they need in the bigger hospitals and when their health condition become stable they can transfer back to the local ICU site, most of the repatriation comes from bigger hospitals to smaller hospitals, you can find it in more details in Table 5.7. Another reason for transfers is when the ICU reaches capacity, necessitating the transfer of patients to a new site. This occurs when there is high demand for admission to ICU site 1, leading to capacity-related transfers.

Different transfer reason rate in each ICU site			
ICU Site	Capacity Transfer From	Medical Reason to	Repatriation Transfer from
1	0.019	0.018	0.006
2	0	0.02	0.003
3	0	0.016	0.01
4	0	0	0
5	0	0.0002	0
6	0	0.013	0.002
7	0	0.003	0.007
8	0.005	0	0

Table 5.7: Different transfer rate comparison based on the simulation result

I conducted a detailed comparison between the rates of medical reason transfers, (HLC), as recorded in the Critical Care Database and the outcomes predicted by our simulation model. This comparison is shown in Table 5.8, which serves as a critical reference point for understanding the alignment between observed data and simulated predictions regarding patient transfers for medical reasons across different ICU sites. It is important to note that the Critical Care Database does not uniformly provide data on the reasons for patient transfers at all ICU locations. In instances where the database lacks information on the transfer reason for a specific ICU site, the analysis pragmatically accounts for these gaps by recording the transfer rate as zero.

Medical Reason Transfer (HLC)		
ICU Site	Medical Reason (HLC) 2019 Data	Medical Reason (HLC) Transfer simulation
1	0.03	0.018
2	0.04	0.02
3	0.06	0.016
4	0	0
5	0.002	0.0002
6	0.03	0.013
7	0.006	0.003
8	0	0

Table 5.8: Medical reason transfer rate 2019 ICU comparison with simulation results eight major hospitals in British Columbia

5.5 Enhancement of Results

The based simulation [102] closely matched the Critical Care Database, with a variance of roughly 4% for ICU site 1. In model with various of error transfer reasons, enhanced accuracy for larger ICU sites, narrowing the discrepancy to approximately 0.78% for ICU site 1. This refinement indicates the simulated mean ICU occupancy aligns almost precisely with the actual data, in Table 5.9 and Figure 5.6. The validation test encompassing all ICU sites is comprehensively illustrated in Figure 5.7.

As shown in Table 5.10, the updated model predicts a significantly higher percentage of time over 90% ICU capacity compared to the based model especially for the hospital 1. This difference is attributed to the updated model's inclusion of more comprehensive transfer reasons across a wider hospital network. Note that minimizing hospital overcapacity is crucial for effective emergency surge preparedness to avoid rejecting new patients.

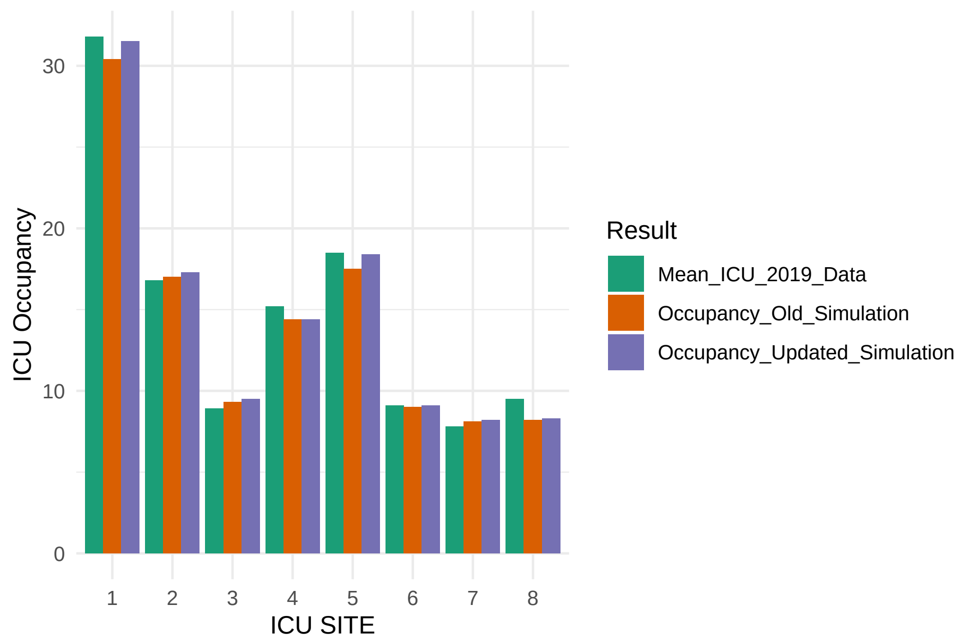


Figure 5.6: Compression of Simulation Results and Mean ICU occupancy 2019 Data

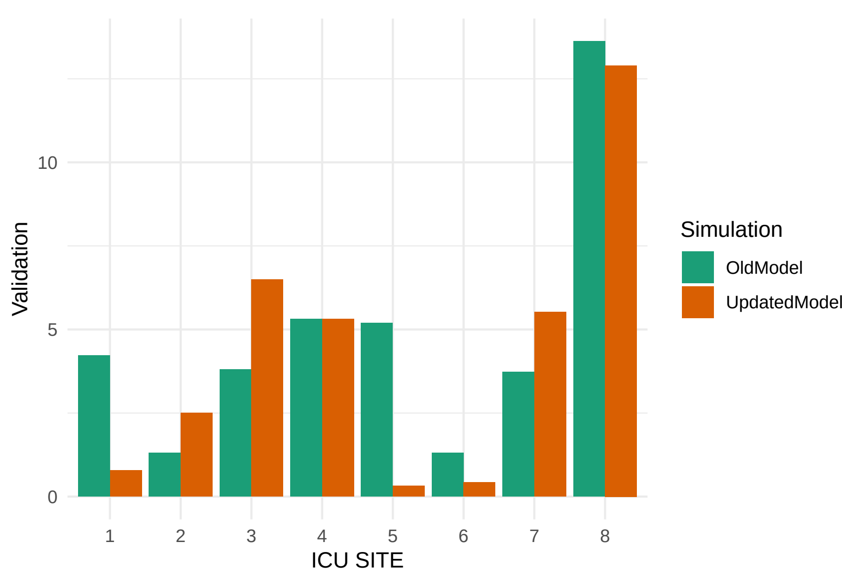


Figure 5.7: Error Validation of old and updated simulation models for each ICU site

ICU Occupancy comparison			
ICU Site	Mean ICU 2019 Data	Based Simulation	Updated Simulation
1	31.8	30.4	31.5
2	16.8	17	17.3
3	8.9	9.3	9.5
4	15.2	14.4	14.4
5	18.5	17.5	18.4
6	9.1	9	9.1
7	7.8	8.1	8.2
8	9.5	8.2	8.3

Table 5.9: Mean ICU Occupancy (ICU and overflow beds) and simulation results comparison with 2019 ICU occupancy from 8 Major Hospitals in British Columbia

ICU Site	2019 Percentage of time over 90% capacity	Simulation Percentage of time over 90% capacity (Based Model) – bumping or transfer	Simulation Percentage of time over 90% capacity (Updated Model) – bumping or transfer
1	92.40	68.4	71.93
2	0	0.04	0.08
3	0	0	0.04
4	0	0.04	0.08
5	0	0	0.04
6	0	0	0
7	0	6.83	5.91
8	0.24	7.63	6.4

Table 5.10: ICU occupancy over 90% capacity and simulation results comparison for based and updated model

Chapter 6

Discussion and conclusions

In British Columbia, the critical care system was already operating at or beyond its capacity before the COVID-19 pandemic. This situation made managing demand surges during the pandemic challenging, underscoring the importance of operational planning for future critical care demand peaks. In this thesis, I developed a simulation model of British Columbia's critical care network by considering three different transfer reasons incorporating their priorities between hospitals to aid in planning for various scenarios, as proposed by the British government Columbia Centre for Disease Control (BCCDC). Despite the difficulty in forecasting future pandemic waves, this model could be instrumental in devising strategies to manage the impacts of new pandemics, like seasonal influenza, without relying heavily on extensive public health measures to limit transmission. The simulation modeling results are valuable for managing ICU resources, analyzing scenarios, and determining mortality rates.

6.1 Discussion of Results

HAUs are becoming increasingly vital in the critical care system, offering a cost-effective way to boost capacity during regular operations and providing additional surge capacity in a crisis. This simulation model seeks to optimize policies for incorporating HAUs into the critical care system, emphasizing their most effective utilization across patient pathways, including step-down, step-up, and full-duration critical care units.

The pandemic led to the addition of extra critical care beds in hospitals, putting immense pressure on ICU staff and resulting in extensive overtime for ICU nurses. Consequently, the mental health of many critical care nurses deteriorated, with post traumatic stress disorder rates among nurses rising from $9-20\%$ before the pandemic to $49-73\%$ during it [20]. Consultations with critical care experts in British Columbia highlighted the need to include medical staff in our simulation model. This integration is believed to enhance the model's utility for future contingency planning, particularly for situations where ICU nurses and other medical staff are transferred between hospitals to manage surges in critical care demand within a network. The model can be used to determine early warning signals for both

COVID-19 and seasonal influenza or upcoming pandemic that could trigger preparation for these contingency measures. Therefore in this model I consider the nurse pool resource in the ICU and HAU to model the critical resources. Also, mechanical ventilation is a key resource, as most ICU patients require it during their stay. The model also accounts for the cleaning time needed after each use of a ventilator. The model is capable of tracking multiple ventilator uses within a single ICU stay.

Previous critical care simulation models mainly focused on single ICUs or hospitals. This model, by encompassing a network of ICUs, enables the examination of scenarios where demand surges in a specific hospital or region can be managed with support from other network hospitals. This approach is particularly relevant in British Columbia, where COVID-19 case peaks were not uniform across the province. This model evaluates the critical care system's capacity to operate as an integrated network, wherein ICUs outside an outbreak region offer additional support to those directly affected. Integrating medical transfers into the model was challenging, as it necessitates simulating aspects of medical decision-making. The model captures three different transfer reasons. The most important transfer reason is medical reason transfer. Patients transferred for medical reasons often need extended ICU stays to recover from critical illnesses. Estimates suggest that $2-11\%$ of critically ill patients require such prolonged ICU care, accounting for $25-45\%$ of total ICU days and consuming a substantial portion of resources [113, 4]. This model uses multiple ICU sites therefore, the model is able to capture medical reason transfer between ICU sites. Patients who required long stay in ICU during their hospital stay comprised only a small proportion of total ICU admissions but they can affect ICU occupancy especially if a new pandemic happens like COVID-19 which brings a high pressure in ICU sites.

In a scenario where an ICU is full and must admit a new patient, "bumping" leads to premature discharge of a current patient, potentially during off-hours. This increases the risk of post-discharge mortality by up to 39%, particularly with nighttime discharges. Bumping in ICUs, where a critically ill patient is refused admission, compromises patient safety due to the lower survival chances when they are cared for in areas without intensive nursing or specialized organ support [98, 37]. In critical care, bumping is a non-ideal but sometimes necessary practice. Patients first compete for ICU beds, then for overflow beds. Bumping is used only as a last resort to make room for new admissions, aiming to minimize its occurrence except during a pandemic, as reflected in the results.

In the original critical care system simulation, ICU occupancy was underestimated for larger hospitals. By incorporating medical transfers and other transfer factors, the simulation's ICU occupancy estimates became more aligned with actual ICU occupancy levels.

6.2 Limitation and Future Works

Future research possibilities may involve integrating a reduction in nurse-to-patient ratios for individuals approaching the conclusion of their treatment. This adjustment could occur in response to new admissions, allowing for increased overflow capacity, or it could be an automatic reduction when patients are deemed ready to leave the ICU and transition to Avoidable or Alternate-Level-of-Care status. In practical settings, nurses are designated either to the ICU or the HAU. Hence, it is essential to maintain separate nurse pools for future assignments. The model does not account for direct admissions to HAU or transfers from HAU to ICU.

In the real world it is possible critical ill patients need to receive higher level of care either in ICU or HAU but this model only considers admission in the ICU and it did not consider direct admision in HAU to receiving close monitoring during their stay. While the model includes the step-down unit post-ICU LOS, it overlooks the step-up process, particularly the transfer of patients from HAU to ICU, which predominantly occurs for patients requiring ventilation due to the unavailability of ventilators in the HAU unit. Ventilation resource is one of the important resources in the hospital and this model currently captures up to five fold ventilation usage during a patient's ICU stay, with potential for further increase in future work and it can improve the model. I considered the bumping from ICU to step down unit but bumping can extend a patient's HAU stay due to ICU transfer and HAU's limited monitoring; however, this model doesn't account for this effect. The model does not incorporate any explicit avoidable time within the HAU as patients may wait for a ward bed before HAU discharge. I treated each ICU visit as independent and didn't consider multiple ICU visits within a single hospitalization.

The VP or ARDS sub-group isn't split by admission due to low patient numbers in each ICU. Future analysis might explore other patient characteristics and expand the patient categories to have more specilize categories such as heart, tumor, etc. This can be helpful as some medical transfer patient in a specific type of patient will transfer to specilize ICU site. In the limitations of my study, while the gamma distribution applied to pre-ventilation times seemed reasonably fitting, it was not an exact match. This leaves room for potential refinement. Subsequent studies might benefit from employing goodness of fit tests to validate the chosen distribution and exploring alternative potentially to be more precise distribution fits to enhance accuracy.

Some mechanical ventilation might be under different codes in the datasets; future research should examine this, and consider them in the model. For ventilation time, I assume each mechanical ventilation instance is independent and identically distributed, even within a single ICU visit. This assumption may not be precise, and future research should delve into the interrelations of subsequent ventilation.

For medical reason transfers and repatriation transfers, the model should account for avoidable time since patients may experience delays in obtaining transfer facilities like ambulances and ICU beds at the new ICU, resulting in waiting times.

6.3 Conclusions

The analysis in this thesis provide a new approach to model critical care system based on various real world scenarios with considering the network between different hospitals in British Columbia. This model incorporates distinct features, including unique state distributions, patient types with specific characteristics, and resource sharing with varied priorities per patient type. By mapping the network between hospitals, it facilitates patient transfers, minimizing the rejection of critically ill patients through the accommodation of diverse scenarios. This model represents a highly accurate version of BC's critical care system. The ICU occupancy rates, derived from simulation outcomes and the 2019 dataset, closely match, indicating significant improvement in model validation—especially in comparison to previous models that either focused on a single ICU or overlooked transfers within and between hospitals. This model facilitates scenario planning to reduce mortality rates caused by factors like bed and ventilator shortages.

The model aims to minimize patient bumping from ICU mid-stay, closely aligning results with the goal of keeping such occurrences near zero.

Bibliography

[1] Canadian hospital rates. `http://www.david-cummings.com/documents/canadian_hospital_rates.htm`. Accessed: 2010-09-30.

[2] Reneé Adomat and Carolyn Hicks. Measuring nursing workload in intensive care: an observational study using closed circuit video cameras. *Journal of Advanced Nursing*, 42(4):402–412, 2003.

[3] Andres Alban, Stephen E Chick, Oleksandra Lvova, and Danielle Sent. A simulation model to evaluate the patient flow in an intensive care unit under different levels of specialization. In *2020 Winter Simulation Conference (WSC)*, pages 922–933. IEEE, 2020.

[4] Yaseen Arabi, S Venkatesh, Samir Haddad, Abdullah al Shimemeri, and Salym al Mallik. A prospective study of prolonged stay in the intensive care unit: predictors and impact on resource utilization. *International Journal for Quality in Health care*, 14(5):403–410, 2002.

[5] Amr Arisha and Wael Rashwan. Modeling of healthcare systems: past, current and future trends. In *2016 Winter Simulation Conference (WSC)*, pages 1523–1534. IEEE, 2016.

[6] Stef Baas, Sander Dijkstra, Aleida Braaksma, Plom van Rooij, Fieke J Snijders, Lars Tiemessen, and Richard J Boucherie. Real-time forecasting of COVID-19 bed occupancy in wards and intensive care units. *Health Care Management Science*, 24:402–419, 2021.

[7] Jie Bai, Andreas Fügener, Jan Schoenfelder, and Jens O Brunner. Operations research in intensive care unit management: a literature review. *Health Care Management Science*, 21:1–24, 2018.

[8] René Bekker, Michiel uit het Broek, and Ger Koole. Modeling COVID-19 hospital admissions and occupancy in the Netherlands. *European Journal of Operational Research*, 304(1):207–218, 2023.

[9] MC Beumer, RM Koch, D Van Beuningen, AM OudeLashof, FL Van de Veerdonk, E Kolwijck, JG van der Hoeven, DC Bergmans, and CWE Hoedemaekers. Influenza virus and factors that are associated with ICU admission, pulmonary co-infections and ICU mortality. *Journal of Critical Care*, 50:59–65, 2019.

[10] Lawrence Brown, Noah Gans, Avishai Mandelbaum, Anat Sakov, Haipeng Shen, Sergey Zeltyn, and Linda Zhao. Statistical analysis of a telephone call center:

A queueing-science perspective. *Journal of the American Statistical Association*, 100(469):36–50, 2005.

[11] Arnold Buss and Ahmed Al Rowaei. A comparison of the accuracy of discrete event and discrete time. In *Proceedings of the 2010 Winter Simulation Conference*, pages 1468–1477. IEEE, 2010.

[12] Andres L Mora Carpio and Jorge I Mora. Ventilator management. In *StatPearls [Internet]*. StatPearls Publishing, 2023.

[13] Kirandeep Chahal and Tillal Eldabi. Hybrid simulation and modes of governance in UK healthcare. *Transforming Government: People, Process and Policy*, 5(2):143–154, 2011.

[14] Kirandeep Chahal, Tillal Eldabi, and Terry Young. A conceptual framework for hybrid system dynamics and discrete event simulation for healthcare. *Journal of Enterprise Information Management*, 26(1/2):50–74, 2013.

[15] Yin-Chi Chan, Eric WM Wong, Gavin Joynt, Paul Lai, and Moshe Zukerman. Overflow models for the admission of intensive care patients. *Health Care Management Science*, 21:554–572, 2018.

[16] Kyoung Won Cho, Seong Min Kim, Young Moon Chae, and Yong Uk Song. Application of queueing theory to the analysis of changes in outpatients' waiting times in hospitals introducing emr. *Healthcare Informatics Research*, 23(1):35–42, 2017.

[17] Marta Cildoz, Amaia Ibarra, and Fermin Mallor. Accumulating priority queues versus pure priority queues for managing patients in emergency departments. *Operations Research for Health Care*, 23:100224, 2019.

[18] Gordon M Clark. Use of polya distributions in approximate solutions to nonstationary m/m/s queues. *Communications of the ACM*, 24(4):206–217, 1981.

[19] Stefan Creemers, Mieke Defraeye, and Inneke Van Nieuwenhuyse. G-rand: A phase-type approximation for the nonstationary g (t)/g (t)/s (t)+ g (t) queue. *Performance Evaluation*, 80:102–123, 2014.

[20] Sarah Crowe, A Fuchsia Howard, and Brandi Vanderspank. The mental health impact of the COVID-19 pandemic on Canadian critical care nurses. *Intensive and Critical Care Nursing*, 71:103241, 2022.

[21] Christine SM Currie, John W Fowler, Kathy Kotiadis, Thomas Monks, Bhakti Stephan Onggo, Duncan A Robertson, and Antuela A Tako. How simulation modelling can help reduce the impact of COVID-19. *Journal of Simulation*, 14(2):83–97, 2020.

[22] Jimmie L Davis, William A Massey, and Ward Whitt. Sensitivity to the service-time distribution in the nonstationary erlang loss model. *Management Science*, 41(6):1107–1116, 1995.

[23] Mieke Defraeye and Inneke Van Nieuwenhuyse. Controlling excessive waiting times in small service systems with time-varying demand: an extension of the ISA algorithm. *Decision Support Systems*, 54(4):1558–1567, 2013.

[24] Mieke Defraeye and Inneke Van Nieuwenhuyse. Personnel scheduling in queues with time-varying arrival rates: Applications of simulation-optimization. *Uncertainty Management in Simulation-Optimization of Complex Systems: Algorithms and Applications*, pages 203–223, 2015.

[25] Mieke Defraeye and Inneke Van Nieuwenhuyse. Staffing and scheduling under nonstationary demand for service: A literature review. *Omega*, 58:4–25, 2016.

[26] Gregory Dobson, Hsiao-Hui Lee, and Edieal Pinker. A model of ICU bumping. *Operations Research*, 58(6):1564–1576, 2010.

[27] Yue Dong, Nicolas W Chbat, Ashish Gupta, Mirsad Hadzikadic, and Ognjen Gajic. Systems modeling and simulation applications for critical care medicine. *Annals of intensive care*, 2(1):1–10, 2012.

[28] Stephen G Eick, William A Massey, and Ward Whitt. The physics of the mt/g/ queue. *Operations Research*, 41(4):731–742, 1993.

[29] Tillal Eldabi, Ray J Paul, and Terry Young. Simulation modelling in healthcare: reviewing legacies and investigating futures. *Journal of the Operational Research Society*, 58:262–270, 2007.

[30] Agner Krarup Erlang. Sandsynlighedsregning og telefonsamtaler. *Nyt Tidsskrift for Matematik*, 20:33–39, 1909.

[31] Neil Ferguson, Daniel Laydon, Gemma Nedjati Gilani, Natsuko Imai, Kylie Ainslie, Marc Baguelin, Sangeeta Bhatia, Adhiratha Boonyasiri, ZULMA Cucunuba Perez, Gina Cuomo-Dannenburg, et al. Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand. 2020.

[32] Daniel Garcia-Vicuña, Laida Esparza, and Fermin Mallor. Hospital preparedness during epidemics using simulation: the case of COVID-19. *Central European Journal of Operations Research*, pages 1–37, 2022.

[33] 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.

[34] RT Noel Gibney, Cynthia Blackman, Melanie Gauthier, Eddy Fan, Robert Fowler, Curtis Johnston, R Jeremy Katulka, Samuel Marcushamer, Kusum Menon, Tracey Miller, et al. COVID-19 pandemic: the impact on canada's intensive care units, 2022.

[35] Jennifer Gillespie, Sally McClean, F FitzGibbons, Bryan Scotney, Frank Dobbs, and BJ Meenan. Do we need stochastic models for healthcare? the case of icats? *Journal of Simulation*, 8:293–303, 2014.

[36] Ken Junyang Goh, Jolin Wong, Jong-Chie Claudia Tien, Shin Yi Ng, Sewa Duu Wen, Ghee Chee Phua, and Carrie Kah-Lai Leong. Preparing your intensive care unit for the COVID-19 pandemic: practical considerations and strategies. *Critical Care*, 24(1):1–12, 2020.

[37] Caroline Goldfrad and Kathy Rowan. Consequences of discharges from intensive care at night. *The Lancet*, 355(9210):1138–1142, 2000.

[38] Giacomo Grasselli, Antonio Pesenti, and Maurizio Cecconi. Critical care utilization for the COVID-19 outbreak in lombardy, italy: early experience and forecast during an emergency response. *Jama*, 323(16):1545–1546, 2020.

[39] Linda Green and Peter Kolesar. The pointwise stationary approximation for queues with nonstationary arrivals. *Management Science*, 37(1):84–97, 1991.

[40] Linda V Green and Peter J Kolesar. On the accuracy of the simple peak hour approximation for markovian queues. *Management Science*, 41(8):1353–1370, 1995.

[41] Linda V Green and Peter J Kolesar. The lagged psa for estimating peak congestion in multiserver markovian queues with periodic arrival rates. *Management Science*, 43(1):80–87, 1997.

[42] Linda V Green, Peter J Kolesar, and Ward Whitt. Coping with time-varying demand when setting staffing requirements for a service system. *Production and Operations Management*, 16(1):13–39, 2007.

[43] Linda V Green, Joao Soares, James F Giglio, and Robert A Green. Using queueing theory to increase the effectiveness of emergency department provider staffing. *Academic Emergency Medicine*, 13(1):61–68, 2006.

[44] Jeff D Griffiths, Mari Jones, Martyn Sinclair Read, and Janet E Williams. A simulation model of bed-occupancy in a critical care unit. *Journal of Simulation*, 4:52–59, 2010.

[45] Jeff D Griffiths, Naomi Price-Lloyd, M Smithies, and Janet E Williams. Modelling the requirement for supplementary nurses in an intensive care unit. *Journal of the Operational Research Society*, 56(2):126–133, 2005.

[46] Murat M Gunal. A guide for building hospital simulation models. *Health Systems*, 1(1):17–25, 2012.

[47] Matthew S Hagen, Jeffrey K Jopling, Timothy G Buchman, and Eva K Lee. Priority queuing models for hospital intensive care units and impacts to severe case patients. In *AMIA Annual Symposium Proceedings*, volume 2013, page 841. American Medical Informatics Association, 2013.

[48] Shlomo Halfin and Ward Whitt. Heavy-traffic limits for queues with many exponential servers. *Operations Research*, 29(3):567–588, 1981.

[49] 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.

[50] Paul R Harper and HM Gamlin. Reduced outpatient waiting times with improved appointment scheduling: a simulation modelling approach. *OR Spectrum*, 25:207–222, 2003.

[51] Gavin H Harris, Kimberly J Rak, Jeremy M Kahn, Derek C Angus, Olivia R Mancing, Julia Driessen, and David J Wallace. Us hospital capacity managers' experiences and concerns regarding preparedness for seasonal influenza and influenza-like illness. *JAMA Network Open*, 4(3):e212382–e212382, 2021.

[52] Michael D Howell. Managing ICU throughput and understanding ICU census. *Current Opinion in Critical Care*, 17(6):626–633, 2011.

[53] Hsin-Chan Huang, Ozgur M Araz, David P Morton, Gregory P Johnson, Paul Damien, Bruce Clements, and Lauren Ancel Meyers. Stockpiling ventilators for influenza pandemics. *Emerging Infectious Diseases*, 23(6):914, 2017.

[54] Peter JH Hulshof, Nikky Kortbeek, Richard J Boucherie, Erwin W Hans, and Piet JM Bakker. Taxonomic classification of planning decisions in health care: a structured review of the state of the art in or/ms. *Health Systems*, 1:129–175, 2012.

[55] Armann Ingolfsson, Elvira Akhmetshina, Susan Budge, Yongyue Li, and Xudong Wu. A survey and experimental comparison of service-level-approximation methods for nonstationary m (t)/m/s (t) queueing systems with exhaustive discipline. *Informs Journal on Computing*, 19(2):201–214, 2007.

[56] Armann Ingolfsson, Fernanda Campello, Xudong Wu, and Edgar Cabral. Combining integer programming and the randomization method to schedule employees. *European Journal of Operational Research*, 202(1):153–163, 2010.

[57] Armann Ingolfsson, Md Amanul Haque, and Alex Umnikov. Accounting for time-varying queueing effects in workforce scheduling. *European Journal of Operational Research*, 139(3):585–597, 2002.

[58] Sheldon H Jacobson, Shane N Hall, and James R Swisher. Discrete-event simulation of health care systems. *Patient Flow: Reducing delay in healthcare delivery*, pages 273–309, 2013.

[59] David L Jagerman. Nonstationary blocking in telephone traffic. *Bell System Technical Journal*, 54(3):625–661, 1975.

[60] Marc Jegers, DL Edbrooke, C Hibbert, D Chalfin, and Hilmar Burchardi. Definitions and methods of cost assessment: an intensivist's guide. *Intensive Care Medicine*, 28:680–685, 2002.

[61] Taylor Kain and Robert Fowler. Preparing intensive care for the next pandemic influenza. *Critical Care*, 23(1):1–9, 2019.

[62] Seung-Chul Kim, Ira Horowitz, Karl K Young, and Thomas A Buckley. Analysis of capacity management of the intensive care unit in a hospital. *European Journal of Operational Research*, 115(1):36–46, 1999.

[63] Patrick Koelling and Michael J Schwandt. Health systems: A dynamic system-benefits from system dynamics. In *Proceedings of the Winter Simulation Conference, 2005.*, pages 7–pp. IEEE, 2005.

[64] Ali Kokangul, Serap Akcan, and Mufide Narli. Optimizing nurse capacity in a teaching hospital neonatal intensive care unit. *Health Care Management Science*, 20:276–285, 2017.

[65] Harold Joseph Kushner and HJ Kushner. *Heavy traffic analysis of controlled queueing and communication networks*, volume 28. Springer, 2001.

[66] LC Lamb, MM DiFiori, J Calafell, CH Comey, DS Shapiro, and JM Feeney. Is a step-down unit safe for patients with mild traumatic intracranial hemorrhages? *European Journal of Trauma and Emergency Surgery*, 44:897–901, 2018.

[67] Averill M Law, W David Kelton, and W David Kelton. *Simulation modeling and analysis*, volume 3. Mcgraw-hill New York, 2007.

[68] Antoine Legrain, Hocine Bouarab, and Nadia Lahrichi. The nurse scheduling problem in real-life. *Journal of Medical Systems*, 39:1–11, 2015.

[69] Qianying Lin, Shi Zhao, Daozhou Gao, Yijun Lou, Shu Yang, Salihu S Musa, Maggie H Wang, Yongli Cai, Weiming Wang, Lin Yang, et al. A conceptual model for the coronavirus disease 2019 (COVID-19) outbreak in wuhan, china with individual reaction and governmental action. *International Journal of Infectious Diseases*, 93:211–216, 2020.

[70] Nelly Litvak, Marleen Van Rijsbergen, Richard J Boucherie, and Mark van Houdenhoven. Managing the overflow of intensive care patients. *European Journal of Operational Research*, 185(3):998–1010, 2008.

[71] Yunan Liu and Ward Whitt. A network of time-varying many-server fluid queues with customer abandonment. *Operations Research*, 59(4):835–846, 2011.

[72] Yunan Liu and Ward Whitt. The g t/gi/st+ gi many-server fluid queue. *Queueing Systems*, 71:405–444, 2012.

[73] Julie C Lowery. Multi-hospital validation of critical care simulation model. In *Proceedings of the 25th Conference on Winter Simulation*, pages 1207–1215, 1993.

[74] Yiruo Lu, Yongpei Guan, Xiang Zhong, Jennifer N Fishe, and Thanh Hogan. Hospital beds planning and admission control policies for COVID-19 pandemic: a hybrid computer simulation approach. In *2021 IEEE 17th International Conference on Automation Science and Engineering (CASE)*, pages 956–961. IEEE, 2021.

[75] Fermín Mallor and Cristina Azcárate. Combining optimization with simulation to obtain credible models for intensive care units. *Annals of Operations Research*, 221:255–271, 2014.

[76] Avi Mandelbaum and William A Massey. Strong approximations for time-dependent queues. *Mathematics of Operations Research*, 20(1):33–64, 1995.

[77] Avishai Mandelbaum and Sergey Zeltyn. Service engineering in action: the palm/erlang-a queue, with applications to call centers. In *Advances in Services Innovations*, pages 17–45. Springer, 2007.

[78] Avishai Mandelbaum and Sergey Zeltyn. Staffing many-server queues with impatient customers: Constraint satisfaction in call centers. *Operations Research*, 57(5):1189–1205, 2009.

[79] Deborah A Marshall, Lina Burgos-Liz, Maarten J IJzerman, Nathaniel D Osgood, William V Padula, Mitchell K Higashi, Peter K Wong, Kalyan S Pasupathy, and William Crown. Applying dynamic simulation modeling methods in health care delivery research—the simulate checklist: report of the ispor simulation modeling emerging good practices task force. *Value in Health*, 18(1):5–16, 2015.

[80] William A Massey and Ward Whitt. An analysis of the modified offered-load approximation for the nonstationary erlang loss model. *The Annals of Applied Probability*, pages 1145–1160, 1994.

[81] Hamish DD Meares and Michael P Jones. When a system breaks: queueing theory model of intensive care bed needs during the COVID-19 pandemic. *The Medical Journal of Australia*, 212(10):470, 2020.

[82] Jyotiprasad Medhi. *Stochastic processes*. New Age International, 1994.

[83] Jyotiprasad Medhi. *Stochastic models in queueing theory*. Elsevier, 2002.

[84] Sanjay Mehrotra, Hamed Rahimian, Masoud Barah, Fengqiao Luo, and Karolina Schantz. A model of supply-chain decisions for resource sharing with an application to ventilator allocation to combat COVID-19. *Naval Research Logistics (NRL)*, 67(5):303–320, 2020.

[85] Martin I Meltzer, Anita Patel, Adebola Ajao, Scott V Nystrom, and Lisa M Koonin. Estimates of the demand for mechanical ventilation in the united states during an influenza pandemic. *Clinical Infectious Diseases*, 60(suppl_1):S52–S57, 2015.

[86] Anish R Mitra, Haljan Gregory, Jennifer P Stevens, et al. How the high acuity unit changes mortality in the intensive care unit: a retrospective before-and-after study. *Canadian Journal of Anesthesia*, 67(11):1507–1514, 2020.

[87] Rym M'Hallah and AH Al-Roomi. The planning and scheduling of operating rooms: A simulation approach. *Computers & Industrial Engineering*, 78:235–248, 2014.

[88] Thomas W Noseworthy, Elsie Konopad, Allan Shustack, Richard Johnston, and Michael P Eng Grace. Cost accounting of adult intensive care: methods and human and capital inputs. *Critical Care Medicine*, 24(7):1168–1172, 1996.

[89] John Paul Oddoye, Dylan F Jones, Mehrdad Tamiz, and P Schmidt. Combining simulation and goal programming for healthcare planning in a medical assessment unit. *European Journal of Operational Research*, 193(1):250–261, 2009.

[90] 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.

[91] Shanshan Qiu, Ratna Babu Chinnam, Alper Murat, Bassam Batarse, Hakimuddin Neemuchwala, and Will Jordan. A cost sensitive inpatient bed reservation approach to reduce emergency department boarding times. *Health Care Management Science*, 18:67–85, 2015.

[92] Sadeem Munawar Qureshi, Nancy Purdy, Asad Mohani, and W Patrick Neumann. Predicting the effect of nurse–patient ratio on nurse workload and care quality using discrete event simulation. *Journal of Nursing Management*, 27(5):971–980, 2019.

[93] Sabrina Ramwadhdoebe, Erik Buskens, Ralph JB Sakkers, and James E Stahl. A tutorial on Discrete-Event Simulation for health policy design and decision making: Optimizing pediatric ultrasound screening for hip dysplasia as an illustration. *Health Policy*, 93(2-3):143–150, 2009.

[94] John Rapoport, Daniel Teres, Yonggang Zhao, and Stanley Lemeshow. Length of stay data as a guide to hospital economic performance for icu patients. *Medical Care*, pages 386–397, 2003.

[95] Tom W Reader, Geetha Reddy, and Stephen J Brett. Impossible decision? an investigation of risk trade-offs in the intensive care unit. *Ergonomics*, 61(1):122–133, 2018.

[96] Andrea Remuzzi and Giuseppe Remuzzi. Covid-19 and italy: what next? *The lancet*, 395(10231):1225–1228, 2020.

[97] JC Ridge, SK Jones, MS Nielsen, and AK Shahani. Capacity planning for intensive care units. *European Journal of Operational Research*, 105(2):346–355, 1998.

[98] René Robert, Jean Reignier, Caroline Tournoux-Facon, Thierry Boulain, Olivier Lesieur, Valérie Gissot, Vincent Souday, Mouldi Hamrouni, Cécile Chapon, and Jean-Paul Gouello. Refusal of intensive care unit admission due to a full unit: impact on mortality. *American Journal of Respiratory and Critical Care Medicine*, 185(10):1081–1087, 2012.

[99] Felipe F Rodrigues, Gregory S Zaric, and DA Stanford. Discrete event simulation model for planning level 2 “step-down” bed needs using nems. *Operations Research for Health Care*, 17:42–54, 2018.

[100] Sheldon M Ross. *Simulation*. Academic Press, 2022.

[101] Michael H Rothkopf and Shmuel S Oren. A closure approximation for the nonstationary m/m/s queue. *Management Science*, 25(6):522–534, 1979.

[102] Alexander R Rutherford, Samantha L Zimmerman, Mina Moeini, Rashid Barket, Steve Ahkioon, and Donald EG Griesdale. Simulation model of a multi-hospital critical care network. In *2022 Winter Simulation Conference (WSC)*, pages 951–960. IEEE, 2022.

[103] Kim Seung-Chul, Horowitz Ira, et al. Flexible bed allocation and performance in the intensive care unit. *Journal of Operations Management*, 18(4):427–443, 2000.

[104] Arjan K Shahani, SA Ridley, and MS Nielsen. Modelling patient flows as an aid to decision making for critical care capacities and organisation. *Anaesthesia*, 63(10):1074–1080, 2008.

[105] Tasnim Sinuff, Kamyar Kahnamoui, Deborah J Cook, John M Luce, Mitchell M Levy, et al. Rationing critical care beds: a systematic review. *Critical Care Medicine*, 32(7):1588–1597, 2004.

[106] David A Stanford, Peter Taylor, and Ilze Ziedins. Waiting time distributions in the accumulating priority queue. *Queueing Systems*, 77:297–330, 2014.

[107] Krisjanis Steins and SM Walther. A generic simulation model for planning critical care resource requirements. *Anaesthesia*, 68(11):1148–1155, 2013.

[108] Michael Thorwarth, Wael Rashwan, and Amr Arisha. An analytical representation of flexible resource allocation in hospitals. *Flexible Services and Manufacturing Journal*, 28(1-2):148–165, 2016.

[109] Marleen van Rijsbergen, Richard J Boucherie, Mark van Houdenhoven, and Nelly Litvak. Managing the overflow of intensive care patients. 2005.

[110] Josephine Varney, Nigel Bean, and Mark Mackay. The self-regulating nature of occupancy in ICUs: stochastic homoeostasis. *Health Care Management Science*, 22:615–634, 2019.

[111] Tao Wang, Alain Guinet, Aissam Belaidi, and Beatrix Besombes. Modelling and simulation of emergency services with aris and arena. case study: the emergency department of saint joseph and saint luc hospital. *Production Planning and Control*, 20(6):484–495, 2009.

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

[113] David T Wong, Manuel Gomez, Glenn P McGuire, and Brian Kavanagh. Utilization of intensive care unit days in a Canadian medical-surgical intensive care unit. *Critical Care Medicine*, 27(7):1319–1324, 1999.

[114] Hannah Wunsch, Jason Wagner, Maximilian Herlim, David H Chong, Andrew A Kramer, and Scott D Halpern. ICU occupancy and mechanical ventilator use in the united states. *Critical Care Medicine*, 41(12):2712–2719, 2013.

[115] Stephanie Zaza, Lisa M Koonin, Adebola Ajao, Scott V Nystrom, Richard Branson, Anita Patel, Bruce Bray, and Michael F Iademarco. A conceptual framework for allocation of federally stockpiled ventilators during large-scale public health emergencies. *Health Security*, 14(1):1–6, 2016.

[116] Zhecheng Zhu, Bee Hoon Hen, and Kiok Liang Teow. Estimating ICU bed capacity using discrete event simulation. *International Journal of Health care Quality Assurance*, 25(2):134–144, 2012.

[117] Samantha L Zimmerman, Alexander R Rutherford, Alexa van der Waall, Monica Norena, and Peter Dodek. A queuing model for ventilator capacity management during the COVID-19 pandemic. *Health Care Management Science*, pages 1–17, 2023.