10-19-2018 3:00 PM

Three Essays on Intensive Care Unit Capacity Planning

Felipe F. Rodrigues
The University of Western Ontario

Supervisor
Zaric, Gregory S.
The University of Western Ontario Co-Supervisor
Stanford, David A.
The University of Western Ontario

Graduate Program in Business
A thesis submitted in partial fulfillment of the requirements for the degree in Doctor of Philosophy

Abstract

Three Essays on Intensive Care Unit Capacity Planning

by Felipe Fontes Rodrigues

The Intensive Care Unit (ICU) is a resource-intensive, costly environment. Data gathered from patients during their stay in the ICU has traditionally been used for clinical purposes, but can have a significant impact on healthcare capacity planning and patient flow. There is a need to study how metrics collected in Canadian ICUs, such as the Multiple Organ Dysfunction Syndrome (MODS) score and the Nine Equivalents of Nursing Manpower Use Score (NEMS) can be used to improve capacity planning decisions. Using discrete-event simulation, statistical, survival and machine learning models, I have built long- and short-term capacity planning models to help hospital administrators better manage patient flows in the ICU. This dissertation consists of three essays that explore the use of these metrics in ICU capacity planning.

In the first essay, I study the incorporation of the nursing manpower score NEMS into a discrete-event simulation model to estimate optimal long-term capacity levels of critical care beds in both Level 3 (ICU) and Level 2 (step-down) units. Using data from London Health Sciences Centre (LHSC) University Hospital, I demonstrate the benefits of simulating patients' daily NEMS changes as triggers for transfer to a step-down unit. This essay also examines ways in which transfer to a step-down unit may improve patient length of stay (LOS), flow and costs.

In the second essay, I demonstrate that the ICU LOS literature shows the predominance of multiple linear regression models for individual patients' ICU LOS and outcome predictions (e.g., death, discharge, long stay). Using data from LHSC's two ICUs, I compare the performances of well known statistical models with contemporary supervised machine learning models in predicting such outcomes. I show that there is no dominant model in terms of individual patients' LOS predictions, but that outcome prediction (death, discharge, long stay) performance can be improved by using supervised machine learning techniques.

In the third essay, I build on the use of NEMS to simulate realistic ICU LOS for long-term capacity planning, and on the use of NEMS and MODS to predict individual ICU LOS in order to improve short-term capacity planning. First, I fit a parametric survival model called the Accelerated Failure Time (Weibull AFT) model with LHSC's UH data. Then I analyze the model's hazard rates, event time ratios and LOS, both at the time of the patient's arrival in the

ICU and after 3 days' stay. Finally, I generate daily patient survival probabilities and pool them to predict future expected ICU occupancy rates. Using survival probability pooling for short-term capacity planning is a novel use of the ATF model, and may be used to accurately predict ICU occupancy.

Keywords: Bed capacity planning, Patient flow, Step-down beds, NEMS, MODS, Length of stay, Discrete-event simulation, Prediction models, Parametric Survival, Accelerated Failure Time, Supervised Machine learning, Survival probability, Patient pooling.

Statement of Co-Authorship

I hereby declare that this thesis incorporates material that is a result of joint research.

Essay 1 was co-authored with Drs. Gregory S. Zaric and David A. Stanford.

Essay 2 was co-authored with Drs. Gregory S. Zaric and John G. Wilson.

Essay 3 was co-authored with Drs. Gregory S. Zaric and John G. Wilson.

As the first author, I was in charge of all aspects of these projects, including formulating research questions, literature review, research design, model formulation and analysis, and preparing the first and final versions of the manuscripts.

With the above exceptions, I certify that this dissertation and the research to which it refers, is fully my own.

This dissertation includes 3 original papers, the first of which was published in 2017 and is reproduced in chapter 2: F. Rodrigues, G.S. Zaric, D.A. Stanford, Discrete event simulation model for planning Level 2 “step-down” bed needs using NEMS, Operations Research for Health Care, 2017, doi.org/10.1016/j.orhc.2017.10.001.

Epigraph

“So that thou incline thine ear unto wisdom, [and] apply thine heart to understanding;”

Proverbs 2:2

"...That knowledge, grounded on accuracy, aided by labour, and promoted by perseverance,

will finally overcome all difficulties, raise ignorance from despair,

and establish happiness in the paths of science."

Book of the Work,

Grand Lodge of A.F. & A.M. of Canada in the Province of Ontario, 2017

Dedication

To my little dude Henrique, lovely wife Beatriz, baby sister Tatiana, my angels Sofia & Catarina, and wonderful parents Edmundo & Terezinha, Jimbo & Ann.

Acknowledgements

My PhD experience can be described as the initiation (not a culmination!) of a life-long search for wisdom—for which both retrospectively and prospectively, I still feel utterly unprepared. Nevertheless, I can endeavor to abstract many lessons from this experience, in however rough a form. In essence, I find the PhD to be an exercise in humility: years of being among uncommonly smart peers and top scholars has instilled in me a different perspective on life and on the personal stoicism and academic discipline required to produce intellectual masterpieces. In my own endeavors, I had to rely in faith that, ultimately, temperance and courage would prevail. The work presented here is an experiment in applied Management Science. Although it is mostly a reflection of my own foray into constructive academia, I have no doubt that the best elements of this experiment either originated from or were supported by the many people who helped and inspired me.

My most sincere appreciation goes to my supervisors, Dr. Greg Zaric and Dr. David Stanford. They have provided academic, professional and personal guidance in equal measure. They have been patient and understanding, have dedicated their time and effort and have taught me so much. Through their own example, they have shown me what it truly means to be an academic. I can only hope to be as dedicated to the advancement of knowledge, the betterment of my students and the development of my community as they are.

I would also like to thank Prof. John Wilson for showing so much interest in and commitment to my research it when I needed it the most, Prof. Fredrik Odegaard for his genuine interest in my progress throughout the doctoral program. Prof. Huber Pun for always making sure I was on the right track, Prof. Rasha Kashef for being such an exemplar of thoughtful teaching and hard work, Profs. Mehmet Begen and Lauren Cipriano for the helpful talks about life during and after the PhD. To all the other members of the Management Science group that have taught me so patiently, thank you for being so helpful.

Prof. June Cotte also played a significant role in my PhD experience. Without her nominating me for the Ontario Trillium Scholarship, I could not have come to Ivey. Thank you to Anna Potrawiak, Carly Vanderheyden, and Paola Ramgren, for having been so helpful and so professional. To Prof. Dave Barret and his team from Ivey Health (IICHI) I offer thanks for funding my research and providing such tremendous opportunities to present it to decision- and policy-makers over the years. On that note, I would also like to thank the Ontario Trillium

Foundation, NSERC, and my *alma-mater,* the Universidade Federal do Parana, particularly the Departamento de Administração Geral e Aplicada.

I thank Prof. Matt Thomson for all his guidance at the time of my arrival and his open-door policy during the following years. If it weren not for Prof. Thomson, I would not have met my dear friends and brethren: Harvey Morrison, Cam Murray, Matt Webster, Ben Dawson, Jaime Cousins, Harold Walker (*in memoriam*), and the fine gentlemen of the Delaware Valley Lodge #358, as well as Mike Steele, Ed Hartman, Jan Gortzak, Chris Henderson, and the brethren of Ashlar Lodge # 610 of the Grand Lodge of Canada (Province of Ontario). There are no words to adequately describe my thanks for their welcoming me into their Lodge, allowing me to continue my travels and helping my family and I to get settled in Canada. They embody the universal values of the Craft; through gavel, chisel and square, they have helped me work on the Rough Ashlar. I thank you sincerely and fraternally.

I extend my thanks to my dear friends (and their families) in Brazil and elsewhere around the globe, particularly: Eduardo Pecora Jr. (a true Brother) & Ana, Bernardo Luz & Priscila, Christiano Barros, João Gabriel & Tia Genoveva (thank you for the visit!), Eduardo Mayrink, Gustavo Fonseca, José Flavio Pedras, Felipe Bretas, Felipe Salim, Ricardo Grandinetti, Francisco Dornas, Ale Megale, Tios Alamar & Toê, Renato & Ruth (for helping me provide for my family during my early days in Canada), Tia Lindamar Peixoto, Tio Tadeu Leite (for being such an exemplar of scholarship and work ethic), Tio Paulo, Tia Cris, Tia Ana & Oswaldo (for the emotional support). From London: Piotrek & Ania, Salah & Fatima, Shaul & Ofer, Bjarni & Kitty, Torres & Ania, Stephen & Kristin, Jason & Harp. From the Brazilian community: Nicacio & Mari, Atila & Andre, Danilo, Carlos, Marco, David and so many others.

I'd also like to thank Dr. Grigori Erenburg, Dr. Sergio Janczak, Dr. Trevor Hunter and Dr. Jen Jeffrey from King's University College for trusting and believing in me, my colleagues Dr. Salar, Dr. Kyle, Dr. Jeff, Dr. Maya, Dr. Yamlak, Dr. Ram, Dr. Vivek, Dr. Dwarka, Dr. Mansour, Dr. Patrick, and soon-to-be Drs. John, Greg, Hongmei, Hadi, Anna, Nicole, Sampath, Fernando, Mirit, and others.

My family, of course, played a pivotal role during this time, and I am eternally grateful to my wonderful parents Edmundo & Terezinha. Their unwavering love and support has brought me here. To Jimbo & Ann Halden, please feel included in the category as well! To Mr. Walker (also a true Brother!) & D. Mazé, and to Tio Guga for all the help and encouragement. To my sisters Sofia & Catarina, I have faith you can see me from above…I miss you always.

I reserve a special thanks to three very special people:

- My baby sister Tatiana, who has shown me the lighter side of life. Being apart from you is particularly painful, and I hope we will see each other more often, very soon!

- My "little dude" Henrique: thank you for showing me every day the meaning of life. I hope you can forgive me for being so absent lately. I promise it will be worthwhile.

- My wife Beatriz: I am hopelessly grateful for your love, companionship, friendship, commitment to our family and tireless hard work. You have supported me emotionally over the past few years and please be sure that I adore you as my wife, admire you as a mother and respect you as a woman. Thank you so very much.

To conclude, I have come to believe that nothing in life is truly random. I have experienced many times the feeling that there is perhaps, faith, hope and love in this world. I have come to terms with the thought that there is an Architect behind all this, and for all I've been and will go through according to His/Her plan, I am truly grateful.

Table of Contents

[1] This work was supported by: The Ontario Trillium Scholarship program (OTS); The Ivey International Centre for Health Innovation (IICHI)/3M Canada MITACS grant; Natural Sciences and Engineering Research Council of Canada (NSERC), and the Universidade Federal do Parana (UFPR). These funding sources had no active role in the study design, collection, analysis, interpretation of data, writing of the report or decision to submit the article for publication. This research is approved by Western University's and LHSC's Research Ethics Board under the file #REB 105583.

[2] This work was supported by: The Ontario Trillium Scholarship program (OTS); The Ivey International Centre for Health Innovation (IICHI)/3M Canada MITACS grant; Natural Sciences and Engineering Research Council of Canada (NSERC), and the Universidade Federal do Parana (UFPR). These funding sources had no active role in the study design, collection, analysis, interpretation of data, writing of the report or decision to submit the article for publication. This research is approved by Western University's and LHSC's Research Ethics Board under the file #REB 105583.

List of Tables

List of Figures

Nomenclature

ADT Admission/Discharge/Transfer temporary entry in pacient management system

AFT Accelerated failure time model

AIC Akaike information criterion

APACHE Acute Physiology and Chronic Health Evaluation

AUC Area under the curve

BIC Bayesian information criterion

CART Classification and regression tree model

CCIS Critical Care Information System

CCTC Critical Care Trauma Centre

CCU Coronary Care Unit

Cox-PH proportional hazards regression model

CR-PH Competing risks proportional hazard model

CSRU Cardiac-Surgical Intensive Care Unit

DES Discrete Event Simulation

ED Emergency Department

ETR Event Time Ratio

GLM Generalized Linear Model

HR Hazard Ratio

ICU Intensive Care Unit

ISPOR-SMDM International Society for Pharmacoeconomics and Outcomes Research - Society

for Medical Decision Making modeling good research practices task force

L2 Level 2 unit

Level 2 Intermediary level of care, usually used as a step-dwon from an Intensive Care Unit

LHSC London Health Science Centre

LOS Length of Stay

MAE Mean absolute error

MODS Multiple Organ Dysfunction Syndrome

MOTP Multi-Organ Transplant Unit

MSICU Medical Surgical Intensive Care Unit

NEMS Nine Equivalents of Nursing Manpower Use Score

NN Neural network

NOBS Neurological Observation Unit

NPV Negative prediction value

NRMSE Normalized Root mean squared error

OR Operating Room

PPV Positive prediction value

RF Random forest classification and regression algorithm

RLM Linear model by robust regression

SAPS Simplified Acute physiology Score

SL Super Learner

SOFA Sequential Organ Function Assessment Score

SVM support vector machine model

TEM treatment effects model

UH University Hospital

Chapter 1

Introduction

ICU beds are among the most expensive resources in a hospital, costing more than $3,500/day ([1, 81]). As a result, hospitals strive to find a balance in capacity that allows them to fulfill demand, while keeping costs under control ([3, 4, 5]). Congestion caused by insufficient capacity creates overstay and triggers "off-service", which refers to transfer to a ward other than the one intended for the patient. Overstays are costly ([109]) and can be detrimental to patient health ([7]).Off-service, meanwhile, creates clinical mismatch with staff, and causes delays and coordination issues ([8]). It is, therefore, natural for hospital administrators to look for cost savings in the ICU through better resource management [71, 10, 13, 52].

Canadian hospitals often experience higher utilization rates than their North American counterparts ([79]). In the ICU, high utilization entails high costs, as there is usually a one-to-one patient/nurse ratio compared to a five-to-one ratio on the wards. Given that the ICU entails the highest level of care in any given hospital, nursing workloads serve as a good proxy for a patient readiness to step down from the ICU.

In Ontario, two measures are collected daily from ICU patients: the Nine Equivalents of Nursing Manpower Use (NEMS) score ([33]) and the Multiple Organ Dysfunction Syndrome, (MODS) score ([59]). NEMS is composed of nine items:

1. Basic monitoring: hourly vital signs, regular record and calculation of fluid balance
2. Intravenous medication
3. Mechanical ventilatory support
4. Supplementary ventilatory care: breathing spontaneously through endotracheal tube
5. Single vasoactive medication
6. Multiple vasoactive medications
7. Dialysis
8. Specific interventions in the ICU: such as, endotracheal intubation, introduction of a pacemaker, cardioversion, endoscopy, emergency operation in the past 24 hours, gastric lavage; routine interventions, such as, X-rays, echocardiography, electrocardiography, dressings, introduction of venous or arterial lines, are not included

9. Specific interventions outside the ICU: such as, surgical intervention or diagnostic procedures; the intervention/procedure is related to the severity of the illness of the patient and makes an extra demand upon manpower in the ICU

Given that NEMS can be used as a proxy for patient readiness to step down to a lower level of care, this makes it an ideal parameter for LOS simulations and capacity planning. MODS, on the other hand, is an organ severity score. It tracks the degree of dysfunction of six vital organs, from "none" to "severe", and is composed of the following:

1. Lungs: a respiratory ratio measured as PaO2/FiO2
2. Kidneys: serum creatinine level
3. Liver: serum bilirubin level
4. Heart and vascular system: pressure adjusted heart ratio
5. Blood: platelet count
6. Brain: Glasgow coma score

We hypothesize that MODS and NEMS may be used to estimate a patient's LOS in the ICU. Ny using these measures, we may not only estimate ICU capacity requirements but also predict patient LOS, which, in turn, can help schedule staff, elective surgeries and downstream bed allocation.

At the London Health Sciences Centre, patients discharged from the ICU are sent directly to the wards, which represents a flow of patients from the highest level of care to the lowest. This flow creates long stays in the ICU as patients remain in an ICU bed until they are healthy enough to go to the ward. Hospital administrators have thus proposed creating an intermediary "step-down" unit in which patients who no longer require ICU care but are not yet ready for transfer to the wards may be cared for. The aim is to shorten ICU LOS in order to allow for the faster release of beds for patients in the greatest need. Estimating step down was a critical issue in properly planning the capacity of the new step-down unit as well as in assessing the appropriate capacity of the ICU. As such, I approach the estimation of ICU LOS (and consequently, step down) from several different perspectives, which are explored in the following chapters.

In Chapter 2, I address the ICU long-term capacity and resource allocation problem by developing a a discrete-event simulation model that estimates Level 2 bed needs for LHSC's University Hospital. There are many studies that approach hospital long-term capacity issues,

using either queuing networks (e.g. [26, 30, 18, 70]) or discrete-event simulation (e.g. [20, 88, 97]). My model extends the literature as it includes the entirety of the hospital's inpatient flow to model overstay and off-service. I also include the ICU's daily stochastic flow based on NEMS to simulate patients health progression over time, triggering either the patient's stay for another 24 hours, step down or death. I found that including daily NEMS measuring creates a realistic simulation of patient LOS, which, in turn, allows for more precise long-term capacity planning. My results show significant gains in terms of patient flow increase, reduction of overstay and off-service and overall cost reduction.

In Chapter 3, I approach the estimation of step-down times via LOS and patient outcome (discharge, death, long stay) prediction models. The current literature relies heavily on multiple regression models using acute physiology scores such as APACHE for covariates (e.g. [61, 78, 112]). To predict patient outcomes, I develop several prediction models based instead on MODS and NEMS measures, which are collected by Canadian hospitals upon the patient's arrival in the ICU. My goal is two-fold: to accurately predict LOS and patient outcomes via MODS and NEMS; and to compare and define the best performing ICU LOS prediction models. The models were built using conventional statistical methods, such as linear regression and parametric and non-parametric survival models. I also explore the use of supervised machine learning models, such as CART, Random Forests, Support vector machines and Neural networks. My results suggest that mortality and long stay may be more accurately predicted using supervised machine learning rather than logistic regression. With regard to LOS predictions, I find no single to be dominant model in all performance metrics.

Chapter 4 also deals with ICU LOS predictions, but focuses on short-term capacity planning. The literature recognizes that predictions made upon arrival in the ICU may be improved by being updated later during the patient's stay and by updating various severity scores (e.g. [31, 9]). I use an Accelerated Failure Time (AFT) model to handle the updating model of the LHSC University Hospital ICU at days 1 and 3. Using MODS and day 3 NEMS measures, I show that day 3 LOS predictions provide better model fit in terms of residuals, but have smaller benefits in terms of performance metrics such as R^2, MAE and RMSE. In this chapter I also develop tools for aggregate prediction of ICU demand. Using the same AFT model fitted with MODS and NEMS day 1 data, I produce individual patients' survival functions and estimate their survival probabilities over time. Then I pool the survival probabilities into daily patient cohorts to predict short-term ICU bed demand. Performance metrics demonstrate the superior

accuracy of the AFT model in predicting the LHSC University Hospital's aggregate ICU demand, with the additional benefit of providing meaningful hazard rates and event time ratios.

In the final chapter, I summarize the main results and managerial insights of each ICU capacity planning model. I also propose new avenues for research and highlight policy implications of the adoption of my models by ICUs across Canada.

Chapter 2

Discrete event simulation model for planning Level 2 “step down” bed needs using *NEMS*

Abstract

In highly congested hospitals it may be common for patients to overstay at Intensive Care Units (ICU↓) due to blockages and imbalances in capacity. This is inadequate clinically, as patients occupy a service they no longer need; operationally, as it disrupts flow from upstream units; and financially as ICU beds are more expensive than ward beds. Step-down beds, also known as Level 2↓ beds, have become an increasingly popular and less expensive alternative to ICU beds to deal with this issue. We developed a discrete event simulation model that estimates Level 2 bed needs for a large university hospital. The model innovates by simulating the entirety of the hospital’s inpatient flow and most importantly, the ICU’s daily stochastic flows based on a nursing workload scoring metrics called "Nine Equivalents of Nursing Manpower Use Score↓" (NEMS). Using data from a large academic hospital, the model shows the benefits of Level 2 beds in improving both patient flow and costs.[1] [2]

[1] This work was supported by: The Ontario Trillium Scholarship program (OTS); The Ivey International Centre for Health Innovation (IICHI); Natural Sciences and Engineering Research Council of Canada (NSERC), and the Universidade Federal do Parana (UFPR). These funding sources had no active role in the study design, collection, analysis, interpretation of data, writing of the report or decision to submit the article for publication. This research is approved by Western University’s and LHSC’s Research Ethics Board under the file # REB 105583.

[2] We would like to acknowledge the contribution of the following: J. Kojlak, Dr. C. Martin and F. Priestap from London Health Sciences Center, London, Ontario, N6A 5W9, Canada

2.1 Introduction

Contemporary hospitals in developed countries strive to provide the best possible patient care while keeping costs at reasonable levels ([28, 29, 30]). Hospital beds are too costly to remain idle, while insufficient beds can be detrimental to in patient care ([31]). Critical care in particular is very expensive: in the USA and Canada, ward beds cost as much as $1,000/day while critical care beds surpass $3,500/day ([32, 81]).

The University Hospital (UH↓) campus of the London Health Sciences Centre (LHSC↓) is a 400 bed hospital responsible for approximately 6,200 surgeries, 60,000 emergency visits, 300,000 ambulatory visits and 17,000 inpatient admissions per year ([34]). It routinely experiences bed utilization rates above 85% which are high compared to the North American average of 67.6% for comparable sized hospitals ([79]). When the wards at UH become congested there is pressure on the Medical-Surgical Intensive care unit (MSICU↓) to take one of two actions: hold some patients in ICU longer than they care (“overstay”), or transfer some patients to a ward other than their intended one ("off-service"). Overstay creates a ripple effect in upstream units such as the Operating Room (OR↓) and the Emergency Department (ED↓), resulting in a disruption in patient flow upstream, delayed surgeries and lengthy ED visits. Off-service is sub-optimal clinically because of staff specialization, such as intensivist nurses and physicians. Off-service is also sub-optimal operationally because specialist doctors must visit different wards to see their patients, creating delays and coordination issues. Thus, off-service treatment should be avoided whenever possible ([36]). LHSC estimates that up to 30% of patients in the specialized Multi-Organ Transplant unit are off-service patients.

To improve patient flow, provide adequate care and reduce costs, UH intends to implement an intermediary care unit between the MSICU and its downstream wards, called "step-down" or, "Level 2" unit (L2↓). These wards usually do not support ventilation, but they can still provide some organ support (see Table 2.1↓). They are less costly in technology and in the patient/nurse ratio, typically two patients per nurse rather than one-on-one found in ICU. Among UH’s primary concerns is the determination of the ideal capacity a new L2 unit should have if such unit were to be employed.

This research assesses the impact of step-down beds on a number of hospital metrics including throughput, length of stay (LOS↓), " off-service" and cost. We develop a DES↓ model to analyze a hospital's L2 bed needs that incorporates the changes in ICU patient health through time, where patient health is modeled by the NEMS. We address the following research questions:

1. What is the impact of a L2 unit on throughput, off-service, inpatient LOS and cost?
2. What is the optimal allocation of MSICU and Level 2 beds for UH?

2.2 Literature Review

2.2.1 Research streams

There ares two main streams of literature related to bed capacity management and planning: queuing models and discrete-event simulation (DES) models ([24]). Queuing models range from analytical queuing methodology such as the use of the $M / M / 1$ ([49]) and Erlang loss models ([48, 86]) to the use of complex network models ([30, 26, 43, 70]). [49] presents a survey of this stream of literature, and taxonomies have been devised by [46, 47, 24].

2.2.2 Discrete Event Simulation in Health Care Capacity Management

DES is a popular alternative to queuing models because it is possible to study applications with large scale and scope and to relax many of the assumptions necessary in queuing models. The DES literature most often focuses on a single unit of a hospital (e.g. ED, OR) and/or on a single type of patients (e.g. trauma, surgery, cardiac). Research is usually focused on designing a new patient flow strategy (early transfers, faster service, better schedules) often in combination with structural improvements, such as pooling, or increased capacity. For example, [49] tested pooling respiratory patients into a single unit similar to a L2 unit. [50] found pooling to show significant improvements in patient throughput and flow balance. [88, 86] share those findings, but stress that pooling patients seems to be particularly beneficial in high variance service time settings such as ICU's. [97] simulate a high dependency unit (HDU) and they found that pooling alone only managed to reduce transfers/off-service but kept similar throughput and utilization levels. They could only achieve better results when pooling was combined with earlier stepping-down

of long stay patients. [105] found that capacity increase alone is not enough to stabilize OR patient flows, often requiring faster service times as well. Comparable results are found by [55, 56, 57] in emergency department settings. [58, 59, 60] investigated congestion by smoothing surgery schedules, which enabled performance gains in ICU utilization, LOS and off-service. [64, 62, 8, 7, 65] suggest that highly congested health care systems may trigger other responses—such as early discharges/transfers/off-service—in order to accommodate higher demands, often with negative results.

2.2.3 Contributions of this paper

Our model attempts to correctly represent the complex flow and interactions present in modern general hospitals without some of the simplifications found in the literature. Our DES model includes "bounce-backs" (patients being transferred back from wards to units upstream), overstay and off-service endogenously. In other words, those phenomena are consequences of congestion as opposed to exogenous parameters of the simulation. Thus, we are able to observe congestion and the impact of changes in capacity and bed mix on congestion. We find a clear trade-off between added capacity and changes in bed mix that might otherwise be absent in previous models due to simplifying assumptions. A model that does not include all these characteristics may provide little help in capacity planning problems.

In addition, we include in the ICU simulation the patient's daily health changes in the form of a death/NEMS scoring routine. This stochastic process provides a precise, realistic simulation of an ICU patient and endogenously creates reliable LOS for bed capacity purposes.

2.3 Materials and Methods

2.3.1 Initial Steps

The first step of the research was to meet with several managers at LHSC to understand the problem and agree upon stakeholder involvement as suggested by [90]. The research objective was defined during the first three exploratory meetings and validated after an initial research proposal draft was presented. The research proposal was reviewed and approved by ethics boards of LHSC and Western University. Management at LHSC were highly involved with the research,

periodically revising goals and methods and validating each step to ensure meaningful and actionable results.

2.3.2 Model Overview

We built the DES model using the software package Simul8®. This software was chosen for three main reasons. First, it has become a popular choice in the healthcare DES literature ([5, 77, 92]). Secondly, its ease of coding allows for flexible modeling, and it features a graphical interface that plays an important role in conveying results to multiple stakeholders. Thirdly, and because of the former two, our institution has experience in using this software for healthcare DES research.

We built the model representing the current capacity allocation of UH as a baseline scenario (Figure 2.1↓; for a detailed model, see 6.1↓). There are six entry points for inpatients: Emergency Department (ED), Operating Room (OR), Clinics, Victoria Hospital (the other major hospital in the LHSC system), OneConsult (inpatient transfers from other hospitals outside of the LHSC system), ADT (Admission/Discharge/Transfer↓). ADT is is a mock entry point the hospital uses to temporarily admit patients while they are not assigned a bed in a ward. Each entry point has its own inter-arrival time distributions (see 6↓). Inpatients flow from the entry points to the remaining units. There are two independent Level 3 units (MSICU and Cardiac-Surgical Intensive Care Unit (CSRU↓), three existing Level 2 units (tailored to other specific patient groups) and twelve specialized wards (Table 6.4↓). Patients exit the hospital via three routes: Discharge, “Signed Out”, or Death.

Since the level of care is closely related to patient/nurse ratio, LHSC has historically used nursing workload as a proxy for patient readiness to step down to a lower level of care. As part of the MSICU’s routine, every patient is scored daily in a 56 point scale known as "Nine equivalents of nursing manpower use score" or "NEMS" ([70]). The NEMS gives a measurement of the workload a nurse has for each patient over time and is closely related to patient health because as the patient’s health improves, less nursing attention is needed, resulting in a lower NEMS. Empirically, LHSC considers a score below 10 to be a "Ward type" patient; scores between 11-25 would be "L2 type" patient, and from 26-56 an "ICU type" patient (see Table 2.1↓).

2.3.3 Patient Flow Data

The model was fit using the most recent one year of data in which UH's bed allocation was stable (i.e., same number of beds in all units over the entire year), from December 1st 2013 to November 30th 2014. Data was gathered from the hospital's patient management system, including:

1. Inpatient arrivals: patient registry number, age, sex, diagnosis, entry point, exit point, service at arrival, service at discharge, discharge category (discharge, death, transfer), dates and time of arrival and of discharge.
2. Inpatient Transfers: all of the above plus the date and time of entry and of exit of patients into each unit of UH, origin and destination unit.
3. Hospital bed capacity: number of available beds in each unit during the research period
4. Nursing workloads: patient registry number, age, sex, diagnosis, discharge category (discharge, death, transfer), time and daily NEMS measurements at MSICU
5. Costs: Estimated daily bed costs at each unit

We estimated length-of-stay (LOS) distributions for each unit, patient outcome distributions and patient transfer matrix to represent transitions between hospital units. Note that LOS is ward-specific but does not depend on patient type. For all cases, several distributions were considered ([20]) and chosen on basis of Akaike information criterion(AIC↓, [3]) and Bayesian information criterion (BIC↓, [95, 53]), as is common in this line of research (e.g. [75, 86]).

2.3.4 Transition Probabilities

There were 17,380 patients representing 42,012 internal movements (an average of 2.41 records/patient) represented in the patient flow matrix (Figure 6.2↓). Each transfer has an unique destination. However, if the intended unit is full, then the practice is to transfer the patient to an alternate unit, causing off-service care. In this way, individual off-service decisions are determined probabilistically. Deaths from the MSICU were modeled separately using a logarithmic function (Figure 6.4↓).

During the patient's stay at MSICU, patients receive a NEMS upon arrival to MSICU, and a revised score every morning during their stay in MSICU. Once the patient reaches a NEMS consistent with a L2 type, she attempts to exit the MSICU and reach the new *L2* unit. In the baseline scenario, patients exit MSICU if they reach a ward type NEMS.

2.3.5 Cost Data

LHSC supplied cost per patient-day for each level of care (Table 2.1↓) as well as capital expenditure estimates for 8 and 15 L2 beds (originated for a previous investment in another site) . We calculated annualized capital expenditures for the entire range from two to 28 L2 beds by linear extrapolation and 10 year linear depreciation, consistent with Canadian accounting practice (Table 6.6↓).

2.3.6 Simulation scenarios and runs

We evaluated the following scenarios:

1. Capacity increase with a L2 unit: Adding a range from 2 to 20 *L2* beds into the existing baseline model.
2. Capacity re-allocation: Maintain a total of 25 beds while shifting capacity from MSICU into the new L2 unit.
3. Capacity re-allocation: Increase the total to 30 beds while shifting capacity from MSICU into the new L2 unit.

Each configuration of each scenario was simulated 200 times, using a one year warm-up period followed by a one year data collection period. A different random seed number was used for each run. Trial run times varied from 20 to 40 minutes using an Intel® Core i5-2400 CPU 3.10GHz 8GB RAM server.

2.4 Results

2.4.1 Model Validation

Our simulation model captures the individual physicians' and nurses' decisions to transfer or discharge individual patients via a macro approach, using LOS distributions for each ward and a probabilistic transition matrix for each patient movement. To validate this approach, we compared patient arrival, throughput, LOS and cost results from the baseline simulation with aggregate empirical data and cost data from publicly available documents such as LHSC's financial statements [67] and the Canadian Institute for Health Information yearly reports [29].

The model is accurate in reproducing entry data, MSICU LOS and cost data (Table 2.2↓). Average throughput is within 1% of empirical data, while total LOS is within 0.4%. MSICU LOS is slightly high (2.9%) but with a lower standard deviation, resulting in no statistically significant difference compared to the empirical data. We concluded that the simulation model is sufficiently valid to address the research questions. Results for all scenarios are summarized in Table 2.4↓.

2.4.2 Scenario 1: Capacity increase with a New *L2* unit

We evaluated the addition of extra beds in a general-purpose "net new capacity" step-down ward. We simulated a range of 2 to 20 L2 beds in a dedicated unit immediately downstream from the MSICU and did not alter the capacity of the MSICU (25 beds). We first assessed the impact of the new capacity on off-service utilization. In the base case (i.e. no new capacity), the existing specialized Level 2 units (MOTP↓, CCU↓, NOBS↓) have a combined off-service load of 573 patients/year. This value drops to 225 patients/year as we add L2 beds. In the base case, the Level 3 units (MSICU and CSRU) have a combined off-service of 621 patients/year. As L2 beds are added, the off-service reduces to approximately 110 patients/year, representing a reduction of 82%. This reduction may represent a significant improvement in terms of patient care, as approximately 500 more Level 3 patients are now able to be transferred to their intended wards.

Next we evaluated the impact of the new L2 beds on throughput. The addition of an L2 unit increases MSICU throughput up until 8-10 new beds where it stabilizes at approximately 1,068 patients/year (Figure 2.2↓). The L2 unit's throughput grows until 12-14 beds are added, reaching 730-732 patients/year. This suggests that until the L2 unit capacity reaches 12 beds, MSICU is still hosting "step-down ready" patients but after that point there is little clinical need for extra beds.

Utilization and LOS have a similar pattern (Figure 2.3↓). The MSICU has a high initial utilization rate (above 85%) that drops dramatically as L2 capacity is increased, eventually stabilizing around 29% at 12 beds. As L2 beds are added, there is a rapid decline in MSICU LOS until we reach 12 beds, where it stabilizes at approximately 59 hours (Figure 2.4↓). Moreover, the percentage of patients who stay more than 21 days in the MSICU reduces to approximately zero after 8 beds. This suggests that additional L2 capacity allows the MSICU to return to its clinical role of intensive care.

Finally, we find that a maximum of 29 total beds (MSICU and L2 beds combined) are ever occupied, which exceeds MSICU's current capacity of 25 beds. This supports further investigation of increased capacity in MSICU in Scenario 3 (Section 2.4.4↓).

2.4.3 Scenario 2: Capacity re-allocation

This scenario involves creating a new L2 unit, but rather than creating new capacity, beds in the existing MSICU would be closed and reallocated to the L2 unit. This scenario would apply in case the hospital does not have additional space to create a new L2 unit or budget for net new beds. Off-service loads are slightly higher than in Scenario 1. The minimum off-service load is reached when there are 15 MSICU and 10 L2 beds, leading to total L3 off-service load of 150 instances per year. This figure represents an improvement in terms of patient care, as approximately 470 patients can now be transferred to their intended wards. Off-service performance then deteriorates as more beds are shifted from MSICU to the L2 unit. MSICU becomes a bottleneck and upstream units are forced to send off-service patients to CSRU. This situation represents a clear clinical misfit, as CSRU is a cardiac surgery unit, where both nurses and physicians are heavily specialized in cardiac care. The treatment of patients intended for MSICU in CSRU could result in deterioration of patient care and disruption of the cardiac surgery patient flow.

MSICU throughput improvements start when there are 4 beds reaching an optimal value of 1,050 patients/year when there are 15 MSICU and 10 L2 beds (Figure 2.2↓). The L2 unit reaches a peak throughput of 720 patient/year when there are 13 MSICU and 12 L2 beds. This is similar to the maximum throughput achieved when we evaluated net new capacity in Scenario 1. After that point, as MSICU beds are converted into L2 beds, the smaller number of MSICU beds becomes a bottleneck to upstream units such as the ED and OR. Patient flow reduces significantly and blockage becomes more frequent in those units due to high utilization rates at MSICU. As the L2 unit is a dedicated downstream unit of MSICU, its throughput is also reduced after 12 L2 beds.

MSICU LOS begins to improve after creating 4 L2 beds. The minimum LOS of 60.66 h/patient occurs when there are 13 MSICU and 12 L2 beds, representing a 63% improvement relative to the base case. As more capacity is shifted to L2 beds, the LOS rises back to the 70 h/patient mark. This reduction represents a gain of at least 2,000 patient-days/year in the

combined MSICU and L2 capacity. This confirms our earlier finding in Scenario 1: a L2 unit provides opportunity for MSICU to go back to its clinical role, with minimum overstay.

This result makes sense due to the drastic reduction in long-stay patients in the MSICU (MSICU LOS above 21 days—Figure 2.5↓). Those patients often reach a L2 NEMS, triggering their stepping-down into the New L2 unit. The result is higher availability of MSICU beds (Figure 2.3↓ (b)) for patients originating from upstream units, thus improving patient flow.

2.4.4 Scenario 3: New capacity and capacity reallocation

In this scenario we evaluated reallocation of beds along with net new capacity of 5 beds. Off-service loads are between the two previous scenarios, with lowest values within a range of 20 to 16 MSICU beds. MSICU throughput is stable at 1,050 patients/year anywhere from 20 to 16 beds reaching a peak of 1.063 patients/year (Figure 2.2↓), while L2 throughput is stable within the range of 10 to 18 beds, peaking at 720 patients/year. Therefore any mix from 20 MSICU and 10 L2 beds to 12 MSICU and 18 L2 beds have comparable results with the Scenario 2 while providing a stable combined throughput. MSICU utilization rates are also significantly lower than in the in Scenario 2, as seen in Figure 2.3↓. With MSICU reaching a minimum slightly below 40% (20 MSICU and 10 L2) and reaching a balanced utilization of approximately 45-47% at 16 MSICU and 14 L2 beds.

Any mix from 20 MSICU and 10 L2 beds to 12 MSICU and 18 L2 beds yield approximately 60h LOS, similar of the previous scenarios (Figure 2.4↓). As in previous analysis, the ability to step down long stay patients with low NEMS plays an important role in improving patient flow (Figure 2.5↓).

2.4.5 Costs

In all three scenarios a significant cost saving was possible relative to the current cost of $3,500/patient-day in MSICU (Figure 2.6↓). Combined MSICU and L2 costs decrease steadily in all scenarios until they reach a minimum of $2,869.46/patient-day at 12 L2 beds under scenario 3. From that point on, under all scenarios, costs escalate, but never reach the current baseline cost. This result can be explained by two factors. First, L2 operational costs represent only 57% of MSICU's. Initial increases in L2 capacity permit a timely step-down and immediate savings occur. Second, after 12 L2 beds, the new L2 unit starts to have idle capacity. This is due to lack

of demand in Scenario 1 and to MSICU constrained flow in Scenarios 2 and 3. Idle L2 beds carry high fixed costs in the form capital expenditure, thus forming the upward half of the curve.

2.4.6 Increased arrivals

By increasing throughput capacity, the hospital may receive more patients. Thus, we simulated an increase in the inpatient flow from ED and OR to see how well our optimal configurations stand a hypothetical surge in demand. For Scenario 1, we focused on ED and OR, where inpatients spend relatively little time waiting for their disposition from ED, or their scheduled surgeries in OR. This is not the wait time to enter the ED, as we simulated only *inpatient* flow. This wait is for patient disposition, i.e. the moment the patient is ready to receive a decision to admit until the true admission and transfer to the intended location. A 10% increase in ED and OR demand, representing an extra 1,200 patients/year, is enough to negate any gains achieved by the introduction of net new L2 capacity (Table 2.3↓).

Next, we focused on MSICU performance in Scenario 3. The inpatient surge is mostly absorbed by MSICU and *L2*, reaching maximums of 1,300 and 930 patients/year respectively (Figure 2.7↓ (a)). There is a gradual shift in the optimum bed mix to 16 MSICU and 14 *L2* beds. Utilization rates increase accordingly, reaching approximately 60% in the optimum throughput bed mix (Figure 2.7↓ (b)). MSICU LOS changes little with the increase in ED and OR demand (Figure 2.8↓(a)). At 30% increase in demand, MSICU LOS rises to approximately 65 hours/patient. In terms of LOS, the optimal configuration shifts slightly to 16 MSICU beds and 14 L2 beds. Thus, the increase in inpatient volume does affect the values of MSICU patient flow indicators but the optimal solution is robust to increased volumes.

Higher utilization in MSICU triggers congestion upstream. Particularly in the ED, at the 30% demand increase, there is an increase of 317% in the use of temporary ED beds (the ED decant ward, with a capacity of 6 beds).

Combined MSICU and L2 patient-day costs remain similar even with a 30% inpatient arrival increase (Figure 2.9↓ (a)), but the minimum shifts slightly from 18 MSICU beds and 12 L2 beds to 16 MSICU beds and 14 L2 beds. Figure 2.9↓ (b) shows that Scenario 3 had a robust range in terms of total cost, with an approximate value of $14.5 million/year for a range of 18 to 12 MSICU beds and 12 to 18 L2 beds. In the 30% demand increase, however, total cost is continuously decreasing, with the optimal mix costing an extra $4.7 million/year, or 33.4% more than Scenario 3. This a direct result of MSICU's diminishing capacity to absorb the increased

demand. However, even a 30% increase in ED and OR volume in the optimal configuration is not enough to return total MSICU and L2 cost to the level of the baseline scenario of $24 million, demonstrating the impact the L2 unit has in UH's cost structure (Figure 2.9↓ (b)).

2.4.7 Management Feedback

Preliminary results from this analysis were presented to a team of managers of LHSC in January 2017. The team consisted of the Vice President of Access and Flow, the Director of Clinical Redesign, the Director of Critical Care, and the City-wide Chair and Chief of Medicine, among others. Our research confirmed their intuition about the need for an L2 unit, but revealed unanticipated findings in terms of the L2 unit's ability to improve flow, reduce MSICU LOS (63% from current levels) and reduce cost by approximately 40%. Implementation of the new L2 unit is likely to occur in the near future.

The managers in attendance stated that our model was the first large scale DES model to be used in UH. Our results led to questions about the need for a clinical study about the MSICU long-stay population and their desired care pathway, as well as about UH's capacity to deal with increased demand. They concluded that our DES model provides support for further L2 capacity studies in other LHSC sites as well, such as Victoria Hospital's L2 clinical redesign.

2.5 Conclusions

We found that there are considerable performance gains to be made with the addition of a step-down unit. In all scenarios, the optimal performance occurs when there are approximately 12 L2 beds yielding MSICU LOS of approximately 60 hours/patient, a cost reduction of 18% per patient-day and 40% in total cost per year (see Table 2.4↓).

It has been recognized for some time in health care simulation literature that implementation does not necessarily follow the recommendations proposed by researchers ([79, 24, 90]). [41] report that from 59 articles surveyed in the literature, only 14 mentioned implementation. Many reasons for this gap are possible, such as lack of client involvement, lack of clear methodology and failure to communicate results properly. To avoid such problems, we followed a general framework of the methodology based on previous literature ([83, 24, 41]) and the ↓ best practices ([86]). In particular, stakeholders were involved right from the beginning of the study, validating and providing input in every step of the research.

Our model has limitations. Our data represents only inpatient arrivals so our model does not consider balking or reneging at any entry points. This means that all ED and OR arrivals are admitted patients and must go through the system. We use a simplified model of the ED and thus our model does not capture ED congestion. However, we believe that this does not have significant impact on our analysis since ED arrivals that eventually visit MSICU are unlikely to be turned down by UH due to their health status. Also, the Death/Stay/Step-down routine has a minor drawback: once the patient is prevented from leaving MSICU due to blockage downstream, the patient has to wait for the next morning to have a new chance to leave the MSICU. In spite of this drawback, the model validation found accurate MSICU LOS.

There are several directions for further research. First, we will explore further the pooling effects that one might have from merging inpatient wards and/or other specialized L2 units. These units are all highly congested and susceptible to blockage, bounce-backs and grid-locks. Also, we modeled all routing and discharge decisions between wards and other hospital units probabilistically. An interesting avenue for future research would be to incorporate decision rules for these occurrences. Second, we can use the data set to create predictive models for LOS based on NEMS. These can then be used to create dynamic staffing models. Finally, we will develop an analytical model that incorporates MSICU's unique position in which it is squeezed between ED/OR's efforts to minimize wait times and the wards efforts to avoid re-admissions. This may involve a combination of queuing and game theory.

Table 2.1 Levels of care characteristics at LHSC

Level of care	Bed characteristics	Patient/nurse ratio	Estimated cost $/patient-day [1]	NEMS [2]
1	Standard Ward bed: No organ support, no ventilation	3 or more to 1	$600	≤ 10
2	Step-down bed: Support single failed organ system, no ventilation	2 to 1	$2,000	11 to 25
3	Intensive care bed: Invasive ventilation and multiple organ support	1 to 1	$3,500	26 to 56

[1] Estimated cost provided by LHSC Management;

[2] Nine equivalents of nursing manpower use score ([87])

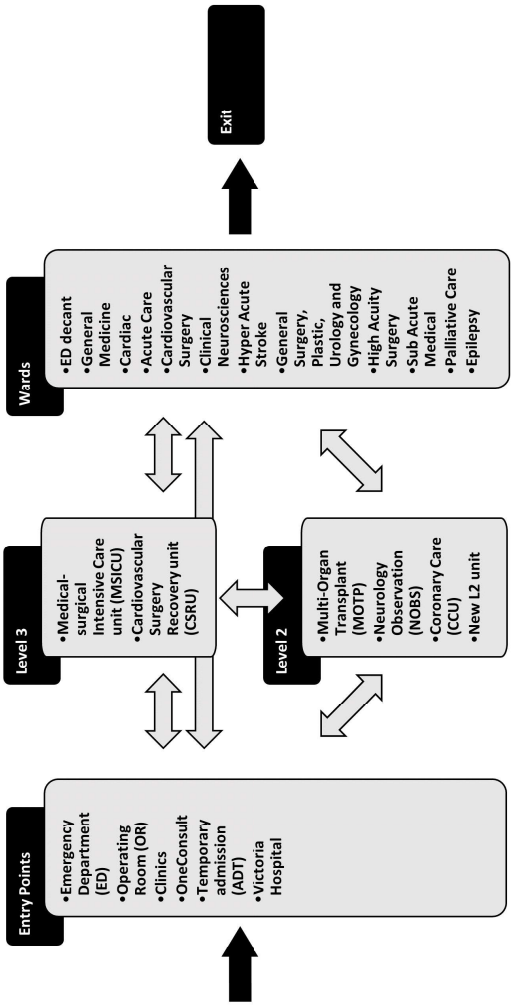


Figure 2.1 Figure 2: *L2*patientflowmodel

Table 2.2 Output and Cost validation

Indicator	Simulation			Empirical data	Difference
	-95% confidence limit	Average	95% confidence limit		
Throughput (patients/year)	17,128.05	17,194.00	17,159.95	17,380.00	-1.07%
Average overall LOS (days/stay)	6.84	6.87	6.90	6.90 ([29])	-0.40%
Cost of hospital stay	$6,347.36	$6,345.41	$6,343.48	$6,123.00 ([29])	3.63%
Total operational cost	$108,717,845	$109,103,000	$109,488,155	$106,417,740 ([67])	2.52%
MSICU Average LOS (hours)	162.12	164.24*	166.36	159.6*	2.91%
MSICU Std Dev of LOS (hours)	174.13	177.96	181.80	201.8	-11.81%
MSICU Long stays (≥ 504 hours)	5.53%	5.26%	4.90%	5%	-0.27%

*P value and statistical significance:The two-tailed P value equals 0.5884

By conventional criteria, this difference is considered to be not statistically significant.

The mean of simulation minus raw input data equals 4.6400

Confidence interval: 95% confidence interval of this difference: From -12.2025 to 21.4825

Intermediate values used in calculations: t = 0.5413 df = 1963 standard error of difference = 8.572

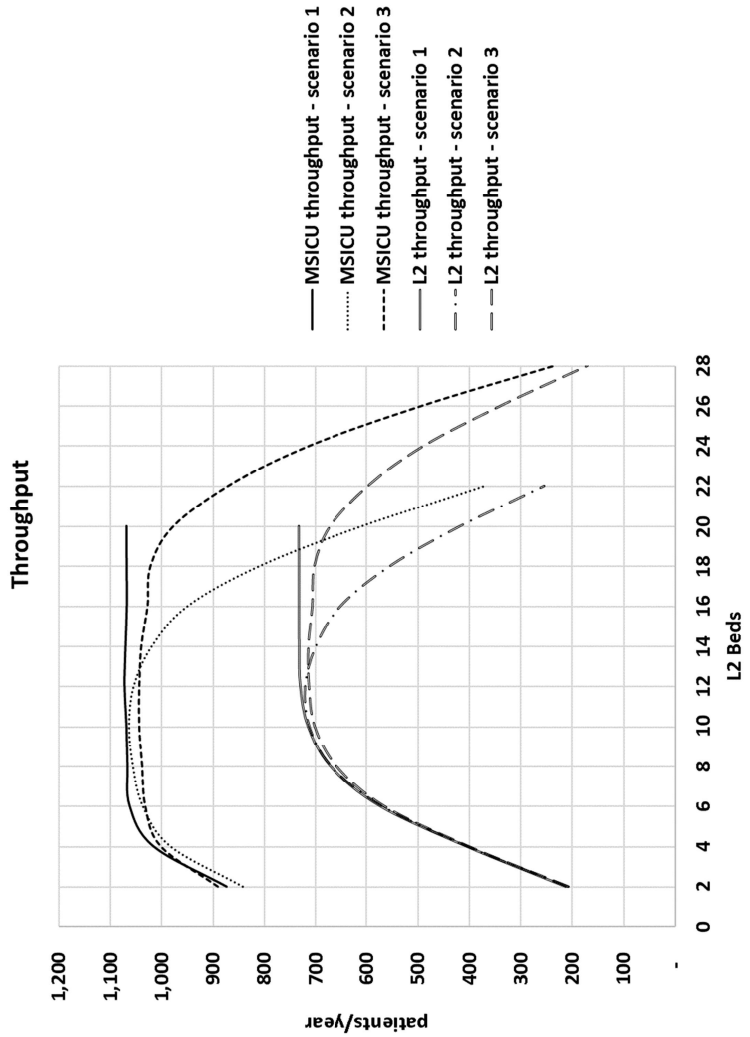


Figure 2.2: MSICU and L2 throughput vs. number of New L2 beds added

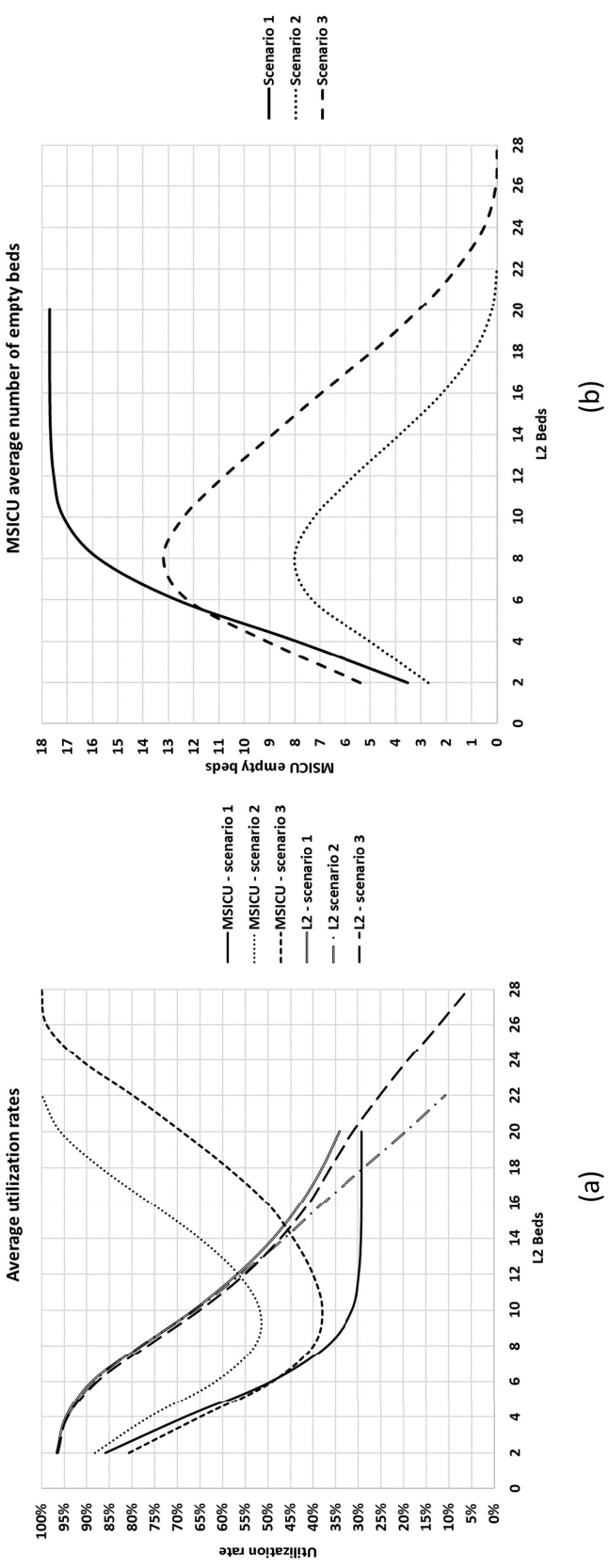


Figure 2.3: MSICU and L2 average utilization rates

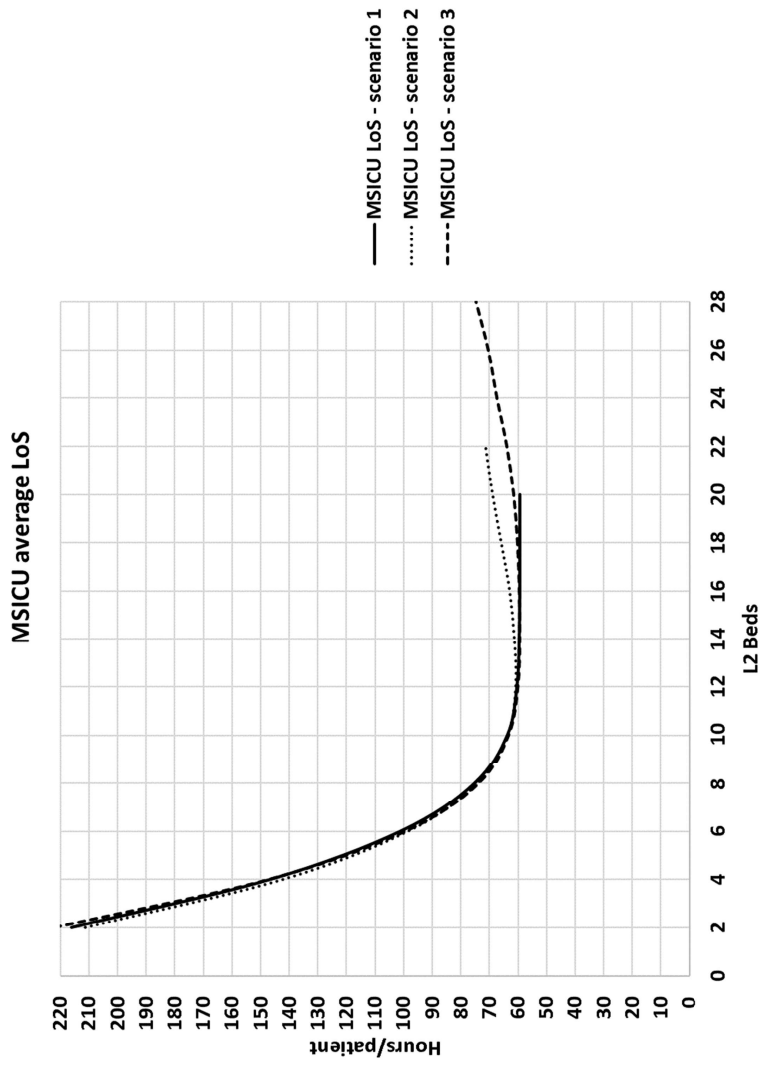


Figure 2.4: MSICU average LOS

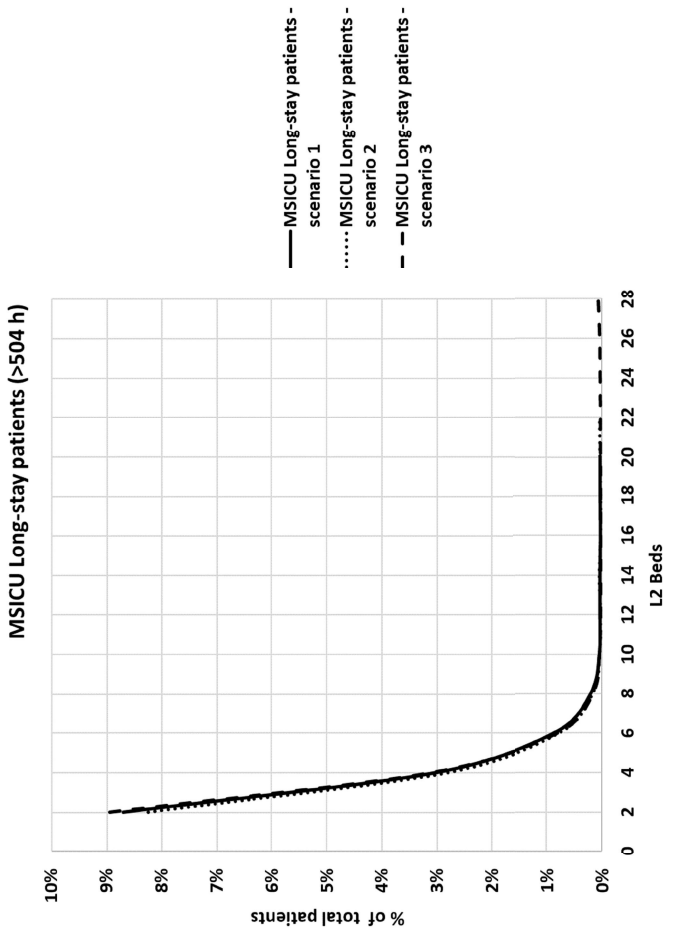


Figure 2.5: MSICU Long-stay patients

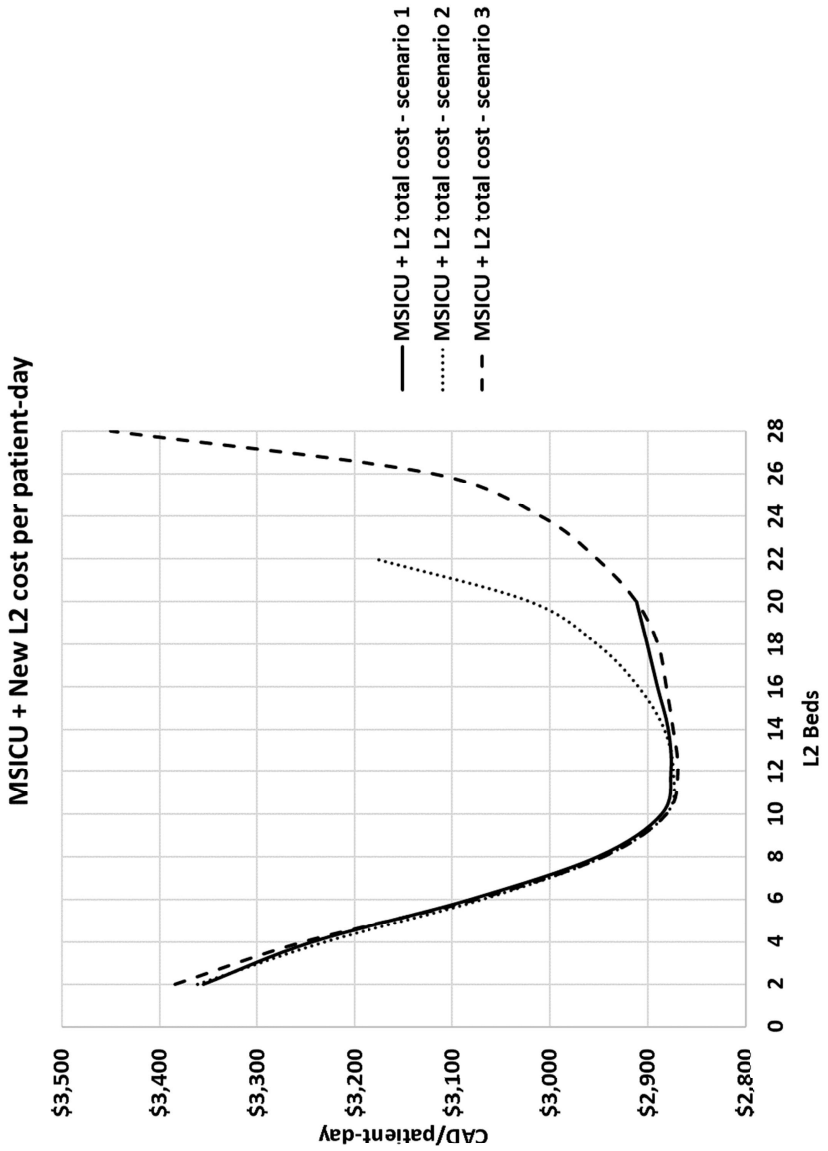


Figure 2.6: Combined MSICU + *L2* cost per patient day

Table 2.3 Sensitivity in *Inpatient* flow

	Wait for Disposition			ED Decant			Queue for WC OR			Total LOS	
Scenario	wait (h)	std (h)	≤ 5 min	LOS (h)	std (LoS)	≤ 1 hour	wait (h)	std (h)	≤ 1 hour	LOS	std (h)
Baseline	0.12	4.99	99%	1.27	7.95	90%	0.43	1.26	87%	164.93	212.83
25 MSICU and12 *L2*	0.22	2.57	98%	2.07	6.33	84%	0.35	1.03	93%	162.69	196.77
5% increase	0.3	2.57	97%	2.08	6.35	82%	0.35	1.03	88%	163.69	194.67
10% increase	1.13	6.34	94%	3.01	7.56	75%	0.75	1.69	79%	164.83	194.2
20% increase	2.09	7.88	87%	3.22	7.36	69%	1.01	1.98	74%	165.2	194.3
30% increase	26.67	50.1	55%	5.26	8.96	52%	1.28	2.24	69%	173.25	189.65

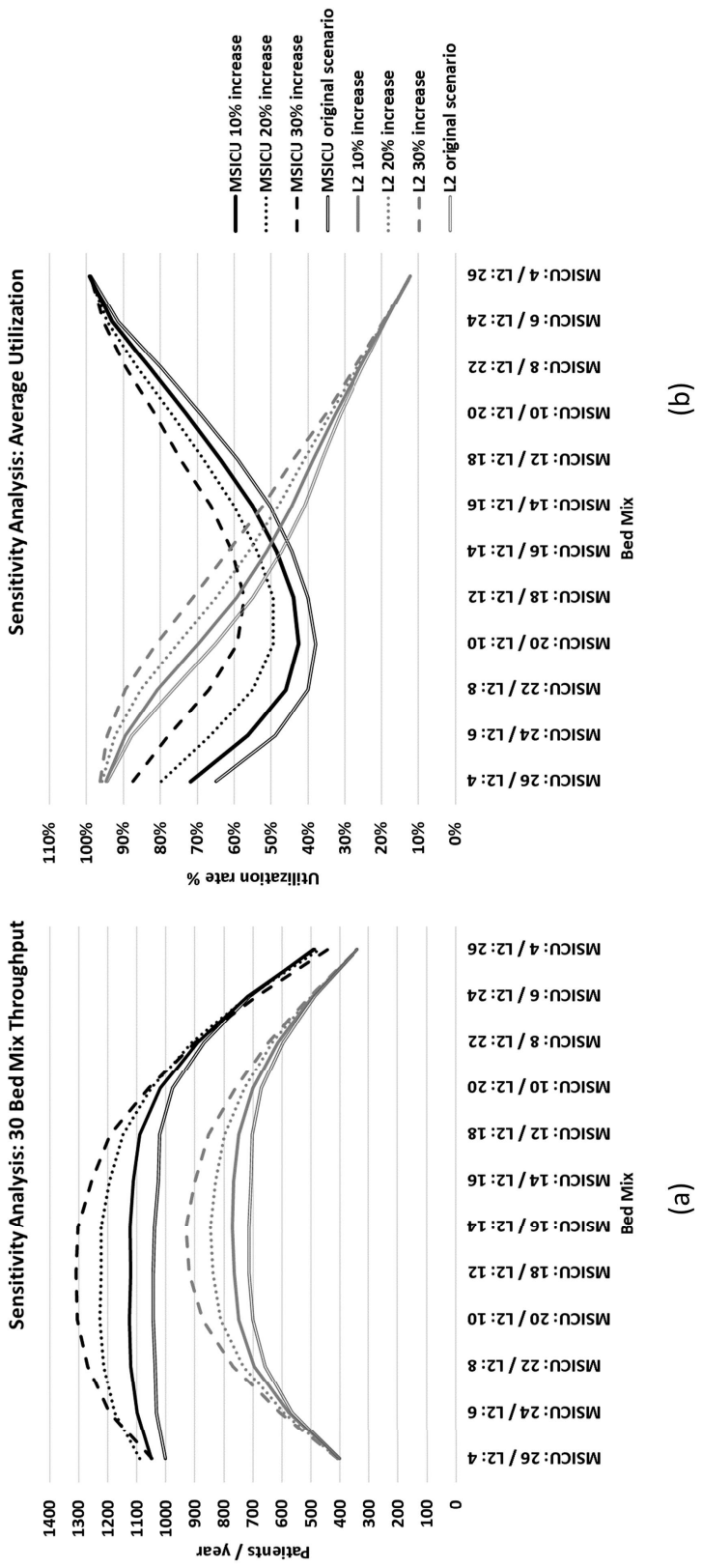


Figure 2.7: Sensitivity Analysis – throughput and utilization

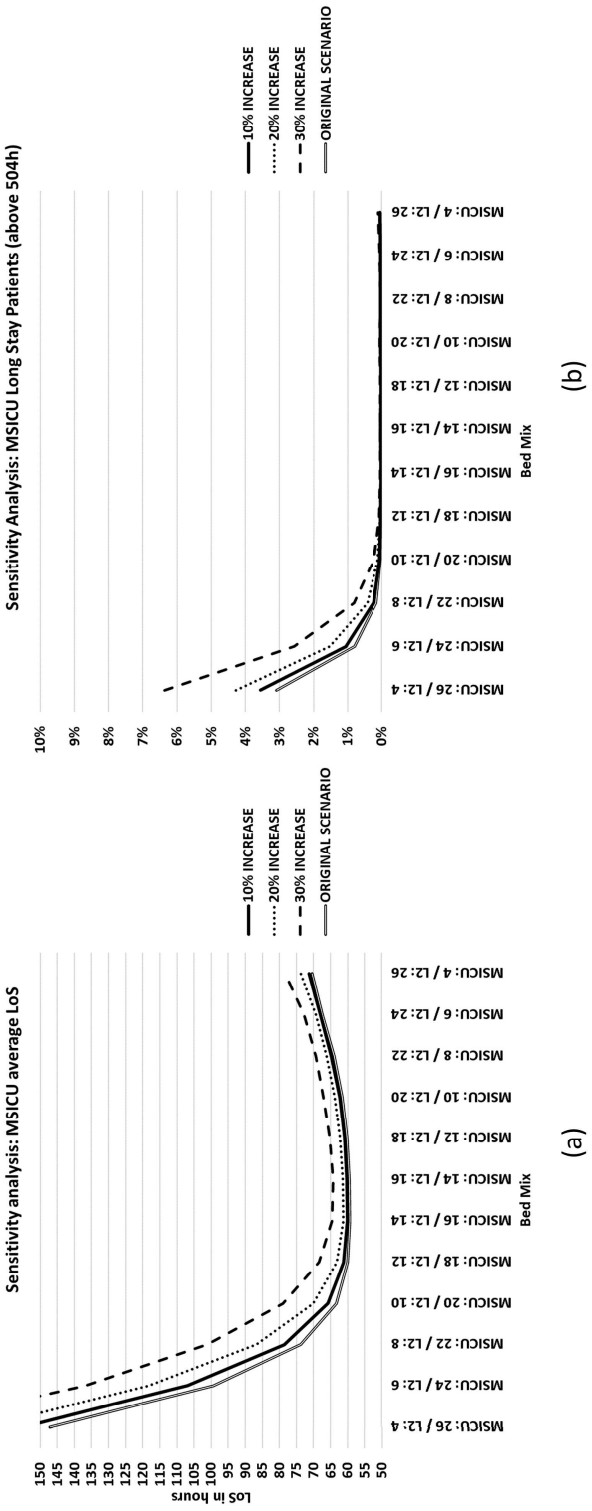


Figure 2.8: Sensitivity Analysis-*LOS* and Long stays

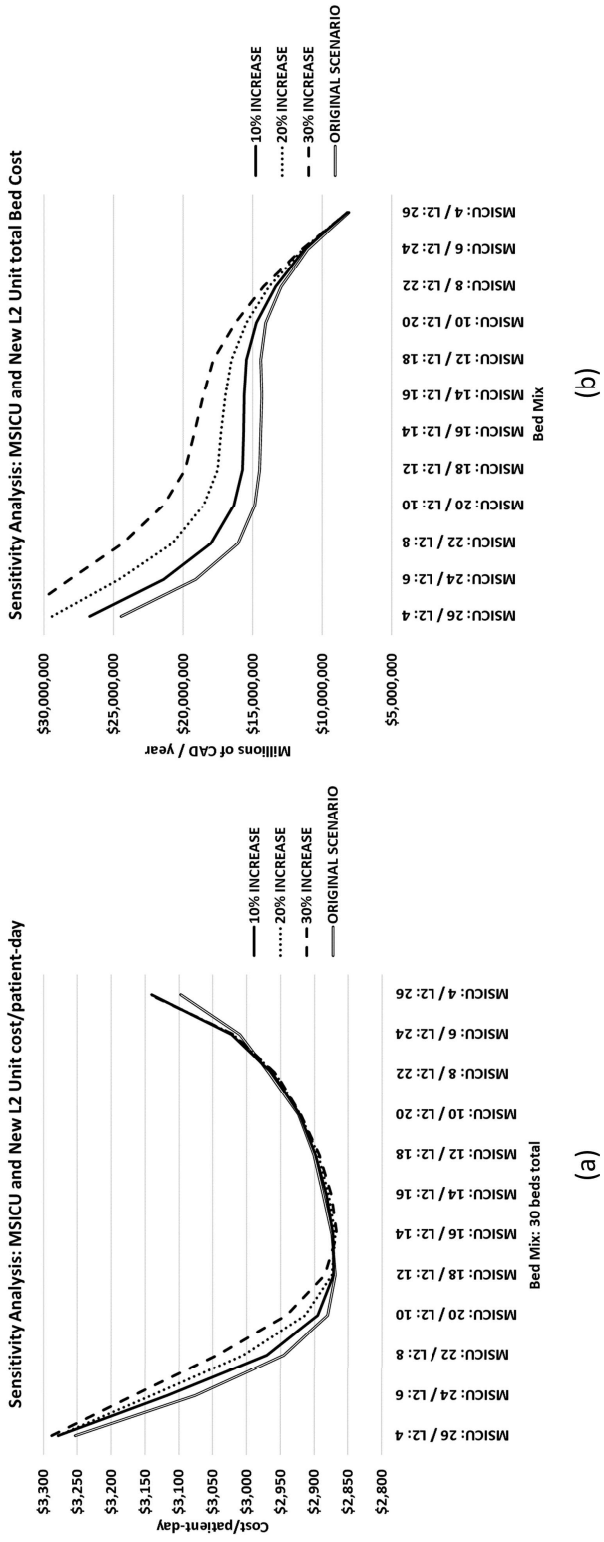


Figure 2.9: Sensitivity Analysis – combined MSICU and *L2* cost per patient day

Table 2.4 Scenario comparison

Indicator	Baseline	Scenario 1	Scenario 2	Scenario 3
MSICU capacity (beds)	25	25	13	18
L2 Capacity (beds)	0	12	12	12
Total Capacity (beds)	25	37	25	30
Mean (beds)	19.1	14.4	14.32	14.29
Median (beds)	19	14	14	14
Mode (beds)	19	14	13	15
Max (beds)	25	29	24	27
Std. dev (beds)	3.28	4.02	3.32	4.33
Average utilization	76.40%	38.92%	57.28%	47.63%
Max utilization	100%	78.38%	96.00%	90.00%
Cumulative frequency below 75%	21	≈17	≈16	≈17
Cumulative frequency below 95%	25	≈25	≈ 20	≈21
LOS in MSICU (h)	164.24	60.37	60.66	60.06
Cost CAD $/patient-day	$3,477.44	$2,876.21	$2,873.83	$2,869.46
Total Cost MSICU+L2 CAD $/year	$24,019,830.00	$14,909,503.75	$14,760,363.22	$14,503,103.34

Chapter 3

ICU Length-of-stay prediction models based on NEMS and MODS

Abstract

Length of stay (LOS↓) is a critical metric for Intensive Care Unit (ICU↓) resource planning. If a hospital can estimate ICU patients' LOS, then it can better schedule both staff and elective surgeries and allocate beds to downstream wards. The "Nine Equivalents of Nursing Manpower Use Score" (NEMS) ↓ and "Multi-Organ Dysfunction Syndrome" (MODS) ↓ score are two commonly used metrics collected daily from ICU patients. Our objective is two-fold. First, we predict patient outcomes (discharge or death) and ICU LOS using prediction models based on NEMS and MODS scores that are assessed at the time of arrival at the ICU. Using data from a large Canadian teaching hospital, we observe that NEMS and MODS behave differently as predictors of LOS, depending on patient outcome (discharge or death). Second, we show how several different techniques, including regressions, survival models and classification models, perform in generating outcome and LOS predictions that can be used in short-term resource planning. While logistic regression, random forest and super learner tended to dominate outcome predictions, no model dominated in terms of LOS predictions.

3.1 Introduction[1]

Intensive care beds are among the most expensive resources in a modern hospital, with costs surpassing $3,500/day and accounting for a significant part of hospital operational costs ([76, 80, 50, 103]). ICU costs are strongly related to LOS ([109]), so it is natural that hospital administrators look for cost savings in the ICU by better managing LOS and discharge processes ([71, 10, 13, 52]). Many models have recently been proposed to help practitioners predict ICU LOS ([66, 109]). These rely predominantly on the Acute Physiology and Chronic Health Evaluation (APACHE↓) score, the Simplified Acute Physiology Score (SAPS↓) and the Sequential Organ Function Assessment (SOFA↓) score ([76, 107, 108, 112]). In Ontario, hospitals are required to collect MODS ([59]) and NEMS ([33]) data for reporting purposes, but do not collect sufficient data to claculate APACHE, SAPS or SOFA. As a result, hospitals in Ontario are compelled to rely on NEMS and MODS models for resource planning purposes despite the fact that these are regression based and do not take advantage of modern machine learning methods.

The two main goals of this study are, therefore:

1. To compare the predictive performance of conventional clinical ICU LOS prediction models versus modern supervised machine learning models.
2. To develop ICU LOS and outcome models relevant to Canada using NEMS and MODS metrics.

3.2 Literature Review

3.2.1 Common methodologies

The literature presents many different methodologies for ICU LOS prediction, such as simulation ([44]), Markov chains ([2, 45]), and multi-stage models ([46]); the most commonly used are OLS multiple regression and GLM models ([108]). Only recently have machine learning models been

[1] This work was supported by: The Ontario Trillium Scholarship program (OTS); an Ivey International Centre for Health Innovation (IICHI)/3M Canada MITACS grant; the Natural Sciences and Engineering Research Council of Canada (NSERC), and the Universidade Federal do Parana (UFPR). These funding sources had no active role in the study design, collection, analysis, or interpretation of data, the writing of the report or the decision to submit the article for publication. This research is approved by Western University's and LHSC's Research Ethics Board under the file # REB 105583.

used in patient ICU LOS prediction ([16, 19, 32, 55, 93]). Different models have proven to be useful in predicting ICU mortality and LOS ([110, 98, 109]). Regression models such as [110] and [61] and, more recently, [57] have shown that most of the variation in LOS may be attributed to patient characteristics, with hospital and intensive care unit characteristics playing a smaller role ([57, 63, 110, 10, 13]).

An advantage of regression models is that they provide straightforward explanatory value for the covariates, which is common in the ICU literature. However, there is no consensus as to interaction terms, non-linear relationships and variable transformations, which makes it difficult to interpret and compare such models ([108]). Machine learning models, on the other hand, have their own strengths and weaknesses. Support vector machines, for instance, because of non-linear kernels, can handle non-linear relationships with ease, but require careful calibration and are prone to overfitting in larger data sets. Neural networks, meanwhile, have only performed well with ICU data when trained in large data sets ([68, 99]). Classification and regression trees tend to be more robust with regard to non-linear relationships and outliers, but are also prone to overfitting ([17]). Ensemble methods, such as random forest and super learner, can alleviate overfitting (often dominating regression models and classification trees), but lack explanatory power ([84, 16]).

3.2.2 Challenge in prediction accuracy involves mortality

A common feature of ICU LOS prediction models is that covariates for acuity and morbidity scores, such as APACHE, behave inversely with regard to patient outcome ([72, 61, 78, 82, 107, 108, 109, 1]). For example, ICU LOS increases amongst surviving patients and decreases amongst deceased patients. Mortality is commonly found to decrease LOS prediction performance ([76, 98, 107, 1, 112]), while survival is associated with a larger mean ICU LOS ([108]). Death in the ICU may be considered an endogenous omitted variable that may cause bias in LOS prediction ([78]) and it has been suggested that prediction models need to account for this phenomenon ([108, 1]). Moran and Solomon (2012) suggest using a treatment effects model to account for the endogeneity of ICU mortality, while survival analysis models can handle mortality as censored data.

Most regression models account for only 5-20% of the individual variation in ICU LOS ([9, 82, 108, 112]), although this is still statistically significant in predicting the likelihood of long stay ([9]) and readmission ([11]). This is one reason why scholars are reticent to use such

models to predict individual patient LOS ([66, 72, 73, 109, 1]) but confident to recommend them for ICU benchmarking ([76, 75]).

3.2.3 Lack of uniformity in study design hinders model comparison

There is little uniformity with regard to study design ([62, 36]), particularly in terms of LOS measurement and exclusion criteria; this makes study comparison somewhat difficult ([72, 62, 51]). Common exclusion criteria are age, minimum ICU LOS thresholds, burn patients, and coronary artery bypass, cardiac valve or heart transplant patients ([71, 27]). Readmission is treated differently depending on the study; it is included in some models and excluded from others ([1, 51]).

Many independent variables have been proposed, with patient characteristics such as age, sex, acuity and morbidity scores being the most common ([78]). Other variables include the type of ventilation, surgery, diagnosis, organ function (e.g., respiratory, renal, hepatic, cardiovascular, hematologic) and neurological dysfunction levels, all of which may be part of the most common acuity scores ([87]). Newer versions of APACHE also include 116 detailed diagnosis groups, as well as comorbidities such as cirrhosis, AIDS, and lymphoma. In some studies, minimum ICU LOS is limited to 1 to 8 hours ([71, 78, 51]) while maximum LOS may be anywhere from 21 to 60 days (the latter equivalent to the 99th percentile) (e.g. [44, 10, 78, 73]). Minimum age thresholds, meanwhile, vary from 14 to 18 years of age ([10, 98, 51]). Some studies also collect race and geographical location ([10]), the hospital's teaching status and size ([73]), LOS prior to ICU time and Glasgow coma scores ([9]).

3.2.4 Mortality prediction

[98] developed a logistic regression model to predict mortality and stated that adding measurements from day 2 to those from day 1 improved mortality prediction. [61, 78] built a logistic regression model for mortality prediction with a data set of over 200,000 patients and found APACHE, diagnosis, age, and surgical status and related interactions to be the most important predictors. [82] also employed logistic regression for their mortality model using 80,000 ICU admissions in 23 Finnish ICUs. [87] is the only study we found that incorporates MODS and NEMS. The researchers used a logistic regression model and found that MODS, NEMS, age, diagnosis, and admission source were significant predictors of mortality for 8,800

Canadian ICU admissions. [83, 84] show the dominance of the machine learning algorithm, super learner, over other machine learning and logistic regression models as applied to a data set of American and French ICU admissions. [83] and [93] found similar results using comparable supervised machine learning algorithms. Superior prediction performance of machine learning methods was also found within patient sub-categories such as cardiac surgery and trauma patients ([4, 69]), an approach defended by [16].

3.2.5 LOS prediction

[112] developed a multiple regression model based on the APACHE classification system for predicting ICU LOS with a cohort of over 130,000 patients from 45 U.S. hospitals. Their model accounts for 21.5% of variation in ICU LOS across individual patients and 62% of variation across ICUs. [107] compared APACHE IV, MPM0 and SAPS II as predictors of ICU LOS in 11,000 Californian ICU admissions, finding APACHE to be the best LOS prediction model for their patient cohort. [61] and [78] compared several regression models in a patient cohort of 230,00 and 111,000 Australian and New Zealand ICUs, finding that ICU mortality was endogenous with respect to LOS prediction and that log-transformed LOS resulted in more consistent LOS predictions. [108] performed similar comparisons with eight different regression and survival models in a Dutch cohort of 32,000 patients. As in [61], they found significant differences in LOS between survivors and non-survivors, suggesting endogeneity, and recommended implementing log-transformed LOS to improve predictions. They also suggest that Cox-PH and GLM log-link regression models yield the best performance ([108]). [42] used logistic regression techniques in a cohort of 9 million Portuguese inpatients to predict long stays (defined as 2 standard deviations from the mean). Their findings suggest that emergency surgeries, comorbidities and hospital types significantly increase the chance of long stays.

[55] found support vector machines to be the best performing models for long-stay predictions in their cohort of of approximately 14,000 patients. Their model used SOFA scores data from the first 5 days of the ICU stay and had an AUC of 0.82 for long stays. Using a subset of surviving non-prolonged stay patients, they suggest that support vector machines also outperform other machine learning methods, such as CART and random forests ([55]). [68] implemented automatic linear modeling and neural networks to predict cardiac surgery ICU LOS in a small cohort of 185 patients. Neural networks explained 53-73% of training variation, suggesting that Neural networks are better candidates for LOS prediction than conventional

linear regression models. [104] develop survival regression trees to predict ICU LOS as they tend to be easier for health practitioners to understand and interpret. Using data from the Veterans Health Administration of approximately 20,000 admissions for congestive heart failure, they show that regression trees, support vector machines and neural networks are able to account for 58-68% of the variation in LOS.

3.3 Materials and Methods

3.3.1 Study design and patient population

Our research is a retrospective study of two ICUs at the London Health Science Centre (LHSC↓), a teaching hospital located in London, Ontario, Canada. The two ICUs are the Medical-Surgical Intensive Care Unit (MSICU↓) at the University Hospital campus and the Critical Care Trauma Centre↓ (CCTC) at the Victoria Hospital campus. Both are adult intensive-care facilities with 25 and 30 beds, respectively, and they care for general medical, surgical, trauma, oncological, neurosurgical, cardiovascular surgery and transplant patients.

Data was gathered from the Critical Care Information System (CCIS↓) of Ontario's Ministry of Health from the period of January 1st 2015 to December 31st 2016. During that time, a total of N = 4,758 patients were admitted to and dischraged from the two ICUs. Total LOS ranged from 4 hours to 276 days (Table 4.1↓). We define Total ICU LOS as the period of time between the patient's admission to and exit from the ICU (as the result of discharge or death). We define Clinical LOS as the period between the patient's admission to the ICU and the physician's disposition decision (i.e., readiness to transfer/discharge); this ranged from 4 hours to 190 days.

As is common in the literature (e.g. [78, 109]) for LOS prediction models, we excluded cases in which total LOS was $\geq$ 60 days; we also excluded patients transferred to external ICUs. These exclusions resulted in N = 4,696 patients with a lower average of 5.1 days (sd = 7.1 days), as indicated in Figures 4.1↓ and 3.3↓.

Based on the literature review of ICU LOS and patient outcome, predictor variables collected upon patient admission to ICU were (see also Table 4.4↓):

- MODS and its components as created by [59] (see full list in Appendix Table 8.1↓), which we categorized as suggested by [87], resulting in five distinct categories based on scores of: 0; 1 to 4; 5 to 8; 9 to 12; 13 and above.
- NEMS and its components as created by [33] (see Appendix Table 8.2↓), which we sorted according to scores of: 0 to 22; 23 to 29; 30 and above, as suggested by [87].
- Patient characteristics: Sex, Age (less than 39; 40 to 79; 80 and above) as suggested by [87].
- Admission characteristics: Campus, admission source, diagnosis group, patient category (medical/surgical), emergency surgery, ward stay prior to admission, readmission to ICU (same stay), and readmission to ICU (different stay).

For the mortality probability models, the binary variable "IsDeceased" accounts for 26.7% of the population. Following the suggestion of LHSC practitioners, long stays were transformed into two separate binary variables, "IslongStay7" for stays longer than 7 days, and "IsLongStay21" for stays longer than 21 days, which is the long-stay threshold used by LHSC. Because long stays are the focus of these models, we decided to use the complete data set, i.e. keeping the patients with stays > 60 days, thus N = 4,758. To avoid imbalance, we attempted to calibrate the models to maximize the sum of sensitivity and specificity.

We also created a variable "Outcome3", as a categorical variable that includes three distinct outcomes: discharged in less than 21 days; deceased in less than 21 days; and unknown outcome with stay above 21 days (see Table 4.4↓). ICU outcome was modeled based on patient type (death, discharge) and type of LOS (≥ 7 days; or ≥ 21 days). We modeled these as binary variables “IsDeceased”, “IsLong-stay7”, “IsLong-stay21”, and also as a categorical variable called “outcome_3” with three possible outcomes: discharge before 21 days; death before 21 days; and unknown outcome with LOS longer than 21 days.

3.3.2 Statistical analysis

ICU LOS was modeled both in raw format and log scaled (base 10). We used the statistical analysis software R ([85]) to implement the following estimators, with parameters as specified in Table 7.2↓ and 3.3↓:

1. Generalized linear regression models:

1. GLM ↓ (generalized linear models).
2. RLM ↓ (linear model by robust regression).
3. TEM ↓(treatment effects model).

2. Survival analysis models:
    1. Parametric survival regression models, in both exponential and Weibull distributions .
    2. Cox-PH ↓(proportional hazards regression model).
    3. CR-PH ↓ (competing risks proportional hazard model).
3. Supervised learning models:
    1. CART ↓ (classification and regression tree model).
    2. RF ↓ (random forest classification and regression algorithm).
    3. NN ↓ (neural network model) and MULTINOM (multinomial regression).
    4. SVM ↓ (support vector machine model).
    5. SL ↓ (super learner).
4. Two-stage models, in which the first stage is patient outcome (death, discharge) prediction model as the dependent variable. The outcome probability is then weighted into the second stage, which is composed of two separate LOS prediction models: one fit for the discharged patient and another for the deceased patient. The weighted sum of the LOS is the final prediction in the model. The first stage is a logistic regression modeled as a GLM with binomial family, while the second stage was modeled as a GLM with Gaussian family and, alternatively, Cox-PH models.

3.3.3 Model Validation

Each model was run on a training set composed of the first 18 months of patient arrivals (January 1st 2015 to June 30th 2016) and validated on a test set composed of patient arrivals during the following six months (July 1st 2016 to December 31st 2016). For the Clinical LOS models, performance was assessed by R^2, calculated as the squared correlation between predicted LOS and observed LOS; MAE↓ (mean absolute error), NRMSE↓ (normalized root mean squared error), rSD (ratio of predicted and observed standard deviations) and PBias (prediction bias), as implemented by [74]. Mortality, long stays and outcome prediction performance were calculated by:

- Accuracy, defined as: (number of true positive + number of true negative)/total population,
- Sensitivity, or true positive rate (TPR): number of true positive/number of condition positive,
- Specificity, or true negative rate (TNR): number of true negative/number of condition negative,
- PPV↓, or positive prediction value: number of true positive/number of predicted condition positive,
- NPV↓, or negative prediction value: number of true negative/number of predicted condition negative,
- AUC↓, or area under the curve: probability that a classifier will rank a randomly chosen positive instance higher than a randomly chosen negative instance.

3.4 Results

3.4.1 Clinical LOS vs Actual LOS; MODS and NEMS as predictors of ICU LOS

LHSC's total ICU LOS is, on average, 0.9 days longer than Clinical ICU LOS[2]. At a rate of $3,500/bed-day and 2,400 patients/year, this represents approximately 2,100 bed-days and $7.5 million/year in unnecessary stays, as indicated in Table 3.4↓. Subsequetly deceased patients stay longer in the ICU (on average 5.72 days, but with larger variance (sd = 7.98) while discharged patients stay, on average 4.98 days (sd = 6.90) (Welch t-test, $p<0.007536$), as indicated in Figures 4.1↓ and 3.3↓.

Figure 3.4↓ presents the distribution of Clinical LOS by MODS and NEMS, and by patient groups (discharged and deceased) with their respective linear regression plots. For those discharged, the higher the MODS and NEMS, the longer the stay, while the opposite holds true for the deceased. Table 4.2↓ also suggests lower average values for both MODS and NEMS in surviving patients, which can be confirmed ($p< 0.00000$) for both measures. Multicollinearity was tested via generalized collinearity diagnostics ([60]), with no major conflicts found (Table 7.1↓).

[2] At the rate of $3,500/bed-days and,2,400 patients/year, this represents approximately 2,100 bed-days and $7.5 million/year in unnecessary stays.

Figure 4.2↓ shows ICU stay probabilities ("survival") and Figure 3.6↓ shows cumulative incidence LOS in days by MODS and NEMS for the two groups (discharged and deceased). The Kaplan-Meier curves in Figure 4.2↓ show the higher probability of longer stays for high-scoring MODS and NEMS patients. However, we can observe censoring (death, marked as circles within the lines) as more prevalent in earlier periods of ICU stay, particularly for “High MODS”. As, Figure 3.6↓ indicates, when MODS is high (bottom 3 dashed lines), the probability of being "deceased" is higher than being "discharged" at the beginning of the patient stay. For “High MODS” plus “Medium” or “Low NEMS” patients, the probabilities meet at around 5 days; for High MODS/High NEMS patients they meet at 15-16 days. This corroborates with the intuition that High MODS/High NEMS patients are at greater risk of mortality at the beginning of their stay.

3.4.2 Outcome prediction models

3.4.2.1 Mortality prediction

We developed mortality prediction models using “IsDeceased” as a binary dependent variable in logistic regression, CART, RF, SVM, NN and SL (Table 3.6↓ and Figures 7.1↓ and 7.2↓). With the exception of the logistic regression, most models experienced significant loss in performance in terms of AUC (Figures 7.1↓ and 7.2↓) with logistic regression, RF and SL performing best overall. All models have a probability rate of accurately predicting death above 70% in the test data sets, with SL, RF and SVM doing so more than 75% of the time. NN is the best performer in specificity (95.6%) but degrades sharply in testing. RF performs best in testing (77%). NN is the best performer in PPV (training at 83%). CART tests better but degrades to 42.9%. All models perform well in NPV, with test results ranging from 87%-91%. While AUC values in the training data sets range from 0.767-0.988, there is a degradation to 0.63-0.76 in the test data sets. The logistic regression model had the smallest degradation in AUC (1.5% loss); this model, together with RF and SL, are the best performers with an AUC of approximately 0.75-0.76.

3.4.2.2 Long stay prediction

For these models, the independent variables are “IsLongstay7”, and “IsLongstay21” (Tables 3.7↓, 3.8↓) . Out of the total patient population of 4,758, 1,020 patients spent more than 7 days in

the ICU (21.44% of the total patient population), while only 229 spent more than 21 days in (4.81%).

First, we compare the models in which the binary variable "IsLongStay7" is the independent variable (Table 3.7↓). Average sensitivity ranged from 65-99.9% in the training data sets, while dropping to 55-82% in the testing set. CART, RF and SL were the best performing models for this criterion in both data sets, while logistic regression was the model that had the most stable performance (for the logistic regression model parameters and variable outputs, see Table 7.7↓). NN and RF also performed well in specificity, however SVM achieved the best performance in specificity testing. PPV presents a challenge to all models, with test values not surpassing 38%. NPV, on the other hand, maintains very high performance in all data sets, with a lower bound of approximately 83%; SL, meanwhile, maintains 91%. As a result of higher sensitivity and specificity, logistic regression, RF and SL outperform the other models in AUC for LOS >7 days, with test values of 70%, 67.7%, and 69.5% respectively.

Similar to the >7 day prediction, the models for LOS > 21 days present mixed results (Table 3.8↓). In this case, RF, SL and SVM have the highest sensitivity (91-100%) in training, but testing performance drops significantly. Specificity trains well, with the exception of logistic regression (60%) and CART (74%). The best performer is SL, with 81.1% specificity in the testing data set. On average, all models have low positive predictive value but truly excel in negative predictive value, with performance rates above 93% in both training and testing sets. As a consequence of good training metrics, AUC is high but testing performance drops significantly to 60%. Thus, the best performers overall are SL, RF and logistic regression.

3.4.3 Clinical LOS prediction models

The comparative performance of Clinical LOS models is seen in Table 4.10↓ and 3.10↓. These models had 31 fixed variables (for the list of variables, see Table 4.4↓; for a list of model tuning parameters, see Table 7.2↓), while the treatment effects model accounted for the variable "IsDeceased" as endogenous to the model (corresponding to an effect of 1.8 days' increase in LOS, $p<0.001$). Two-stage models had the variable "IsDeceased" as the weighted factor for LOS predictions.

There was no dominant raw-scale Clinical LOS model across all performance measures. While NN, RF and SI achieved high average R^2 in the training data sets, they failed to replicate the same performance in the test data sets. For example, the random forest algorithm had a

degradation of R^2=0.86 to R^2=0.08. Super learner retained more of its prediction power, with an average training R^2=0.61 and test R^2=0.10. Regression and survival models were more stable in terms of R^2, with performances ranging from R^2=0.10 to R^2=0.07 . Two-stage GLM and Cox-PH models did not show significant improvement over their single-stage counterparts, with an equivalent testing R^2=0.10. For detailed parameters and variable outputs of each individual model, see the CART model in Figure 7.11↓, the Cox-PH model in Table 7.8↓, the GLM model in Table 7.9↓, the neural network model in Table 7.10↓ and the random forest model in Table 7.11↓.

In terms of MAE, performance ranged from 2.28-5.37 days in the training data set to 3.84-5.8 days in the test data set. SVM, RLM, SL, Cox-PH and Survival (Weibull) tend to err the least, with the SVM performing best at 3.84 days in the test data set. NRMSE tends to penalize models with higher rates of error, with regression and survival models performing best at the range of 92-95%. SVM, RLM and SL tend to underestimate the predictions as measured in terms of percentage bias (PBIAS). On the other hand, the competing risks model overestimates the predictions (~46% bias) with the treatment effects, random forest and neural networks models doing so moderately. In terms of the ratio of standard deviations, the rSD measure shows that only the survival models are able to generate the variance observed in the Clinical ICU LOS, with few models achieving more than a 0.8 ratio of the observed standard deviation.

Log transforming the Clinical LOS showed improvements over the raw LOS, in the form of higher R^2 values in testing (see standardized residual plots in Figures 7.3↓, 7.4↓, 7.5↓ and 7.6↓). Here, machine learning models train significantly better than other models, with RF (R^2=0.80), NN (R^2=0.58), and SL (R^2=0.60) as the best performers. But, again, performance dropped significantly in the test data sets, in which these models fail to reach values higher than R^2=0.11, the best performer being the treatment effects model at R^2=0.12. Regression and survival models are more consistent in terms of R^2, with smaller performance losses from training to testing (compared to machine learning models).

In terms of MAE performance, the log scaling yielded results ranging from 2.38-4.64 days (RF offered the best best training performance, followed by NN and SL)[3]. R^2, MAE and NRMSE were computed on the back-transformed "day" scale using Duan's smearing estimator . The testing of MAE performance was fairly uniform across all models, with regression and machine learning models erring the least (4.01-4.13 days). The supervised machine learning

[3] R^2, MAE and NRMSE were computed on the back-transformed "day" scale using Duan's smearing estimator

models train better in terms of NRMSE (particularly NN and RF, with 65%), but test similarly to the regression and survival models, which range from 95-163%.

3.4.4 Patient outcome prediction

As an alternative to predicting death/discharge and short/long stay separately, we also developed the categorical variable "Outcome_3", in which we have either discharge in under 21 days, death in under 21 days or a long stay > 21 days (with no prediction of death/discharge). The performance of these models is indicated in Table 3.11↓. Again, RF seems to perform best with an accuracy rate of approximately 99% (in training) and 75.5% (in testing), followed closely by multinomial regression and CART.

3.5 Discussion

In this paper we developed NEMS- and MODS-based ICU outcome and LOS prediction models using several different methodologies, including well known statistical models and supervised machine learning techniques. Our first finding was that, on average, deceased LHSC ICU patients had longer LOS than discharged patients. This difference is significant and unexpected. It contradicts most of the literature on ICU LOS prediction, in which deceased patients are estimated to have shorter stays ([35, 13, 72, 27, 112, 61, 108, 109]). Overall, our models suggest that MODS and NEMS measures provide reasonable predictions in terms of mortality as suggested by [87], with random forest and super learner performing significantly better in training sets and at least as accurately as logistic regression. Similar performance can be seen for long stays >7 days, but not for longstays >21 days or individual clinical LOS.

Long stay prediction models are the least explored in the literature, and represent the category of patient stay that logistic regression models are least able to predict ([16]). Despite predicting long stay LOS thresholds ($\geq$ 7 days; or $\geq$ 21 days), the models diverge in terms of significant predictors. Most notably, in the logistic regression model, neither MODS nor the ICU admission source is applied to the 21 day model. The absence of MODS seems counter-intuitive as severity scores ([55]) and admission source ([32]) are commonly present and extremely influential in the generation of such predictions.

Our results show no clear dominant LOS prediction model, with different metrics privileging different models. Log transformation of the Clinical ICU LOS seems to improve

MAE and NRMSE in general, while mortality in the ICU may be considered an omitted endogenous variable affecting LOS [78], and limiting the interpretation of traditional regression models. The supervised machine learning models perform very well in training sets but do not differ from traditional models in testing set performance. As noted in [108, 109], due to differences in covariates and exclusion criteria, comparison between different studies is often inappropriate and, when such comparisons are made, it is often by way of R^2 measures. Common R^2 values range anywhere from 0.05-0.25, which is consistent with our findings. Careful analysis of the models based on the latter scores reveals the significant weight of detailed diagnostics and comorbidity variables ([42, 78, 84, 107, 108, 112]), which our models carry only in the form of a simplified diagnostics group. We expect prediction performance improvements if such variables become available in Canadian ICUs. Nevertheless, our models incorporating MODS and NEMS measures are already suitable replacements for APACHE, SAPS and SOFA scores for prediction purposes.

We have also noted that no LOS prediction model is able to reach the entire range of the observed clinical LOS and that there is a gradual increase in residuals for long-stay patients (as seen by the low overall rSD measures and residual plots). The difficulty in predicting the long tail is one drawback of the regression models ([108]), but its impact on supervised machine learning models has not been sufficiently explored.

Our research presents shortcomings and indicates that further challenges need to be addressed. MODS and NEMS collected upon patient arrival in the ICU may provide masked information about the patient's true health status. For the sake of improved performance, it may be better to include 24-hour or 72-hour measurements that may be carried out after the administration of the first intensive treatments, as suggested by [98]. We expect that such added measurements would provide a more reliable source of information about the patient's actual health status, which may be unclear upon admission to the ICU. Finally, our data set is relatively small with less than 5,000 patients, while other cohorts include 100,000-200,000 patients ([61, 78]). Our intention is to run the same analysis in a province-wide data set, which may help with degradation issues.

3.6 Conclusion

This research aimed to provide managerial decision support prediction models that utilize metrics common to Canadian hospitals such as MODS and NEMS, which are not commonly available in the literature. We did so by incorporating other methods than the most commonly adopted linear regressions, such as survival models and supervised machine learning models.

We have shown that MODS and NEMS can be used as clinical ICU LOS and outcome predictors and that there is much to be gained by adding the survival and supervised learning models to ICU LOS prediction literature and practice.

Table 3.1 ICU LOS Descriptive statistics

Variable	N	Mean	St. Dev.	Min	Pctl(25)	Median	Pctl(75)	Max
Total ICU LOS	4,758	6.4	11.1	0.2	1.7	3.4	7.1	276.8
Clinical ICU LOS (before exclusions)	4,758	5.5	9.2	0.2	1.1	2.5	5.9	190.9
Clinical ICU LOS (after exclusions)	4,696	5.1	7.1	0.2	1.1	2.5	5.8	59.0

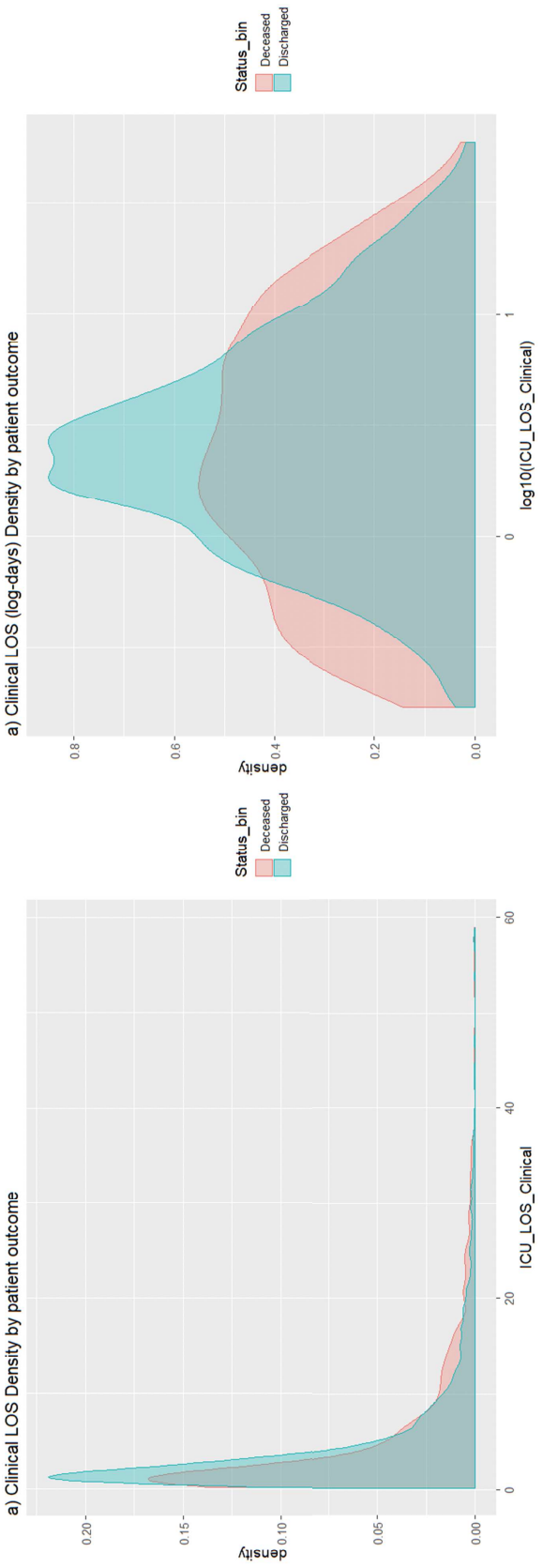


Figure 3.1: Clinical LOS by patient outcome

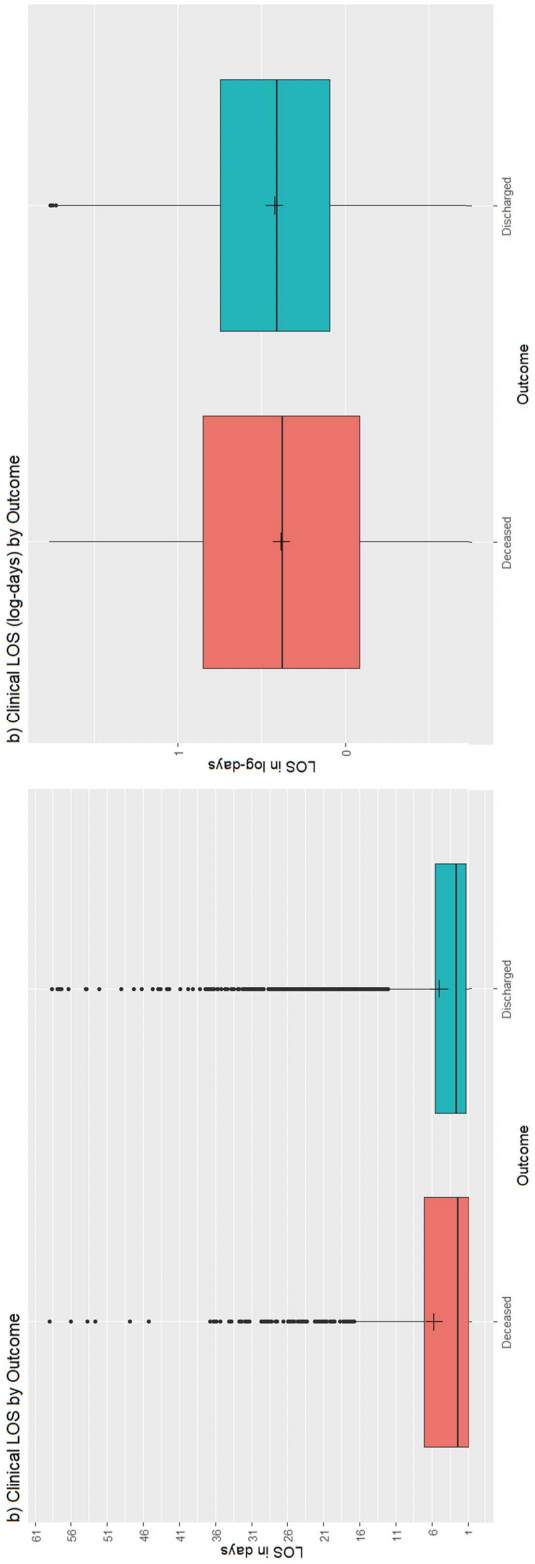


Figure 3.2: Clinical LOS by patient outcome (Box-Plot)

Table 3.2 List of variables

Variable type	Subcategory	Variable	Nature (units)	Description
Outcomes	LOS	ICU_LOS_Clinical	continuous	clinical length of stay in days
	Death	IsDeceased	binary	death in ICU
	Long stay	IsLongStay7	binary	LOS > 7 days
	Long stay	IsLongStay21	binary	LOS > 21 days
	disposition	outcome_3	categorical	discharged < 21 days, deceased <21 days, long stay (above 21 days)
Covariates	MODS	MODSGlasgowComaScoreCat	categorical	Glasgow Coma Scale score, (None, Minimal, Mild, Moderate, Severe)
		MODSHaematologicCat	categorical	Platelet count, (None, Minimal, Mild, Moderate, Severe)
		MODSHepaticCat	categorical	Serum bilirubin level, (None, Minimal, Mild, Moderate, Severe)
		MODSPressureAdjustedHeartRateCat	categorical	Pressure-adjusted heart rate, (None, Minimal, Mild, Moderate, Severe)
		MODSRenalCat	categorical	Serum creatinine level, (None, Minimal, Mild, Moderate, Severe)
		MODSRespiratoryRatioCat	categorical	PaO2/FIO2 ratio, (None, Minimal, Mild, Moderate, Severe)
		MODSScoreCategory	categorical	a-0 score, b-1 to 4, c-5 to 8, d-9 to 12, e-13 and above
	NEMS	IsArterialLine	binary	introduction of arterial lines
		IsbasicMonitoring	binary	monitoring of basic vital signs
		IsCentralVenous	binary	central venous medication
		IsDialysis	binary	any dialysis techniques
		IsDischarged	binary	discharged from ICU
		IsExtraCorporealMembrane	binary	extra-corporeal membrane life support
		IsInterventionOutside	binary	Specific interventions outside the ICU: such as surgical intervention or diagnostic procedure;
		IsIntraAorticBalloonPump	binary	Intra-aortic balloon pump
		IsIntracranial	binary	Intracranial pressure monitoring
		IsIntravenous	binary	intravenous medication
		IsOtherIntervention	binary	other interventions such as: endotracheal intubation, pacemaker, cardioversion, endoscopy, emergency operation in the past 24 h, gastric lavage; X-rays, echocardiography, electrocardiography, dressings
		IsOtherIntravenous	binary	other types of intravenous medication
		IsVentilation	binary	mechanical ventilation
		NEMSScoreCategory	categorical	a-0 to 22, b-23 to 29, c-30 and

			above
Patient characteristics	Age_category	categorical	18-39, 40-80, 80 and above
	Sex	binary	male (0), female (1)
Admission characteristics	IsEmergencySurgery	binary	emergency surgery
	IsLOS_before	binary	hospital admission to other prior to ICU
	ReadmissionDifferentStay	binary	readmission to ICU from previous hospital admission into ICU
	ReadmissionWithinSingleStay	binary	readmission to ICU in the same hospital stay
	Campus	categorical	MSICU, CCTC (ICU's from different hospital sites)
	ICUAdmissionDiagnosis_group	categorical	\strikeout Cardiovascular/Cardiac/Vascular, Gastrointestinal, Neurological, Other, Respiratory, Trauma
	ICUAdmissionSource	categorical	Emergency Dept., OR, Other Hospital, Stepdown Unit, Ward
	PatientCategory	categorical	Medical, Surgical

Table 3.3 Model References

Generalized linear regression models

GLM (generalized linear models), with log link and both Gaussian and Gamma families, and step-wise backward selection as implemented in R Stats package [85] .

RLM (linear model by robust regression) using an M estimator fitted via iterated re-weighted least squares as described by Huber [56] and implemented in the R MASS package [111].

TEM (treatment effects model), to estimate linear regression parameters and treatment effects ("deceased-in-ICU") in the presence of endogeneity.
This model is based on Heckman [54] and implemented by Spieker in the R Endogenous package [12].

Survival analysis models

Parametric survival regression models, in both exponential and weibull distributions as described and implemented by Therneau & Grambsch in the R Survival package [101, 100].

Cox-PH (proportional hazards regression model), as formulated by Andersen & Gill [6]

and implemented by Therneau & Grambsch in the R Survival package [101, 100].

CR-PH (competing risks proportional hazard model), as described by Gray [47] and Fine & Gray [40],

and implemented in the R Cmprsk package by Gray [23] and Scrucca et al. [96].

Supervised learning models

CART (classification and regression tree model), a regression tree model fit by binary recursive partitioning

as implemented in the R Rpart package by Ripley [28] and Therneau, Atkinson & Ripley [102].

RF (random forest classification and regression algorithm) as described by Breiman [25]

and implemented in the R RandomForest package by Liaw & Wiener [14].

NN (neural network model) and MULTINOM (multinomial regression), as implemented in the R NN package by Venables & Ripley [111].

SVM (support vector machine model), as implemented in the R E1071 package by Meyer et al [34].

SL (super learner), as implemented in the Super Learner package by Polley et al [38] (for continuous and binary variables), we used the same GLM, CART, RF, SVM algorithms as above, which should guarantee performance improvements as shown in [106].

Table 3.4 Clinical LOS by outcome

Outcome	N	Mean	SD	Min.	1st Qu.	Median	3rd Qu	95th pctl	99th pctl	Max
Deceased	992	5.72	7.98	0.17	0.81	2.37	7.12	23.39	35.36	58.98
Discharged	3,704	4.98	6.90	0.17	1.23	2.56	5.56	18.83	34.31	58.67

Welch Two Sample t-test t = -2.676, df = 1412.9, p-value = 0.007536

Table 3.5: MODS and NEMS by outcome

	N	Mean MODS	SD	Min	Pctl(25)	Median	Pctl(75)	Max
Deceased	992	6.7	3.1	0	4	6	9	19
Discharged	3,704	4.6	2.8	0	3	4	6	16

Welch Two Sample t-test t = 19.14, df = 1445.2, p-value < 2.2e-16

NEMS

	N	Mean NEMS	SD	Min	Pctl(25)	Median	Pctl(75)	Max
Deceased	992	36.8	8.5	15	32	38.5	44	56
Discharged	3,704	31.6	8.5	0	27	32	38	56

Welch Two Sample t-test t = 17.168, df = 1561.9, p-value < 2.2e-16

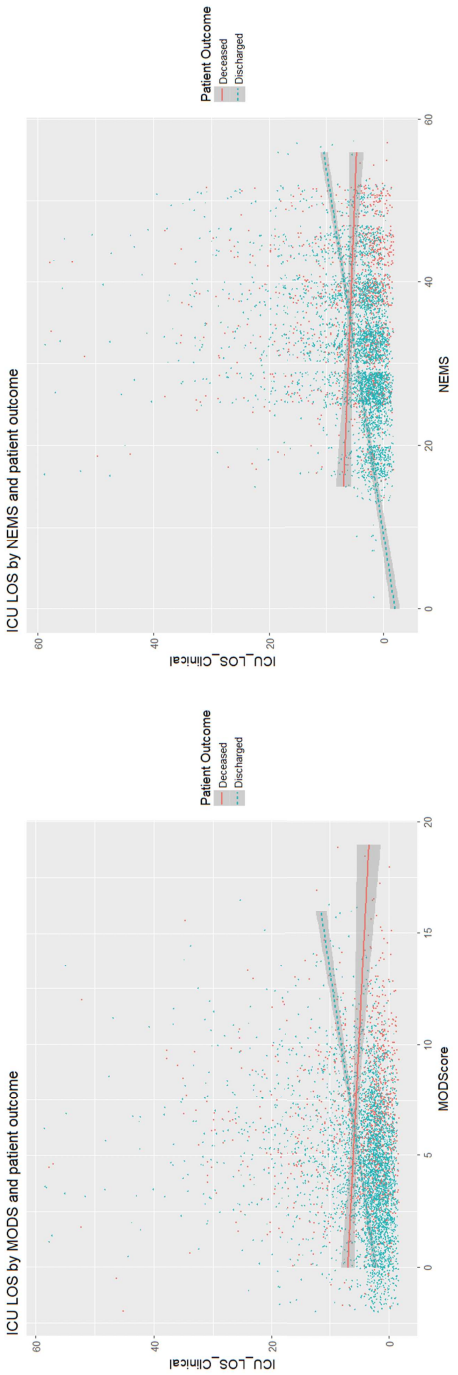


Figure 3.3: Raw scale Clinical LOS with regards to MODS and NEMS by patient outcome

Table 3.6 Mortality prediction average performance

Model	Threshold*	Sensitivity		Specificity		PPV		NPV		AUC	
		train	test	train	test	train	test	train	test	train	test
Logistic Regression	0.210	0.739	0.718	0.680	0.715	0.378	0.408	0.909	0.892	0.779	0.767
CART	0.263	0.695	0.650	0.670	0.704	0.403	0.429	0.903	0.889	0.767	0.742
Neural Network	0.386	0.806	0.707	0.956	0.531	0.831	0.284	0.947	0.877	0.954	0.638
Random Forest	0.224	0.843	0.665	0.741	0.774	0.468	0.412	0.945	0.889	0.877	0.751
Super Learner	0.316	0.970	0.767	0.931	0.325	0.792	0.345	0.991	0.912	0.988	0.759
Support Vector Machine	0.042	0.864	0.671	0.816	0.663	0.549	0.365	0.957	0.887	0.882	0.708

* Threshold corresponding to the best sum of sensitivity and specificity

Table 3.7 Long-stay prediction average performance (>7 days)

IsLongStay > 7 days	Threshold*	Sensitivity		Specificity		PPV		NPV		AUC	
		train	test	train	test	train	test	train	test	train	test
Logistic Regression	0.229	0.653	0.689	0.694	0.658	0.359	0.381	0.884	0.874	0.723	0.701
CART	0.192	0.770	0.725	0.555	0.563	0.313	0.336	0.902	0.870	0.699	0.646
Neural Network	0.415	0.827	0.696	0.936	0.506	0.772	0.301	0.954	0.845	0.946	0.606
Random Forest	0.362	0.997	0.637	0.999	0.647	0.996	0.355	0.999	0.854	1.000	0.677
Super Learner	0.246	0.975	0.824	0.847	0.485	0.627	0.305	0.992	0.910	0.968	0.695
Support Vector Machine	0.041	0.862	0.557	0.856	0.678	0.611	0.345	0.959	0.834	0.888	0.628

* Threshold corresponding to the best sum of sensitivity and specificity

Table 3.8 Long-stay prediction average performance (>21 days)

IsLongStay > 21 days	Threshold*	Sensitivity		Specificity		PPV		NPV		AUC	
		train	test	train	test	train	test	train	test	train	test
Logistic Regression	0.037	0.813	0.574	0.606	0.704	0.088	0.100	0.987	0.951	0.773	0.635
CART	0.042	0.661	0.450	0.747	0.691	0.174	0.094	1.007	0.978	0.805	0.612
Neural Network	0.368	0.76	0.474	0.996	0.649	0.898	0.072	0.989	0.955	0.884	0.526
Random Forest	0.237	0.998	0.508	1.000	0.738	0.996	0.093	0.999	0.938	1.000	0.622
Super Learner	0.144	1.000	0.410	1.000	0.811	1.000	0.111	1.000	0.960	1.000	0.630
Support Vector Machine	0.020	0.913	0.517	0.957	0.667	0.542	0.086	1.002	0.943	0.974	0.575

* Threshold corresponding to the best sum of sensitivity and specificity

Table 3.9 Clinical ICU LOS (days) performance comparisons for different models

Model	R2		MAE		NRMSE (%)		PBIAS (%)		rSD	
	train	test	train	test	train	test	train	test	train	test
CART	0.12	0.06	4.16	4.49	94.00%	97.30%	0.0%	-1.5%	0.34	0.37
Cox-Ph	0.09	0.10	4.07	4.28	95.30%	95.20%	-2.9%	-5.1%	0.33	0.35
CR-ph	0.07	0.08	5.37	5.5	111.20%	109.80%	44.3%	41.7%	0.71	0.73
GLM (Gmma)	0.09	0.09	4.14	4.37	95.20%	95.30%	0.3%	-2.1%	0.33	0.34
GLM (Gaussian)	0.10	0.09	4.17	4.4	95.20%	95.30%	0.0%	-3.3%	0.31	0.32
Neural Network	0.49	0.03	3.44	5.84	71.20%	120.30%	1.0%	1.7%	0.72	0.85
Random Forest	0.86	0.08	2.28	4.43	52.70%	96.20%	2.8%	-1.1%	0.56	0.34
RLM	0.09	0.11	3.69	3.91	98.60%	98.50%	-31.7%	-33.9%	0.19	0.19
Super Learner	0.61	0.10	3.1	4.21	77.00%	95.00%	-7.8%	9.5%	0.33	0.27
Support Vector Machine	0.12	0.10	3.49	3.84	100.30%	100.90%	-43.8%	-45.6%	0.18	0.18
Survival (Exponential)	0.09	0.10	4.14	4.36	95.30%	95.10%	0.5%	-1.9%	0.34	0.35
Survival (Weibull)	0.10	0.10	4.07	4.28	95.30%	95.10%	-2.9%	-5.1%	0.43	0.44
Treatment Effects	0.10	0.15	4.3	4.47	5.20%	92.50%	6.6%	11.4%	0.33	0.44
Two-stage Model (Cox-PH)	0.09	0.11	4.1	4.28	95.20%	94.20%	-1.4%	-4.2%	0.32	0.32
Two-stage Model (GLM)	0.09	0.10	4.18	4.37	95.40%	94.80%	20.0%	-3.5%	0.3	0.31

Table 3.10 Clinical ICU (days) performance comparisons for different log-transformed models

Model*	R^2*		MAE (days)*		NRMSE (%)*		PBIAS (%)		rSD	
	train	test	train	test	train	test	train	test	train	test
CART	0.10	0.12	4.10	4.15	99.2%	94.4%	-26.9%	10.6%	0.22	0.27
Cox-Ph	0.07	0.09	4.25	4.34	97.4%	95.1%	-16.3%	-27.2%	0.27	0.30
CR-ph	0.09	0.10	4.14	4.64	97.6%	93.1%	-30.3%	-16.4%	0.30	0.33
GLM (Gmma)	0.06	0.08	4.15	4.13	100.1%	96.0%	-28.3%	-29.8%	0.25	0.26
GLM (Gaussian)	0.08	0.11	4.09	4.10	99.7%	95.1%	-29.0%	-25.7%	0.20	0.22
Neural Network	0.58	0.06	2.82	6.59	65.0%	163.1%	-14.6%	-24.1%	0.88	1.42
Random Forest	0.80	0.11	2.38	4.01	65.2%	95.1%	-12.9%	3.3%	0.44	0.21
RLM	0.08	0.10	4.07	4.09	99.5%	96.7%	-31.4%	18.8%	0.21	0.23
Super Learner	0.60	0.10	3.16	4.07	83.6%	95.9%	-28.0%	22.4%	0.26	0.20
Support Vector Machine	0.16	0.05	3.37	4.34	94.1%	110.9%	-30.7%	-34.5%	0.28	0.32
Survival (Exponential)	0.07	0.09	4.13	4.14	98.1%	96.4%	-13.9%	-51.9%	0.34	0.35
Survival (Weibull)	0.08	0.09	4.25	4.34	97.4%	95.0%	-15.5%	-28.7%	0.43	0.44
Treatment Effects	0.09	0.12	3.95	4.15	5.3%	96.1%	-23.9%	-21.9%	0.33	0.44
Two-stage Model (Cox-PH)	0.07	0.11	4.24	4.28	97.3%	93.5%	-16.5%	79.8%	0.32	0.32
Two-stage Model (GLM)	0.07	0.10	3.90	4.30	99.5%	101.1%	-33.4%	-151.6%	0.30	0.31

R^2, MAE and RMSE were computed on the back-transformed "day" scale using Duan's smearing estimator

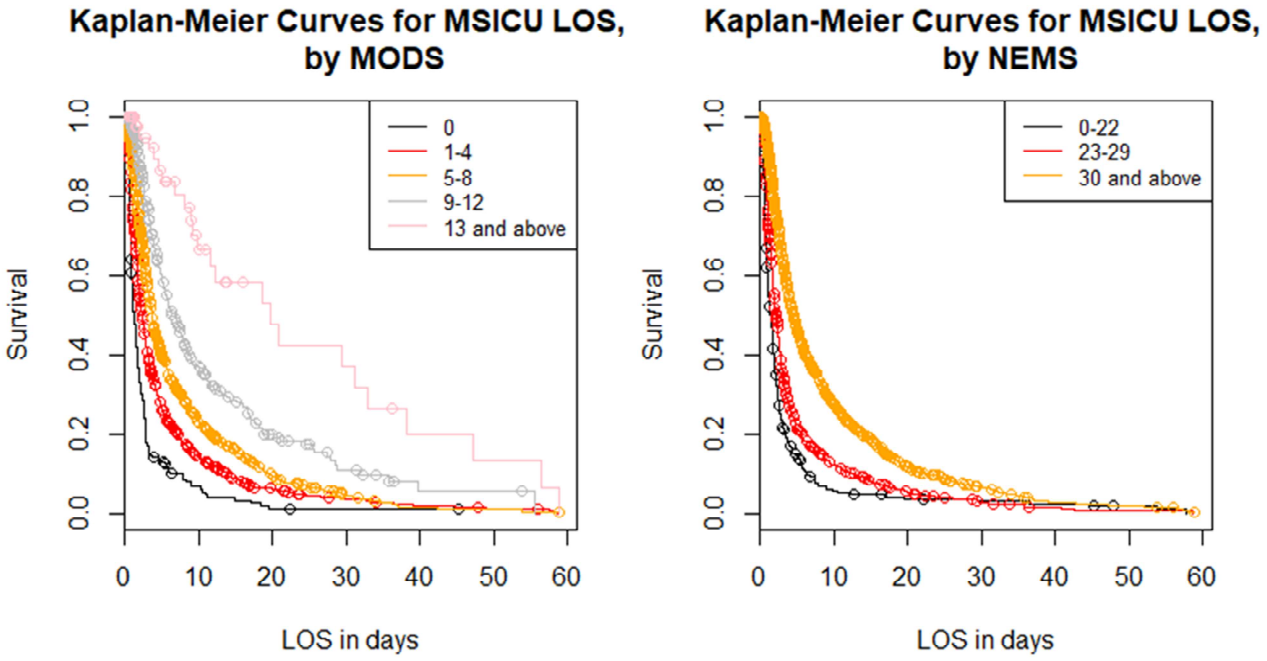


Figure 3.4: Kaplan-Meier Curves for Survival Models, by MODS and NEMS

Table 3.11 Patient outcome prediction average performance

Model	Accuracy	
	train	test
CART	0.815	0.740
Multinomial regression	0.769	0.754
Random Forest	0.989	0.755
Neural Network	0.881	0.662
Support Vector Machine	0.938	0.701

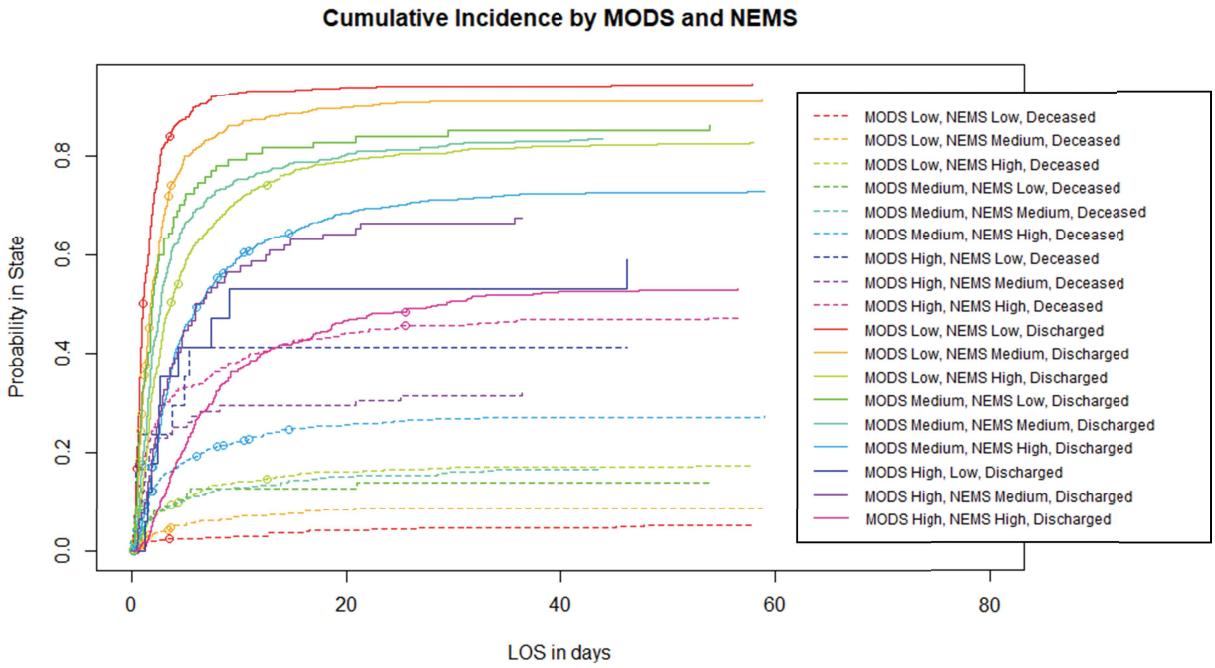


Figure 3.5: Competing risks for Clinical LOS prediction

Chapter 4

ICU Length of Stay: Individual predictions and short term capacity planning using a Parametric Accelerated Failure Time Model

Abstract

Prediction of a patient's length of stay (LOS↓) may help Intensive Care Units (ICU↓) to plan future bed allocations and staffing levels and to schedule surgeries. Present-day ICUs collect a number of patient data and aggregate it in the form of severity scores such as APACHE and SOFA, which have been used in the literature for individual patient LOS predictions and LOS benchmarking. In Canada, however, patient severity is often measured by the "Multi-Organ Dysfunction Syndrome" (MODS) ↓ score and nursing workloads by the "Nine Equivalents of Nursing Manpower Use Score" (NEMS)↓. To the best of our knowledge, these have yet to be used for LOS prediction and capacity planning. Using MODS and NEMS as covariates, we developed a parametric Accelerated Failure Time model (Weibull AFT survival model) serving two purposes. First we assess LOS prediction performance considering ICU arrival (day 1) measures, and compare it to an updated prediction for patients whose ICU stay reaches day 3. Second, we use each patient's individual survival function to generate individual survival probabilities and pool the patient cohort probabilities to form expected short-term bed needs. Using data from a large Canadian university hospital, we show that, although individual LOS prediction is prone to significant error, aggregate bed occupancy is more predictable and can be used reliably for short-term resource capacity planning purposes.

4.1 Introduction[1]

ICU beds are usually the most expensive beds in any hospital and congestion in capacity-constrianed ICUs has been linked to increased mortality ([18, 31]). This likelihood prompts hospitals to plan ICU patient discharge as efficiently as possible ([278, 22, 50, 281]). The University Hospital (UH) campus of the London Health Sciences Centre (LHSC) ↓ is no different, with ICU beds costing as much as $3,500/day and being highly utilized ([67]). UH is a 400-bed Canadian teaching hospital with roughly 17,000 admissions per year, of which 1,000-1,200 are medical-surgical intensive care unit (MSICU) ↓ patients ([283]).

Several models have been proposed in the literature to predict ICU LOS, such as [108, 109]. Most are based on the Acute Physiology and Chronic Health Evaluation (APACHE↓) score, the Simplified Acute Physiology Score (SAPS↓) and the Sequential Organ Function Assessment (SOFA↓) score, as noted in [76, 93, 107, 108, 112]. None of these are useful to LHSC or other hospitals in Ontario, which are required by the provincial government to collect for reporting purposes, upon patient ICU admission, the Multiple Organ Dysfunction Syndrome (MODS) score ([59]), and, daily, the Nine Equivalents of Nursing Manpower Use Score (NEMS) ([33]). This essay addresses the following questions:

1. Can NEMS and MODS be used in a parametric survival model for individual patient ICU LOS prediction?
2. Does updating ICU patient metrics after ICU arrival improve ICU LOS prediction?
3. Can the individual predictions based on NEMS and MODS be used for short-term bed capacity planning?

In order to answer these three questions, we have developed a parametric Accelerated Failure Time (AFT↓) model using MODS and NEMS. Our model not only estimates an individual patient's ICU LOS and daily expected survival probability, but also pools thesse probabilities to predict short-term ICU bed capacity requirements.

[1] This work was supported by: The Ontario Trillium Scholarship program (OTS); an Ivey International Centre for Health Innovation (IICHI)/3M Canada MITACS grant; the Natural Sciences and Engineering Research Council of Canada (NSERC), and the Universidade Federal do Parana (UFPR). These funding sources had no active role in the study design, collection, analysis, interpretation of data, writing of the report or decision to submit the article for publication. This research is approved by Western University's and LHSC's Research Ethics Board under the file # REB 105583.

4.2 Literature Review

4.2.1 ICU Bed Capacity Management

The ICU bed capacity management literature focuses on queuing models, as in [15, 49], queuing networks, as developed by [26, 30], dynamic programming (e.g. [31]) and discrete-event simulation models ([39]), as seen in recent extensive surveys by [24, 300, 301]. [26] developed a heuristic for an ICU/step-down queuing network with blocking and found optimal capacity allocations based on blocking probability minimization. [31] on the other hand, study the impact of congestion on discharge decisions in capacity-constrained ICUs, and provide optimal discharge policies based on risk of readmission. [21] focus on bed specialization and develop a bed capacity optimization model based on hospital wing partitioning. They find that, under certain conditions, wing formation may lead to increased specialization and shorter LOS. [15] use queuing to optimize distribution of ICU and step-down beds and to find optimal allocations depending on patient step-down readiness and nurse-patient ratios. [18] and [65] observe that ICU congestion influences patient admission and outcomes originating in the Emergency department (ED) ↓. Thus, they developed optimal ICU admission policies based on severity scores to reduce readmission and LOS. [39] use nursing workload scores (NEMS) to simulate individual patients' ICU LOS under congestion. They find that simulating daily NEMS transitions accurately captures patient readiness to step down and can be used for long-term bed capacity planning purposes. [43] and [97] and [89] have also successfully built models based on different severity scores for capacity planning and patient flow management.

4.2.2 LOS prediction

LOS prediction models can be traced back to the development of the APACHE model for hospital mortality risk prediction [312, 313, 110]. [112] later used APACHE scores in a multiple linear regression model for predicting ICU LOS. APACHE (in its current iterations APACHE III and IV) has outperformed other scores such as SAPS II in LOS prediction, in part, due to the inclusion of an extensive list of diagnostics and comorbidities ([107]). Multiple linear regression APACHE-based models have also been used by [61, 78], who have noted significant differences in performance between discharged and deceased patients. They claim ICU death increases ICU LOS variance, thereby reducing the accuracy of LOS prediction models.

[108, 109] were among the first scholars to challenge the use and adequacy of multiple linear regression models for ICU LOS prediction. They compared those models with other techniques such as log-link generalized linear models and non-parametric survival models and found the latter to be dominant over OLS multiple regressions.

Later, [9] and [31] recognized the benefit of updating measures for prediction of prolonged stays, claiming that LOS predictions may require better information than the one available upon patient arrival in the ICU. Their model updates severity score measurements at day 5 of the ICU stay, significantly improving prediction accuracy.

More recent methodologies, such as logistic regression as developed by [42], support vector machines as described by [55], neural networks ([99, 68, 93]), classification trees ([104]), random forests ([89]) and super learner ([84, 93]), have also been used for ICU LOS prediction ([93]), particularly to identify mortality and long-stay patients.

4.2.3 ICU survival analysis and AFT models

Survival analysis models applied to ICUs tend to focus on hazard ratios, as opposed to LOS prediction ([333, 334, 335, 336, 337, 338, 339]). [340] find age, Glasgow Coma Scores and Abbreviated Injury Scores to be significant predictors of ICU discharge and mortality, with variable effects during prolonged ICU stays. [341] build on Clark and Ryans' work and suggest the use of competing risks models if patients have multiple discharge destinations. [342] compare Cox and AFT models and advocate for the inclusion of time-varying covariates in ICU survival analysis as they show that hazard ratios vary over time. [343] developed a Cox proportional hazards model to estimate the impact of ventilation and renal replacement therapy on ICU survival and found APACHE II scores and ventilation to be inversely related to ICU survival. [344] observed the significance of severity scores in ICU mortality but failed to find an association between the SOFA score and post-hospital mortality. Finally, [345] use a multi-state proportional hazards model and found age, comorbidities, and ICU LOS to be associated with higher rates of post-ICU mortality, long-term care and readmission.

4.2.4 Contributions of this paper

Our Weibull Accelerated Failure Time (AFT) model contributes to the ICU capacity management literature and practice in a number of important ways. To our knowledge, we are

the first to use NEMS and MODS as covariates in an AFT model for ICU LOS prediction. Second, we use the Weibull AFT model not with the intention of finding a variable impact on hazard rates, but to provide conditional daily survival estimates and expected time to event (LOS discharge) for each patient upon arrival in the ICU. Third, we show that covariate updating performed on prolonged-stay patients provides a better prediction of LOS. Finally, we show how the survival probabilities of the AFT model can be pooled to estimate future bed occupancy. This novel approach to the AFT model can, in turn, be used to improve short-term bed capacity management.

4.3 Materials and Methods

4.3.1 Study design and patient population

Our research is a retrospective study of the Medical-Surgical Intensive Care Unit (MSICU↓) at the London Health Science Centre's (LHSC↓) University Hospital (UH) campus. UH is a teaching hospital located in London, Ontario, Canada. The MSICU is an adult intensive-care facility with 25 beds, caring for general medical, surgical, trauma, oncological, neurosurgical, cardiovascular surgery and transplant patients.

Data was gathered from the Critical Care Information System (CCIS↓) of Ontario's Ministry of Health from the period of January 1st 2015 to December 31st 2016. There were a total of N = 2,195 patients admitted into the ICU with a total LOS ranging from 0 to 276 days (Table 4.1↑). We define total ICU LOS as the period of time between patient admission to the ICU and patient exit from the ICU (due to discharge or death). We define Clinical LOS as the period between admission to the ICU and the physician's disposition decision (i.e., readiness to transfer/discharge/death), which ranges from 0 hours to 190 days, averaging 5.63 days (sd = 10.14 days).

As is common in the literature (e.g. [78, 109]), we excluded cases in which total LOS $\geq$ 60 days. These exclusions resulted in N = 2,176 patients with a lower LOS average of 5.05 days (sd = 7.40 days), as seen in Table 4.1↑ and Figure 4.1↑.

MODS and NEMS are distributed differently according to patient outcome (Table 4.2↑), with deceased patients scoring higher than their discharged counterparts in both measures

(p<000). Similarly, medical patients tend to have higher MODS and NEMS measures than surgical patients (see Table 4.3↓) and apparently experience longer stays (see Figure 4.1↑).

Based on the literature, we used the following predictor variables available upon patient admission to ICU (day 1) and after completion of median time (3 days of ICU stay) (see also Table 4.4↑):

- MODS and its components, including: platelet count, serum bilirubin, serum creatinine, pressure adjusted heart rate, Glasgow coma score, respiratory ratio. These variables were categorized (see Table 8.1↓) according to the scale created by [59]. The total MODS is based on a 24-point scale (see Table 8.1↓) which we sorted, as suggested by [87], into five distinct categories based on the following scores; 0; 1 to 4; 5 to 8; 9 to 12; 13 and above.
- NEMS and its components, collected upon admission (Day 1) and after 3 complete days (Day 3), including: arterial line, intravenous medication, intracranial membrane, dialysis, intervention inside the ICU, intervention outside the ICU, central venous line, mechanical ventilation. All of these variables are treated as binary variables and become components of NEMS, which is based on a 56-point scale developed by [33] (Table 8.2↓). We also sorted NEMS into three categories based on the following scores: 0 to 22; 23 to 29; 30 and above, as suggested by [87].
- Patient characteristics: sex, age (less than 39; 40 to 79; 80 and above) as suggested by [87].
- Admission characteristics: admission source, diagnosis group, patient category (medical/surgical), emergency surgery, ward stay prior to admission, readmission to ICU (same stay), readmission to ICU (different stay).

4.3.2 Statistical analysis

4.3.2.1 Weibull AFT model

We used the statistical analysis software R ([85]) to implement a Weibull AFT model (parametric survival regression) as described and implemented by Therneau & Grambsch in the R *Survival* package ([101, 100]). The Weibull AFT model has a distribution of time to event T as a function of covariates written as in Equation 5.1↓, where β_0 is the intercept, β_n is the coefficient

of the covariates x_{ij} for patient i and time j, σ is the shape parameter for ε which follows the extreme minimum value distribution $G(0, \sigma)$.

$$(5.1)\ ln(T) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} \ldots + \beta_n x_{ij} + \sigma\varepsilon$$

The Weibull regression model allows for the description of the covariates in terms of Hazard rates ↓(HR), i.e., the instantaneous risk of ICU exit given survival up to time t, per Equation 5.2↓:

$$(5.2)\ h(t, x_{ij}, \beta_i, \lambda) = \lambda\gamma t^{\lambda - 1} e^{-\lambda\beta_i x_{ij}} = h_0(t)\, e^{\theta_i \mathbf{x}_{ij}}$$

The model also allows for and Event time ratios ↓(ETR), i.e., relative change in survival time, per Equation 5.3↓:

$$(5.3)\ h(t, x_{ij}, \beta_i, \lambda) = \lambda\gamma(t\, e^{-\beta_i x_{ij}})^{\lambda - 1} e^{-\beta_i x_{ij}}$$

Where $\gamma = e^{(-\beta_0)/(\sigma)} = e^{\theta_0}$, $\theta_i = -(\beta_i)/(\sigma)$ and the baseline hazard function is $h_0(t) = \lambda\gamma t^{\lambda - 1}$, and scale parameter$\lambda = 1 / \sigma$.

Model development and variable selection was made via an ANOVA test (Figure 4.3↓), step-wise backward elimination using Harrell's *RMS* package ([356]) *fastw()* function with p-value as the elimination criterion, and, most importantly, with the final input from the MSICU clinical staff. The final model was built converting the Weibull distribution parameters to Hazard rates and Event time ratios and were extracted via Hubeaux and Rufibach's *SurvRegCensCov* package (see [357]). Diagnostic and predictive plots were made with Brostrom's *EHA* package ([358]) and Lesaffre et al.'s *smoothSurv* package ([37]).

Following the need to investigate the impact of variable updating in LOS predictions, the time threshold for the model update was chosen based on consultation with LHSC's clinical staff. The threshold decided upon was 3 complete days, which is the empirical median LOS in the MSICU. The updated model includes only patients that survived to have measurements taken on day 3 and is comprised of the original covariates collected upon ICU admission with the addition of the updated NEMS and its components at day 3.

4.3.2.2 Short-term capacity planning model.

To predict the expected occupancy of the ICU, each individual patient who arrives in the ICU must have her own survival function calculated and pooled with her current patient cohort as follows:

- A patient $i = [1, ...n]$ who arrives at the ICU on day t_{d1} (where d is the date of arrival), has a Weibull survival function given by Equation 5.4↓(probability of survival):

$$(5.4)\ S_i(t_{d1}) = e^{-\gamma t_{d1}^{\lambda}}$$

- Equivalently, for any subsequent day $k = [1, ...m]$ of patient i, the Weibull survival function is:

$$(5.5)\ S_i(t_{dk}) = e^{-\gamma t_{dk}^{\lambda}}$$

Where the $S_i(t_{d1}) = 1$ and $S_i(t_{d\infty}) = 0$ and is monotonically decreasing.

- Therefore, the expected total number of current patient cohort at time t_{dk+1} that will remain in the ICU in any given day $T > t_{dk}$ is:

$$(5.6)\ E[ICU\ occupancy\ at\ day\ T \mid T > t_{dk}] = \sum_{i=1}^{n} P[S_i(T) \mid T > t_{dk}]$$

4.3.3 Model Validation

The Weibull AFT (day 1) model was fitted on a training set (first 18 months of the data set) and validated on the following 6 months. For the updated (day 3) model, we followed the same procedure, but included only patients that survived three complete days in the ICU.

Model accuracy was assessed by R^2, calculated as the squared correlation between predicted LOS (extracted from Equation 5.2↑) and observed LOS; MAE↓ (mean absolute error), NRMSE↓ (normalized root mean squared error), rSD (ratio of predicted and observed standard deviations), RSR (ratio of the RMSE between simulated and observed values to the standard deviation of the observations) and PBias (prediction bias), as implemented by [74].

The short-term capacity planning model was tested on the last 6 months of the data set, discarding the first and last 14 days to avoid underestimation of observed LOS. Model accuracy was assessed by ICU occupancy MAE, NRMSE, rSD, RSR,PBias and comparative means t-test.

4.4 Results

4.4.1 Weibull AFT model (day 1)

Figure 4.2↑ shows ICU stay probabilities ("Survival") by MODS, NEMS, patient category and ventilation. The Kaplan-Meier curves in Figure 4.2↑ show the higher probability of longer stays for high scoring MODS and NEMS patients, as well as medical, emergency, and ventilated younger patients.

The list of variables can be found in Table 4.4↑) while the resulting model can be found in Table 4.5↓. Converted Weibull parameters and variable coefficients can be found in Table 4.6↓, hazard rates in Table 4.7↓ and event time ratios in Table 4.8↓.

As seen in Table 4.7↓, the higher the Glasgow Coma Score, the higher the risk of a shorter ICU LOS. The same behaviour can be found for age: patients 80+ have a 144% higher risk of a shorter ICU LOS than their 18-39 year-old counterparts. Similarly, surgical patients are 129% more likely to have a shorter LOS than medical patients. Higher MODS patients are 64% (HR of 0.36) more likely to have a longer stay than Lower MODS patients. Likewise, higher NEMS patients are more likely to have a longer stay (HR of 0.93 - although not statistically significant at the $p = 0.05$ level), as are ventilated patients (HR 0.62) and Level 3 patients (those who originated from other ICU units - HR of 0.32).

Perhaps more intuitive is the interpretation of event time ratios (Table 4.8↓). For example, a 13+ MODS is linked to an increase in LOS of 200%, while advanced age (e.g., 80+) tends to reduce LOS by 33% (ETR of 0.67). Similarly, we find that Level 3 patients have a 240% longer LOS than emergency patients and that ventilated patients can expect to have an ICU LOS 66% longer than that of non-ventilated patients. Notice that the score 30+ in NEMS yields only a 7% increase in LOS.

Next, we compared the cumulative hazards functions for the non-parametric Cox proportional hazards model with our Weibull AFT model (see Figure 4.4↓-a) and an Exponential AFT model distribution (see Figure 4.4↓-b). The Cox proportional hazards model (dashed lines) fits closer to the observed data as no distribution function is assumed for the baseline hazard function. However, the proportionality of the hazards cannot be assumed ($\chi^2 = 79.3217$, $p < 0.000$); therefore, the Weibull AFT model provides the closest fit.

4.4.2 Weibull AFT model (day 3)

In a manner similar to that described in the preceding section, we fit a Weibull AFT model for patients that have remained in the ICU for three complete days. This model considered not only the variables collected at day 1 (e.g., MODS, NEMS), but also updated values collected at day 3 of the ICU stay, namely NEMS and its components. For example, a patient might be under mechanical ventilation upon arrival at the ICU, but after three days no longer needs mechanical ventilation, resulting in a change of the variable "IsVentilation", and, consequently, of the NEMS at day 3.

We performed an ANOVA test, step-wise variable selection and clinical validation for the Day 3 model as well. Variable impact in LOS variance is shown in Figure 4.3↓, and the resulting model can be seen in the ETR Table 4.5↓.

The updated Day 3 model adds the following variables to the previous model: haematologic and renal scores, sex, NEMS (day 3 score).

We will limit our discussion to a few significant ETR values for the Weibull AFT model (day 3). For example, males have an estimated reduction in LOS in of 22% compared to females, a relationship that was absent from the previous model. Higher MODS values maintain their increasing effect in ETR, ranging from 35% to 96% longer LOS than the baseline MODS. Notice though, that the day 1 model had the higher MODS scores increasing LOS from 70% to 200% (ETR from 1.7 to 3.0) compared to the baseline MODS. Age has a similar impact on LOS to that observed in the previous model (e.g., an ETR of 0.89 for the age group 40-80 and 0.65 for those aged 80 and above). Level 3 patients are expected to have a 90% longer LOS than emergency patients (ETR of 1.907) and patients that remain ventilated at day 3 can expect an ICU LOS 30% longer than that of non-ventilated patients (ETR of 1.309). Notice that both NEMS (day 1) and NEMS (day 3) are now part of the model. NEMS (day 1) scores have a diminishing effect on LOS (around 60%), compared to NEMS (day 3) scores (between 76 and 100% of the LOS of baseline NEMS).

4.4.3 Weibull AFT model prediction performance

Table 4.10↑ presents the prediction performance of both Day 1 and Day 3 models. Figure 4.6↓ shows the standardized residual plots of the fitted models. The Day 1 model, as fitted considering all the patients in our data set, represents measures and individual characteristics

collected upon arrival in the ICU. Predicted average LOS was 5.7 days compared to 5.04 days in the training data set. We can observe that there is a consistent loss in performance in R^2 (from 0.09 to 0.07), in MAE (4.68 days to 4.76 days), and in NRMSE (96% to 98.6%) from the training to test data sets. The Day 1 model tends to overestimate predictions from 13-19.6% and and only reaches 39-42% of the standard deviation of the observed data.

When using the fitted Day 1 model to predict the remaining LOS of the patients who had already reached LOS $\geq$ 3 days, we can see that R^2, MAE and NRMSE worsen, with NRMSE reaching 106% of observed values (compared to 96-98% of the complete data set). Additionally, this model greatly underestimates LOS. The predicted average remaining LOS is 6.68 days, compared to 10.27 days of the observed LOS, a bias of -34.9%.

Comparatively, the Day 3 model seems to be more accurate in terms of training R^2. Although MAE sees a similar increase (reaching 6.44-6.73 days), NRMSE ranging from 95.5% to 101.6% is better than the 106% found in the Day 1 model. Note that the model is positively biased at 13.1-14.8% (contrasting with the negative bias of -34.9% in the previous model), while reaching a slightly larger variance, as evidenced by the rSD values of 0.46 (training) and 0.43 (testing).

The residual plots show an improvement in prediction accuracy of the Day 3 model compared to the Day 1 model (fitted with LOS $\geq$ 3 days patients). The Day 3 model is notably less dispersed than the Day 1 model for both the complete data set as well as the LOS $\geq$ 3 days data set (see Figure 4.6↓).

4.4.4 Short-term capacity planning model

As seen in the previously described models, individual predictions tend to err approximately as much as the average observed LOS (NRMSE ranging from 95.5-106%) for most models. By pooling individual survival probabilities on a given day, one can focus on the short-term bed needs of the ICU as a whole. Therefore, our model reduces overall error while providing a reliable means of ICU capacity planning. As stated previously (Section 4.3.3↑), the short-term capacity planning model was tested on data from July 15^{th,} 2016 to December 17th, 2016. Table 4.11↓ provides a depiction of observed and expected LOS generated by our model. Observed and expected occupancy average \thickapprox19 beds, or roughly 76% utilization. Observed occupancy has a larger variance, ranging from 7 to 25 beds, while expected occupancy ranges

from \thickapprox12 to 24 beds. Table 4.12↓ summarizes the performance metrics. First and foremost, we can see that the aggregation of individual survival probabilities into overall ICU occupancy greatly reduces prediction error, with NRMSE of only 15.1% and R^2=0.51 (recall that individual prediction NRMSE approximates 100% error, and R^2\thickapprox0.10).

Table 4.13↓ provides an example of a typical day in the ICU (October 1st, 2016). The model required 40 patients to estimate the occupancy of October 1st, of whom only 14 were observed to be in the ICU. Curiously, one observed patient had a 0.1% survival probability to be in the ICU on that particular date.

Figure 4.7↓ shows density plots of observed and expected occupancy; both curves are right-skewed and the observed occupancy has a higher density at the extremes. The two-sample t-test shows no significant difference between the two distributions (p=value = 0.6995). Figure 4.8↓ shows a time series chart comparing observed occupancy and expected occupancy (+/- 1 standard deviation), confirming good prediction fit.

4.5 Discussion

In this paper we have developed individual ICU LOS and aggregate occupancy predictions based on a Weibull AFT model. First and foremost, the use of an AFT model for LOS predictions provides good fit and avoids the strong assumption of normality or errors (as suggested by [108, 109]) required by the multiple regression models developed earlier by [61, 78, 112], as well as the proportionality of hazards (Cox models) as in [108]. Overall, our models suggest that MODS and NEMS measures provide reasonable predictions in terms of ICU LOS, with R^2 values similar to those found in other models described in the literature by [108, 109]. Interestingly, patient diagnosis was not relevant to our AFT model, despite being the main strength of current APACHE models (see [42, 78, 84, 107, 108, 112]). This may be due to the simpler nature of our model's diagnosis, which contains only 6 categories: cardiovascular/cardiac/vascular, gastrointestinal, neurological, other, respiratory, trauma. High MODS, NEMS, Glasgow coma scores, and mechanical ventilation have increasing HR and, thus, increasing LOS (contradicting, in part [375, 376]), while age and surgical origin are linked to shorter stays (perhaps due to mortality as suggested by [377, 378, 87, 110, 98, 61, 107]).

Severity scores collected upon patient arrival in the ICU may provide masked information about the patient's true health status. We also note that individual LOS predictions

tend to improve when taken at a later time (i.e, after ICU arrival and with score updating, as suggested by [31, 9]). Although the choice of update time is unclear in the literature (e.g. [44, 387, 98]), NEMS and Ventilation updating at day 3 provided a more accurate model, both in terms of NRMSE and residuals (see Figure 4.6↓). We expect that MODS updating, if it had been available, may have provided an even better fit. Our model may also benefit from a larger data set since other cohorts can reach 100,000-200,000 patients, as in [61, 78, 93].

Although we did not address congestion and readmission (see [15, 26, 394, 65, 396, 30]), our short-term capacity planning model provided precise occupancy estimations. While individual LOS predictions err considerably (NRMSE of 100% or more), the survival probabilities of a daily cohort of patients creates a pooling effect, which reduces the overall NRMSE to 15%, representing, in our example, no more than 2 or 3 beds per day.

4.6 Conclusion

This research has successfully addressed several shortcomings in the ICU prediction model literature, including the implementation of MODS and NEMS as worthy replacements for APACHE and other metrics in the modeling of ICU LOS. The Weibull AFT model solves the problem of assumption violation (e.g., normality of errors, time-dependent variable effects, lack of hazard proportionality, residual trends) of traditional regression models as well as the lack of interpretability of the currently proposed machine learning models.

The Weibull AFT model achieves such objectives not only by providing comparably accurate LOS estimates, but also by generating helpful survival probabilities, meaningful hazard rates and intuitive event time ratios. Provided that the prediction is not needed upon the arrival of the patient in the ICU, updating patient information may help reduce the problem of individual LOS prediction error.

It is precisely this individual prediction error that has prevented practitioners from adopting ICU LOS predictions for capacity planning purposes. Therefore, with a simple mathematical formulation and novel use of the Weibull AFT model, we have demonstrated that pooling the future survival probabilities of an ICU patient cohort may provide a reliable estimation of short-term bed needs. Our model allows practitioners to finally implement ICU LOS predictions to start using patient information in order to better manage staff, downstream beds, and elective surgeries.

Table 4.1: ICU LOS Descriptive Statistics

Variable	N	Mean	St. Dev.	Min	Pctl(25)	Median	Pctl(75)	Max
Total ICU LOS	2,195	6.66	13.57	0.01	1.53	3.07	7.06	276.78
Clinical ICU LOS (before exclusions)	2,195	5.63	10.14	0.01	0.99	2.32	6.01	190.86
Clinical ICU LOS (after exclusions)	2,176	5.05	7.40	0.03	0.98	2.28	5.91	58.98

Table 4.2 MODS and NEMS By Outcome

MODS

	N	Mean MODS	SD	Min	Pctl(25)	Median	Pctl(75)	Max
Deceased	475	6.41	3.03	0	4	6	8	19
Discharged	1,701	4.43	2.80	0	2	4	6	17

Welch Two Sample t = -12.78, df = 715.1, p-value < 2.2e-16

NEMS

	N	Mean NEMS	SD	Min	Pctl(25)	Median	Pctl(75)	Max
Deceased	475	37.39	8.46	15	32	39	44	56
Discharged	1,701	31.61	8.75	0	27	32	39	56

Welch Two Sample t = -13.081, df = 780.1, p-value < 2.2e-16

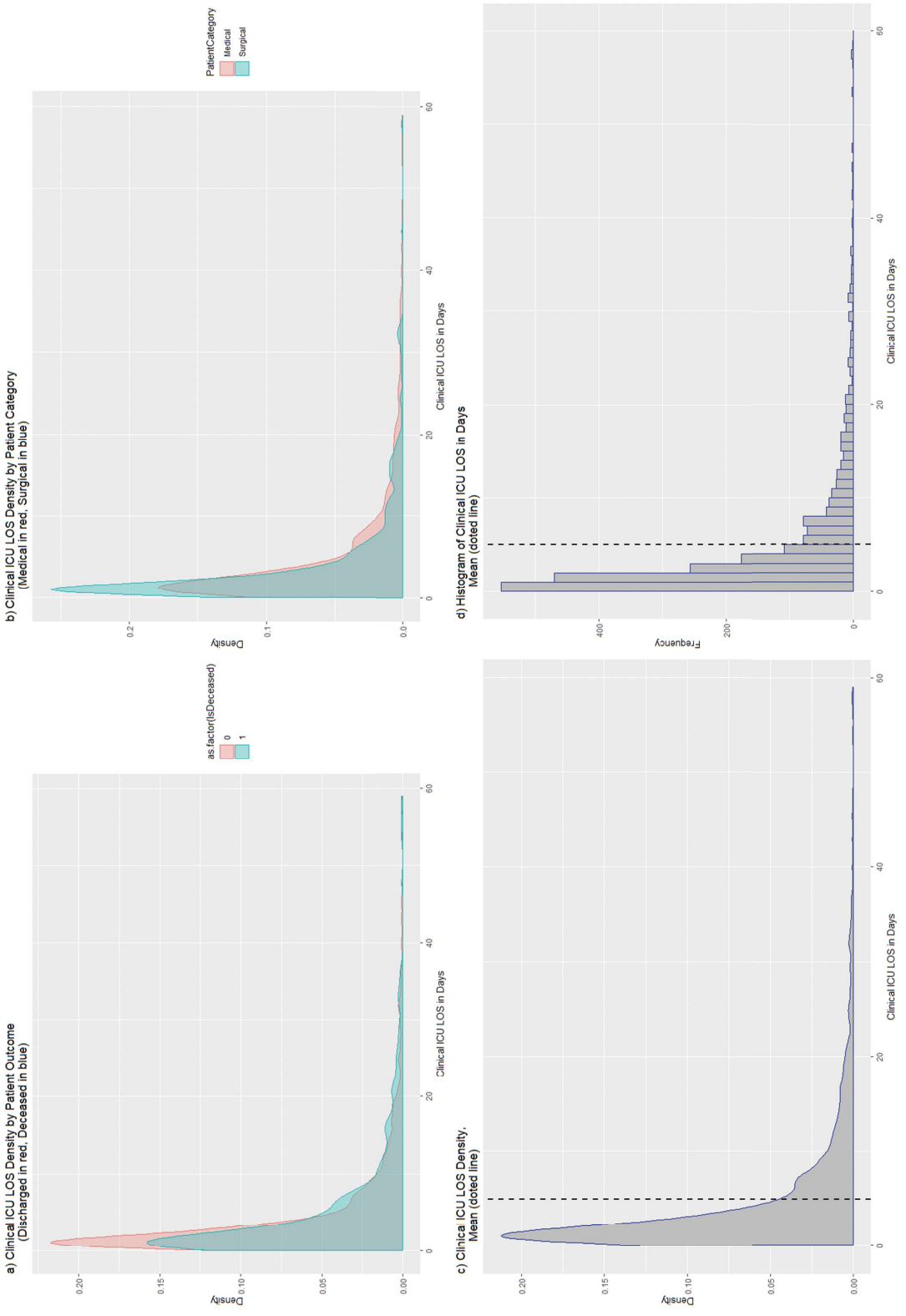


Figure 4.1: Clinical LOS By Patient By Outcome

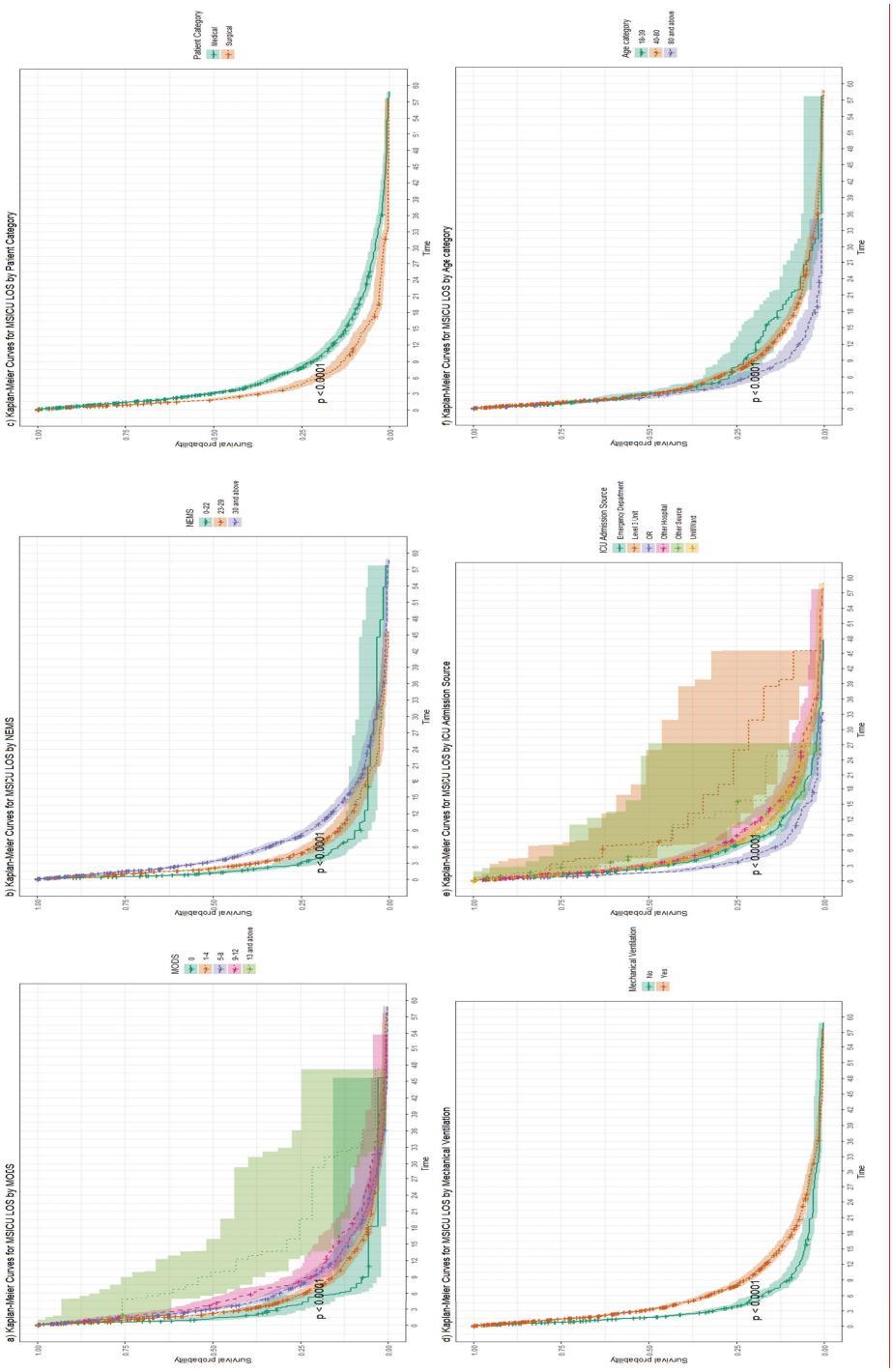


Figure 4.2: Kaplan-Meier Curves for Survival Models, Various Covariates

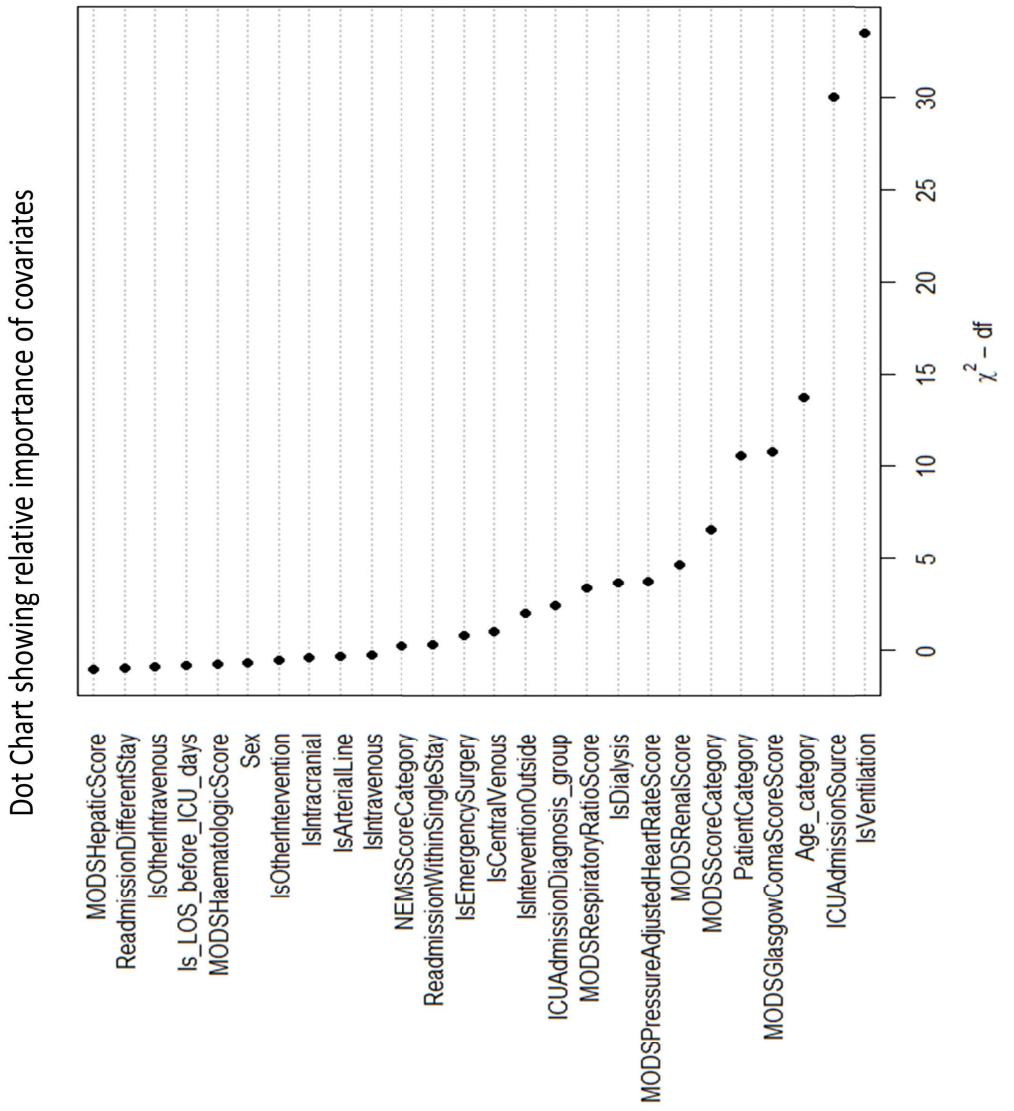


Figure 4.3: Variable Selection, Dot Chart - Weibull AFT Model (Day 1)

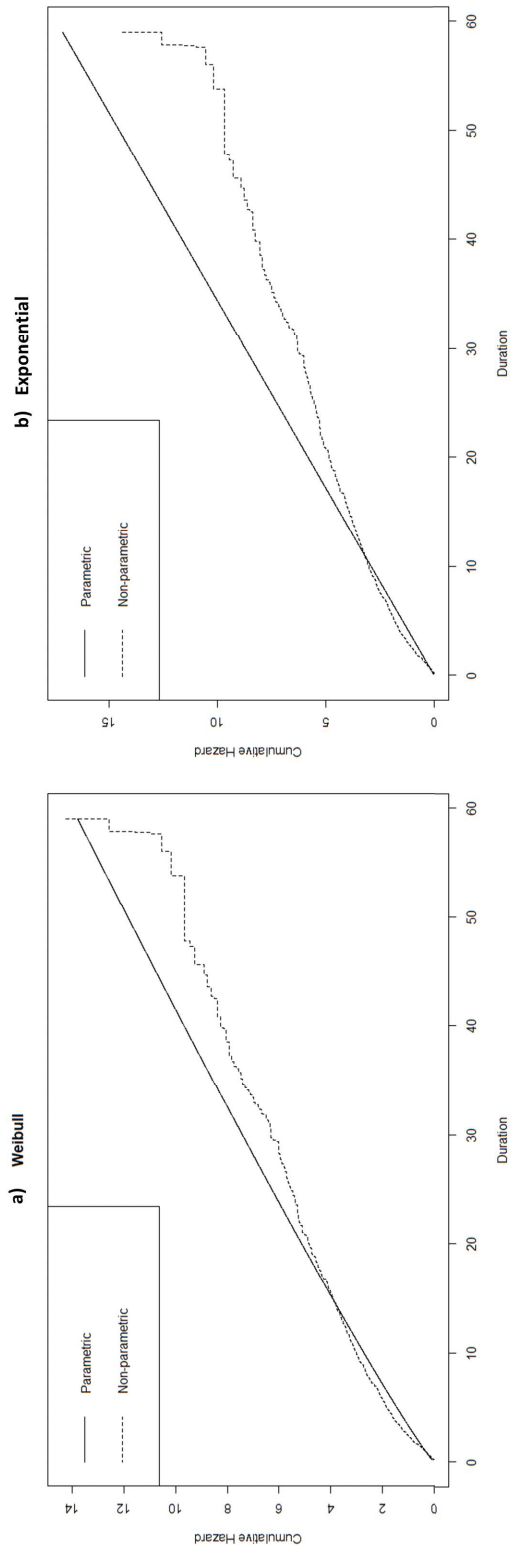


Figure 4.4: Cumulative Hazards Functions Comparison

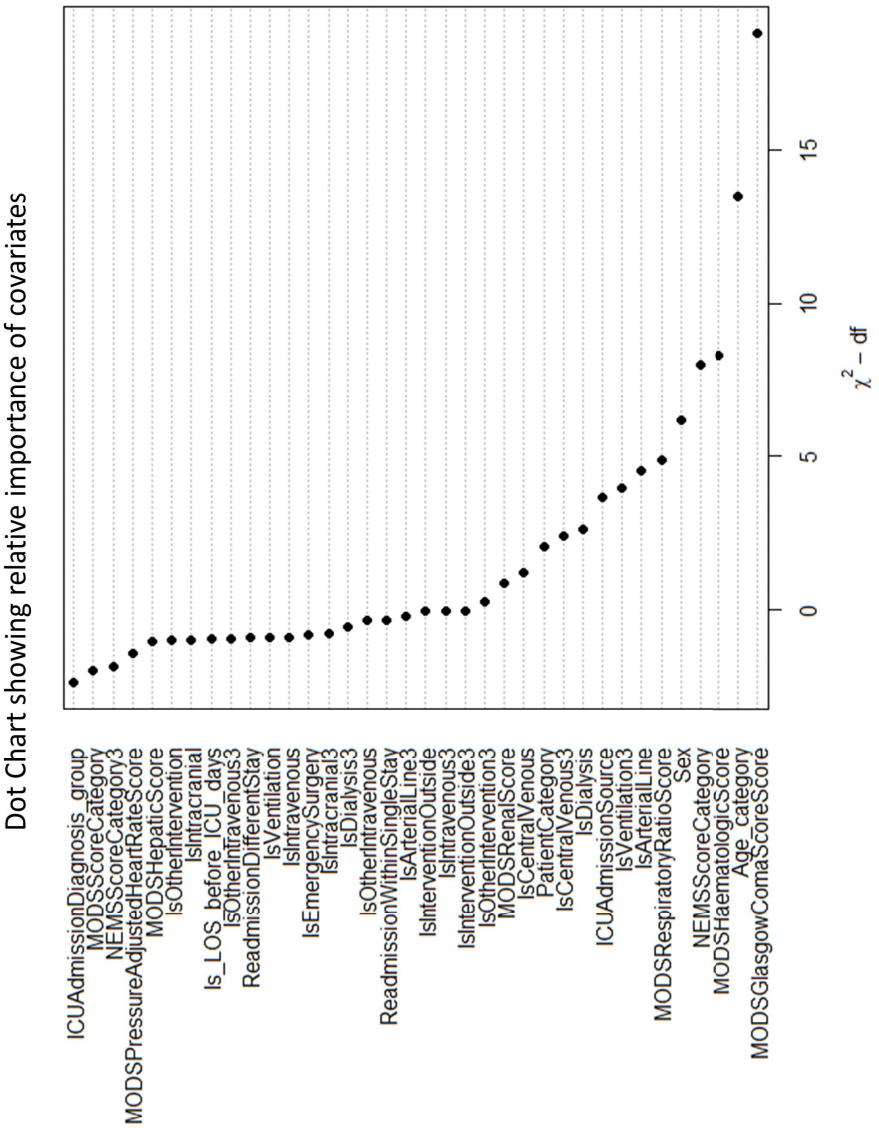


Figure 4.5: Variable Selection, Dot Chart - Weibull AFT Model (Da

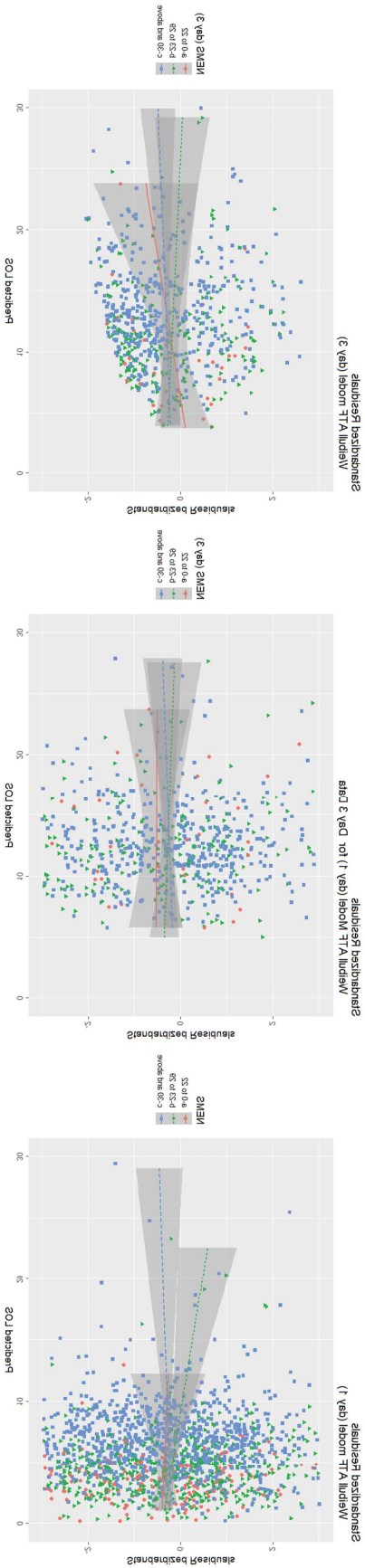


Figure 4.6: Residual Plots – Weibull AFT Models

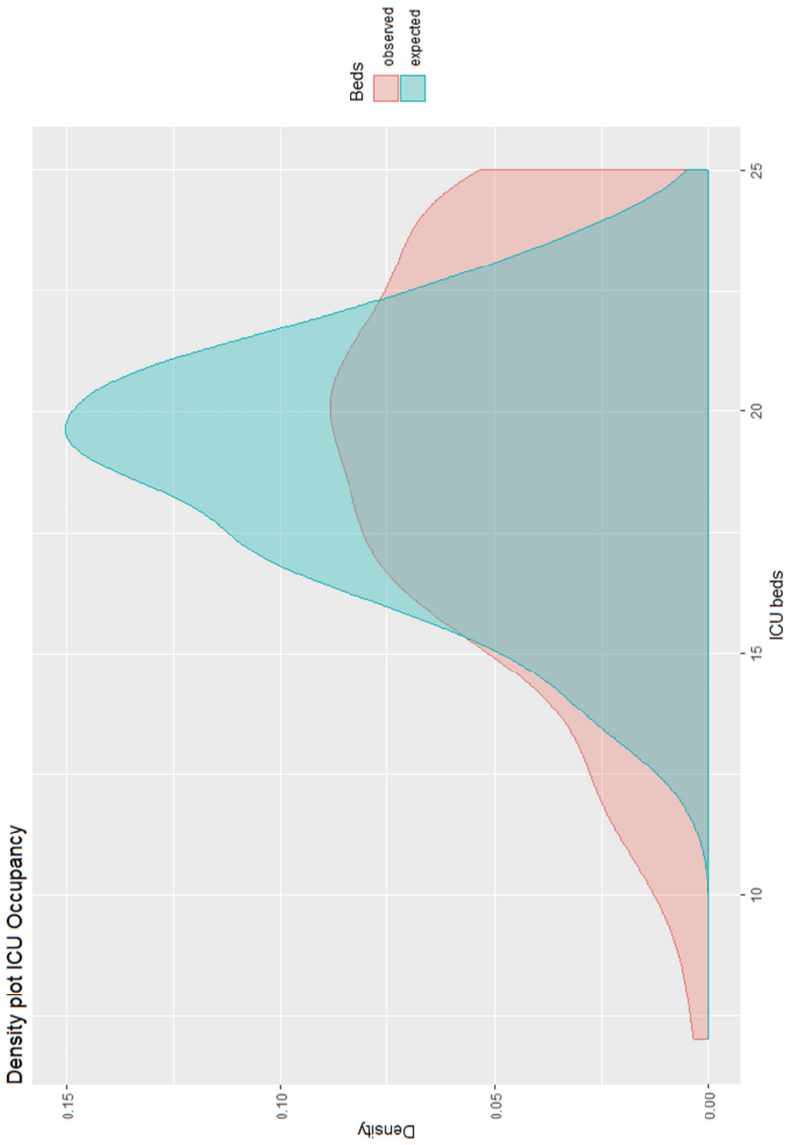


Figure 4.7: Comparative Density Plot - Short Term Capacity Planning Model

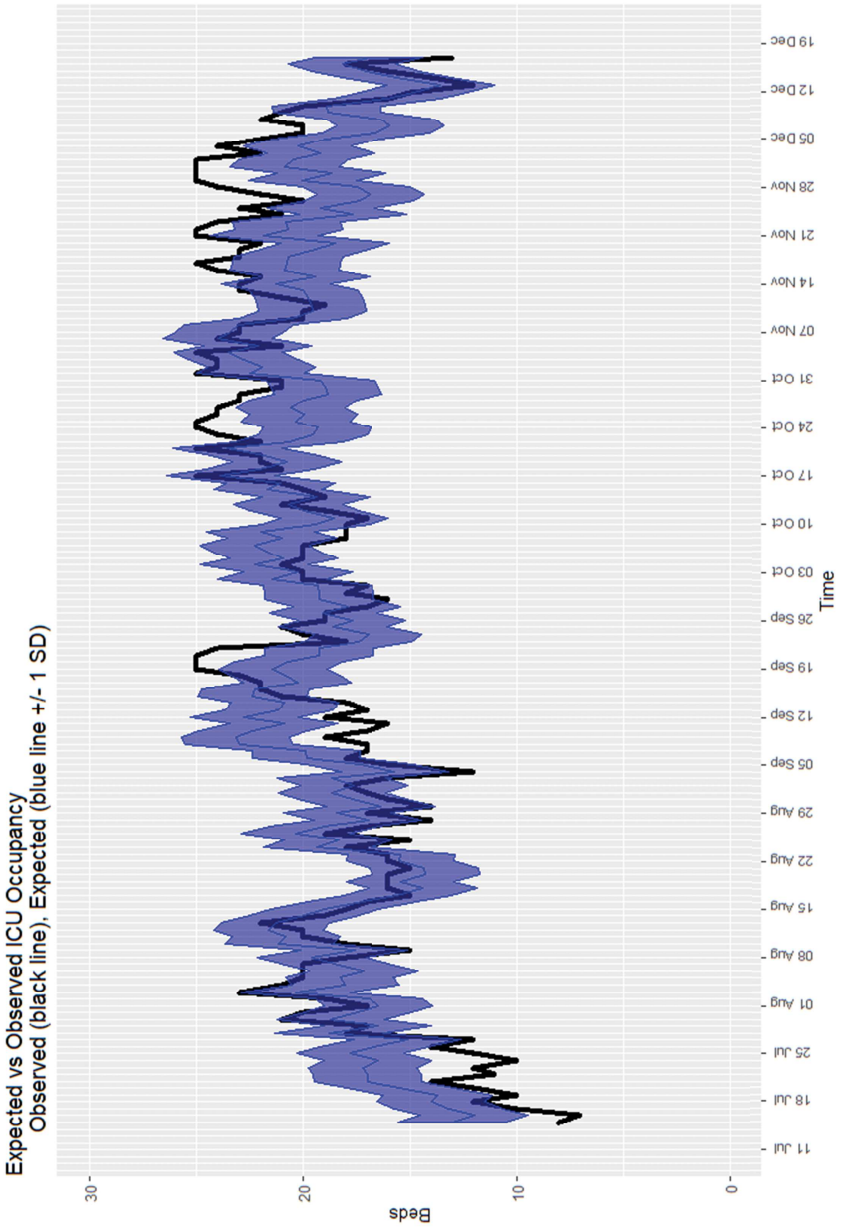


Figure 4.8: Time series – Short Term Capacity Planning Model

Table 4.3 MODS and NEMS by Patient Category

MODS

	N	Mean MODS	SD	Min	Pctl(25)	Median	Pctl(75)	Max
Medical	1,459	5.10	3.02	0	3	5	7	19
Surgical	717	4.39	2.78	0	2	4	6	14

Welch Two Sample t = 5.4517, df = 1532.7, p-value = 5.805e-08

NEMS

	N	Mean NEMS	SD	Min	Pctl(25)	Median	Pctl(75)	Max
Medical	1,459	33.25	9.13	0	27	32	39	56
Surgical	717	32.09	8.69	12	27	32	39	51

Welch Two Sample t = 2.8719, df = 1488.1, p-value = 0.004138

Table 4.4 List of Variables

Variable type	Subcategory	Variable	Nature (units)	Description
Outcomes	LOS	ICU_LOS_Clinical	continuous	clinical length of stay in days
Covariates	MODS	MODSGlasgowComaScoreCat	categorical	Glasgow Coma Scale score, (None, Minimal, Mild, Moderate, Severe)
		MODSHaematologicCat	categorical	Platelet count, (None, Minimal, Mild, Moderate, Severe)
		MODSHepaticCat	categorical	Serum bilirubin level, (None, Minimal, Mild, Moderate, Severe)
		MODSPressureAdjustedHeartRateCat	categorical	Pressure-adjusted heart rate, (None, Minimal, Mild, Moderate, Severe)
		MODSRenalCat	categorical	Serum creatinine level, (None, Minimal, Mild, Moderate, Severe)
		MODSRespiratoryRatioCat	categorical	PaO2/FIO2 ratio, (None, Minimal, Mild, Moderate, Severe)
		MODSScoreCategory	categorical	a-0 score, b-1 to 4, c-5 to 8, d-9 to 12, e-13 and above
	NEMS	IsArterialLine	binary	introduction of arterial lines
	(at day 1 and day 3)	IsbasicMonitoring	binary	monitoring of basic vital signs
		IsCentralVenous	binary	central venous medication
		IsDialysis	binary	any dialysis techniques
		IsDischarged	binary	discharged from ICU
		IsExtraCorporealMembrane	binary	extra-corporeal membrane life support
		IsInterventionOutside	binary	Specific interventions outside the ICU: such as surgical intervention or diagnostic procedure;
		IsIntraAorticBalloonPump	binary	Intra-aortic balloon pump
		IsIntracranial	binary	Intracranial pressure monitoring
		IsIntravenous	binary	intravenous medication
		IsOtherIntervention	binary	other interventions such as: endotracheal intubation, pacemaker, cardioversion, endoscopy, emergency operation in the past 24 h, gastric lavage; X-rays, echocardiography, electrocardiography, dressings
		IsOtherIntravenous	binary	other types of intravenous medication
		IsVentilation	binary	mechanical ventilation
		NEMSScoreCategory	categorical	a-0 to 22, b-23 to 29, c-30 and above
	Patient characteristics	Age_category	categorical	18-39, 40-80, 80 and above
		Sex	binary	male (0), female (1)

Admission characteristics	IsEmergencySurgery	binary	emergency surgery
	IsLOS_before	binary	hospital admission to other prior to ICU
	ReadmissionDifferentStay	binary	readmission to ICU from previous hospital admission into ICU
	ReadmissionWithinSingleStay	binary	readmission to ICU in the same hospital stay
	Campus	categorical	MSICU, CCTC (ICU's from different hospital sites)
	ICUAdmissionDiagnosis_group	categorical	Cardiovascular/Cardiac/Vascular, Gastrointestinal, Neurological, Other, Respiratory, Trauma
	ICUAdmissionSource	categorical	Emergency Dept., OR, Other Hospital, Stepdown Unit, Ward
	\strikeout off\uuline off\uwave offPatientCategory	\strikeout off\uuline off\uwave offcategorical	Medical, Surgical

Table 4.5 Weibull AFT Model (Day 1)

Variable	Value	Std.Error	z	p
(Intercept)	0.870	0.181	4.820	0.000
Glasgow Coma Score (1)	0.156	0.120	1.300	0.195
Glasgow Coma Score (2)	0.244	0.097	2.520	0.012
Glasgow Coma Score (3)	-0.029	0.108	-0.270	0.786
Glasgow Coma Score (4)	-0.162	0.087	-1.870	0.062
MODS (1 to 4)	0.535	0.148	3.620	0.000
MODS (5 to 8)	0.722	0.155	4.680	0.000
MODS (9 to 12)	0.757	0.179	4.220	0.000
MODS (13 and above)	1.103	0.275	4.020	0.000
NEMS (23 to 29)	-0.116	0.119	-0.980	0.330
NEMS (30 and above)	0.072	0.126	0.570	0.570
Is Intervention Outside	0.114	0.077	1.480	0.138
Is Ventilation	0.509	0.086	5.890	0.000
Age (40 to 80)	-0.149	0.102	f-1.460	0.144
Age (80 and above)	-0.400	0.121	f-3.290	0.001
ICU Admission Source (Level 3 Unit)	1.225	0.285	4.300	0.000
ICU Admission Source (OR)	-0.320	0.117	-2.730	0.006
ICU Admission Source (Other Hospital)	0.106	0.090	1.180	0.240
ICU Admission Source (Other Source)	0.387	0.310	1.250	0.213
ICU Admission Source (Unit/Ward)	0.209	0.084	2.490	0.013
Patient Category (Surgical)	-0.283	0.083	-3.400	0.001
Log(scale)	0.087	0.019	4.510	0.000
Scale=1.09				
Weibull distribution				
Loglik(model)=-3773.8	Loglik(intercept only)=-3895.4			
"Chisq=243.22 on 20 degrees of freedom"	p=2.7E-40			
Number of Newton-Raphson Iterations: 5	n=1648			

Table 4.6 Converted Weibull Parameters and Variable Coefficients - Weibull AFT Model (Day 1)

Variable	Estimate	SE
lambda	0.450	0.075
gamma	0.917	0.018
Glasgow Coma Score (1)	-0.143	0.110
Glasgow Coma Score (2)	-0.224	0.089
Glasgow Coma Score (3)	0.027	0.099
Glasgow Coma Score (4)	0.148	0.079
MODS (1 to 4)	-0.490	0.136
MODS (5 to 8)	-0.662	0.143
MODS (9 to 12)	-0.695	0.166
MODS (13 and above)	-1.012	0.253
NEMS (23 to 29)	0.106	0.109
NEMS (30 and above)	-0.066	0.115
Is Intervention Outside	-0.105	0.071
Is Ventilation	-0.467	0.080
Age (40 to 80)	0.136	0.093
Age (80 and above)	0.367	0.111
CU Admission Source (Level 3 Unit)	-1.124	0.262
ICU Admission Source (OR)	0.294	0.108
U Admission Source (Other Hospital)	-0.097	0.083
U Admission Source (Other Source)	-0.355	0.285
CU Admission Source (Unit/Ward)	-0.192	0.077
Patient Category (Surgical)	0.260	0.076

Table 4.7 Hazard Rates - Weibull AFT Model (Day 1)

Variable	Hazard Rate	Confidence Interval (95%)	
		Lower Bound	Upper Bound
Glasgow Coma Score (1)	0.867	0.699	1.076
Glasgow Coma Score (2)	0.799	0.672	0.952
Glasgow Coma Score (3)	1.027	0.846	1.247
Glasgow Coma Score (4)	1.160	0.993	1.355
MODS (1 to 4)	0.612	0.469	0.800
MODS (5 to 8)	0.516	0.390	0.682
MODS (9 to 12)	0.499	0.361	0.691
MODS (13 and above)	0.364	0.221	0.597
NEMS (23 to 29)	1.112	0.898	1.377
NEMS (30 and above)	0.936	0.747	1.174
Is Intervention Outside	0.901	0.784	1.034
Is Ventilation	0.627	0.537	0.733
Age (40 to 80)	1.146	0.955	1.376
Age (80 and above)	1.443	1.160	1.795
ICU Admission Source (Level 3 Unit)	0.325	0.194	0.544
ICU Admission Source (OR)	1.341	1.086	1.657
ICU Admission Source (Other Hospital)	0.907	0.772	1.067
ICU Admission Source (Other Source)	0.701	0.402	1.225
ICU Admission Source (Unit/Ward)	0.825	0.710	0.960
Patient Category (Surgical)	1.296	1.116	1.506

Table 4.8 Event Time Ratio - Weibull AFT Model (Day 1)

Variable	Event Time Ratio	Confidence Interval (95%)	
		Lower Bound	Upper Bound
Glasgow Coma Score (1)	1.168	0.923	1.478
Glasgow Coma Score (2)	1.276	1.056	1.543
Glasgow Coma Score (3)	0.971	0.786	1.200
Glasgow Coma Score (4)	0.851	0.718	1.008
MODS (1 to 4)	1.707	1.278	2.279
MODS (5 to 8)	2.059	1.521	2.787
MODS (9 to 12)	2.133	1.501	3.031
MODS (13 and above)	3.013	1.759	5.161
NEMS (23 to 29)	0.891	0.706	1.124
NEMS (30 and above)	1.074	0.839	1.375
Is Intervention Outside	1.121	0.964	1.303
Is Ventilation	1.663	1.404	1.970
Age (40 to 80)	0.862	0.706	1.052
Age (80 and above)	0.670	0.529	0.850
ICU Admission Source (Level 3 Unit)	3.405	1.948	5.952
ICU Admission Source (OR)	0.726	0.577	0.914
ICU Admission Source (Other Hospital)	1.112	0.932	1.327
ICU Admission Source (Other Source)	1.472	0.801	2.704
ICU Admission Source (Unit/Ward)	1.233	1.045	1.453
Patient Category (Surgical)	0.754	0.640	0.887

Table 4.9 Event Time Ratio - Weibull AFT Model (Day 3)

Variable	Event Time Ratio	Confidence Interval	
		Lower Bound	Upper Bound
Haematologic Score (1)	0.850	0.718	1.006
Haematologic Score (2)	0.568	0.450	0.716
Haematologic Score (3)	0.701	0.511	0.961
Haematologic Score (4)	0.480	0.310	0.745
Renal Score (1)	0.874	0.765	0.998
Renal Score (2)	0.843	0.688	1.032
Renal Score (3)	0.621	0.467	0.825
Renal Score (4)	0.862	0.590	1.262
Glasgow Coma Score (1)	1.223	0.975	1.535
Glasgow Coma Score (2)	1.207	1.009	1.444
Glasgow Coma Score (3)	0.822	0.682	0.991
Glasgow Coma Score (4)	0.741	0.629	0.872
MODS (1 to 4)	1.350	0.870	2.095
MODS (5 to 8)	1.623	1.034	2.548
MODS (9 to 12)	1.960	1.198	3.207
MODS (13 and above)	3.068	1.673	5.625
NEMS day 1 (23 to 29)	0.627	0.477	0.825
NEMS day 1 (30 and above)	0.621	0.476	0.810
NEMS day 3 (0 to22)	0.769	0.571	1.036
NEMS day 3 (23 to 29)	0.844	0.742	0.960
NEMS day 3 (30 and above)	1.000	1.000	1.000

Is Ventilation (day 3)	1.309	1.095	1.564
Age (40 to 80)	0.893	0.734	1.085
Age (80 and above)	0.658	0.521	0.832
Sex (Male)	0.873	0.782	0.974
ICU Admission Source (Level 3 Unit)	1.907	1.288	2.824
ICU Admission Source (OR)	0.944	0.755	1.179
ICU Admission Source (Other Hospital)	1.181	1.008	1.385
ICU Admission Source (Other Source)	1.282	0.742	2.214
ICU Admission Source (Unit/Ward)	1.165	1.004	1.352
Patient Category (Surgical)	0.872	0.751	1.014

Table 4.10 Clinical ICU LOS (Days) Performance Comparisons For Different Models

Model	R^2		MAE (days)		NRMSE (%)		PBIAS (%)		rSD	
	train	test	train	test	train	test	train	test	rain	test
Day 1 (full data set)	0.09	0.07	4.68	4.76	96.0%	98.6%	13.2%	19.6%	0.39	0.42
Day 1 (LOS ≥ 3 days)		0.05		6.93		106.0%		-34.9%		0.39
Day 3 (LOS ≥ 3 days)	0.12	0.04	6.44	6.73	95.5%	101.6%	13.1%	14.8%	0.46	0.43

Table 4.11 Short Term Capacity Planning Model - Descriptive Statistics

Descriptive Statistics (beds)	ICU Occupancy	
	Observed	Expected
Minimum	7.00	11.97
Average	19.28	19.08
Median	20	19.28
Maximum	25.00	24.02
SD	4.04	2.49
Variance	16.31	6.20

Welch Two Sample t-test t = 0.38644, df = 271.22, p-value = 0.6995

Table 4.12 Short Term Capacity Planning Model - Performance Prediction Statistics

Performance Measure	Value
MAE	2.37 beds
MAPE	12.6%
RMSE	2.88 beds
NRMSE	15.1%
PBias	-0.80%
rSD	0.62
rSR	0.71
R^2	0.51

Table 4.13 Example of Short Term Capacity Planning Model - Predicction Date: October 1st, 2016

Patient index	ICU Admission Date	Date of Stay	Survival Probability
1	2016-08-27	2016/10/01	0.1%
2	2016-09-17	2016/10/01	21%
3	2016-09-24	2016/10/01	7.0%
4	2016-09-24	2016/10/01	50.1%
5	2016-09-25	2016/10/01	18.8%
6	2016-09-27	2016/10/01	32.5%
7	2016-09-28	2016/10/01	44.4%
8	2016-09-29	2016/10/01	73.5%
9	2016-09-29	2016/10/01	59.6%
10	2016-09-29	2016/10/01	56.1%
11	2016-09-29	2016/10/01	69.3%
12	2016-09-30	2016/10/01	71.8%
13	2016-09-30	2016/10/01	86.5%
14	2016-09-30	2016/10/01	83.10%
+ 26 patients with survival probabilities ranging from 0.01% to 71.8%			9.59 (estimated days)
Observed occupancy (not included arrivals from 2016/10/01)	14		
Expected occupancy (not included arrivals from 2016/10/01)			16.33

Chapter 5

Conclusions and Future Directions

In this thesis, I have examined the ICU capacity management problem and explored ways in which to incorporate Canadian metrics such as the Multi-Organ Dysfunction Syndrome (MODS) score and the Nine Equivalents of Nursing Manpower Score (NEMS) into models designed to aid in long- and short term capacity planning. I have extended the ICU capacity management literature by incorporating patient-specific daily stochastic changes in NEMS to trigger patient step down from the ICU.

In Chapter 2, I developed a discrete-event simulation model to study the application of daily NEMS scoring processes as a means to trigger patient step down and examined the impact of step-down units on ICU patient flow. In Chapter 3, I examined the use of MODS and NEMS data as a means of predicting individual patients' LOS and outcome. I first assessed the suitability of MODS and NEMS scores and then compared the performance of multiple statistical and supervised machine learning methodologies in terms of LOS and outcome (e.g., death, long stay) prediction. In Chapter 4, I examined how MODS and NEMS can be used to predict a patient's ICU LOS at two different times: upon arrival and after three complete days in the ICU. I also introduced the use of the AFT model as means of estimating future aggregated ICU occupancy via pooled daily survival probabilities.

5.1 Managerial Insights

The findings in Chapter 2 provide valuable insights for hospital decision makers facing congested ICUs. I have shown that capturing daily NEMS scores and using them as a proxy for step-down readiness is a useful approach to estimating step-down time. NEMS is measured and collected daily on all ICU patients in Ontario, so managers can track patient progression via NEMS and respond with appropriate downstream resource allocation decisions. My findings suggest that, under certain conditions, a step-down unit can be beneficial to patient flow. First, its 2:1 patient/nurse ratio may reduce operational costs, while still providing the appropriate level of care. Second, and perhaps counter-intuitively, ICU beds can be converted into step-down beds, while maintaining overall capacity and patient throughput. Finally, provided that both ICU and

step-down beds are properly allocated, a step-down unit can work as a buffer that reduces ICU overstay and off-service, ultimately improving patient care.

Chapter 3 provided evidence that NEMS and MODS can be used to predict ICU LOS and patient outcome as reliably as more popular metrics, such as APACHE, SAPS and SOFA. Combining MODS (a severity score) with NEMS (a nursing workload score) is a useful practice, and given that both measures are already being collected for provincial reporting purposes, it is a logical next step to assess LOS and outcome estimates amongst patients arriving in the ICU. I evaluated several types of models. Given that no model was dominant in LOS prediction, this would indicate that decision makers may choose from a host of plausible models that do not require the limiting assumptions that are common in APACHE-based multiple linear regression models. For example, survival analysis models and CART provide both LOS predictions and variable explanatory power, while a random forest model may be more promising in terms of prediction but lacks explanatory value. Regarding patient outcomes, our results suggest that supervised machine learning models such as super learner and random forest seem to yield more accurate predictions but lack explanatory power

In Chapter 4 that the Accelerated Failure Time (AFT) parametric survival model provided hazard rates that are useful in explaining the instantaneous impact of each covariate in the patient's likelihood of leaving the ICU; it also provided event time ratios that indicate the impact of each covariate in total estimated LOS *vis-a-vis* the baseline case. By analyzing these relationships, decision makers will be better suited to plan patient care from the moment the patient arrives in the ICU. The survival probability pooling made possible by the AFT model can also be useful for short-term capacity planning. Decision makers can look at their current patient cohort and make a reliable estimate of short-term bed needs and staffing requirements as well as availability for elective surgery. This may provide hospitals with the flexibility needed to accommodate patient surges and calculate bed needs downstream.

5.2 Future Research

The studies in this thesis are based on models that rely on limiting assumptions. The long-range capacity planning discrete simulation model in Chapter 2 was developed to fit LHSC's needs; a natural extension of this would be to try to find generalizations and/or analytical models describing the relationship between the ICU and step-down units given the stochastic nature of

patient NEMS. Different methodologies can be explored, such as the Markov decision process, queuing networks with accumulating priority queues and game theory, to name a few.

LOS prediction models may be further explored through the inclusion of more detailed variables such as diagnostics, comorbidities and frailty scores. Larger data sets may also help in finding location-specific characteristics such as clinical practices that may play a role in ICU LOS, as well as in further refining the supervised machine-learning models.

The short-term capacity planning model may be extended in different directions. Another methodology that generates survival probabilities is the survival tree, which may be usefully compared to our current model. Future patient arrivals may also be simulated so the model can be extended to longer planning horizons. MODS and NEMS can also be updated on a daily basis, which may also serve to improve individual predictions .

Bibliography

[1] A. W. Woods, F. N. MacKirdy, B. M. Livingston, J. Norrie, J. C. Howie: "Evaluation of predicted and actual length of stay in 22 Scottish intensive care units using the APACHE III system", *Anaesthesia*, pp. 1058—1065, 2000.

[2] Adriana Pérez, Wenyaw Chan, Rodolfo J. Dennis: "Predicting the length of stay of patients admitted for intensive care using a first step analysis", *Health Services and Outcomes Research Methodology*, pp. 127—138, 2006.

[3] H. Akaike: "A new look at the statistical model identification", *IEEE Transactions on Automatic Control*, pp. 716-723, 1974.

[4] Jerome Allyn, Nicolas Allou, Pascal Augustin, Ivan Philip, Olivier Martinet, Myriem Belghiti, Sophie Provenchere, Philippe Montravers, Cyril Ferdynus: "A Comparison of a Machine Learning Model with EuroSCORE II in Predicting Mortality after Elective Cardiac Surgery: A Decision Curve Analysis", *PLOS ONE*, pp. e0169772, 2017.

[5] A. Almashrafi, L. Vanderbloemen: "Quantifying the effect of complications on patient flow, costs and surgical throughputs", *BMC Medical Informatics and Decision Making*, pp. 1-15, 2016.

[6] P. K. Andersen, R. D. Gill: "Cox Regression Model for Counting Processes: A Large Sample Study", *The Annals of Statistics*, pp. 1100-1120, 1982.

[7] David Anderson, Bruce Golden, Wolfgang Jank, Edward Wasil: "The impact of hospital utilization on patient readmission rate", *Health Care Management Science*, pp. 29-36, 2012. URL http://dx.doi.org/10.1007/s10729-011-9178-3.

[8] David Anderson, Carter Price, Bruce Golden, Wolfgang Jank, Edward Wasil: "Examining the discharge practices of surgeons at a large medical center", *Health Care Management Science*, pp. 338-347, 2011.

[9] Andrew A Kramer, Jack E Zimmerman: "A predictive model for the early identification of patients at risk for a prolonged intensive care unit length of stay", *BMC Medical Informatics and Decision Making*, 2010.

[10] Andrew A. Kramer, Jack E. Zimmerman: "The relationship between hospital and intensive care unit length of stay", *Critical Care Medicine*, pp. 1015—1022, 2011.

[11] Andrew A. Kramer, Thomas L. Higgins, Jack E. Zimmerman: "Intensive care unit readmissions in U.S. hospitals", *Critical Care Medicine*, pp. 3—10, 2012.

[12] Andrew J. Spieker: *endogenous: Classical Simultaneous Equation Models*. 2016. URL https://CRAN.R-project.org/package=endogenous. R package version 1.0.

[13] Andrew L. Rosenberg, Jack E. Zimmerman, Carlos Alzola, Elizabeth A. Draper, William A. Knaus: “Intensive care unit length of stay: Recent changes and future challenges”, *Critical Care Medicine*, pp. 3465—3473, 2000.

[14] Andy Liaw, Matthew Wiener: “Classification and Regression by randomForest”, *R News*, pp. 18-22, 2002. URL http://CRAN.R-project.org/doc/Rnews/.

[15] Mor Armony, Carri W. Chan, Bo Zhu: “Critical Care Capacity Management: Understanding the Role of a Step Down Unit”, *Production and Operations Management*, pp. 859-883, 2018.

[16] Aya Awad, Mohamed Bader-El-Den, James McNicholas: “Patient length of stay and mortality prediction: A survey”, *Health Services Management Research*, pp. 105-120, 2017.

[17] Bart Baesens: *Analytics in a big data world: the essential guide to data science and its applications*. John Wiley & Sons, Inc, 2014.

[18] Sean M. Bagshaw, Xioaming Wang, David A. Zygun, Dan Zuege, Peter Dodek, Allan Garland, Damon C. Scales, Luc Berthiaume, Peter Faris, Guanmin Chen, Dawn Opgenorth, Henry T. Stelfox: “Association between strained capacity and mortality among patients admitted to intensive care: A path-analysis modeling strategy”, *Journal of Critical Care*, pp. 81-87, 2018.

[19] Sebastien Bailly, Geert Meyfroidt, Jean-François Timsit: “What is new in ICU in 2050: big data and machine learning”, *Intensive Care Medicine*, pp. 1-4, 2017.

[20] Jerry Banks: *Discrete-event system simulation*. Pearson Prentice Hall, 2005.

[21] Thomas J. Best, Burhaneddin Sandikci, Donald D. Eisenstein, David O. Meltzer: “Managing Hospital Inpatient Bed Capacity Through Partitioning Care into Focused Wings”, *Manufacturing & Service Operations Management*, pp. 157-176, 2015.

[22] Otavio Bittencourt, Vedat Verter, Morty Yalovsky: “Hospital capacity management based on the queueing theory”, *International Journal of Productivity and Performance Management*, pp. 224-238, 2018.

[23] Bob Gray: *cmprsk: Subdistribution Analysis of Competing Risks*. 2014. URL https://CRAN.R-project.org/package=cmprsk. R package version 2.2-7.

[24] Theologos Bountourelis, M.Yasin Ulukus, JeffreyP. Kharoufeh, SpencerG. Nabors: *The Modeling, Analysis, and Management of Intensive Care Units* in *Handbook of Healthcare Operations Management* (Denton, Brian T., ed.). Springer New York, 2013.

[25] Leo Breiman: “Random Forests”, *Machine Learning*, pp. 5-32, 2001.

[26] Kurt M. Bretthauer, H. Sebastian Heese, Hubert Pun, Edwin Coe: “Blocking in Healthcare Operations: A New Heuristic and an Application”, *Production and Operations Management*, pp. 375—391, 2011. URL http://dx.doi.org/10.1111/j.1937-5956.2011.01230.x.

[27] Brian M. Livingston, Fiona N. MacKirdy, J. Cameron Howie, Ray Jones, John D. Norrie: "Assessment of the performance of five intensive care scoring models within a large Scottish database", *Critical Care Medicine*, pp. 1820—1827, 2000.

[28] Brian Ripley: *tree: Classification and Regression Trees*. 2016. URL https://CRAN.R-project.org/package=tree. R package version 1.0-37.

[29] CIHI: *Canadian Institute for Health Information*. http://yourhealthsystem.cihi.ca, 2015. URL http://yourhealthsystem.cihi.ca.

[30] Carolina Osorio, Michel Bierlaire: "An analytic finite capacity queueing network model capturing the propagation of congestion and blocking", *European Journal of Operational Research*, pp. 996 - 1007, 2009. URL http://www.sciencedirect.com/science/article/pii/S0377221708003871.

[31] Carri W. Chan, Vivek F. Farias, Nicholas Bambos, Gabriel J. Escobar: "Optimizing Intensive Care Unit Discharge Decisions with Patient Readmissions", *Operations Research*, pp. 1323—1341, 2012. URL http://www.jstor.org/stable/23323703.

[32] Mao-Te Chuang, Ya-han Hu, Chia-Lun Lo: "Predicting the prolonged length of stay of general surgery patients: a supervised learning approach", *International Transactions in Operational Research*, pp. 75-90, 2016.

[33] D. Reis Miranda, R. Moreno, G. Iapichino: "Nine equivalents of nursing manpower use score (NEMS)", *Intensive Care Medicine*, pp. 760—765, 1997.

[34] David Meyer, Evgenia Dimitriadou, Kurt Hornik, Andreas Weingessel, Friedrich Leisch: *e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien*. 2017. URL https://CRAN.R-project.org/package=e1071. R package version 1.6-8.

[35] David T. Wong, Sally L. Crofts, Manuel Gomez, Glenn P. McGuire, Robert J. Byrick: "Evaluation of predictive ability of APACHE II system and hospital outcome in Canadian intensive care unit patients", *Critical Care Medicine*, pp. 1177—1183, 1995.

[36] E.W. Steyerberg: *Study design for prediction models* in *Statistics for Biology and Health*. Springer Nature, 2008.

[37] Emmanuel Lesaffre, Arnošt Komárek, Dominique Declerck: "Statistical Methods in Medical Research", *An overview of methods for interval-censored data with an emphasis on applications in dentistry*, pp. 539-552, 2005.

[38] Eric Polley, Erin LeDell, Chris Kennedy, Mark van der Laan: *SuperLearner: Super Learner Prediction*. 2018. URL https://CRAN.R-project.org/package=SuperLearner. R package version 2.0-23.

[39] F. Rodrigues, G.S. Zaric, D.A. Stanford: "Discrete event simulation model for planning Level 2 'step-down' bed needs using NEMS", *Operations Research for Health Care*, pp. 42 - 54,

2018. URL http://www.sciencedirect.com/science/article/pii/S2211692316301564. Operational Research Applied to Health Services 2016.

[40] Jason P. Fine, Robert J. Gray: "A Proportional Hazards Model for the Subdistribution of a Competing Risk", *Journal of the American Statistical Association*, pp. 496-509, 1999.

[41] Helena Hvitfeldt Forsberg, Håkan Aronsson, Christina Keller, Staffan Lindblad: "Managing Health Care Decisions and Improvement Through Simulation Modeling", *Quality Management in Health Care*, pp. 15-29, 2011.

[42] Alberto Freitas, Tiago Silva-Costa, Fernando Lopes, Isabel Garcia-Lema, Armando Teixeira-Pinto, Pavel Brazdil, Altamiro Costa-Pereira: "Factors influencing hospital high length of stay outliers", *BMC health services research*, pp. 265-265, 2012.

[43] Adam Garland, David Ashton-Cleary, Ray Sinclair: "Inpatient illness severity surveys provide essential data for planning capacity and managing patient flow in the acute hospital setting", *Journal of the Intensive Care Society*, pp. 196-201, 2016.

[44] Gilles Clermont, Vladimir Kaplan, Rui Moreno, Jean-Louis Vincent, Walter T. Linde-Zwirble, Ben Van Hout, Derek C. Angus: "Dynamic microsimulation to model multiple outcomes in cohorts of critically ill patients", *Intensive Care Medicine*, pp. 2237—2244, 2004.

[45] Andrew S. Gordon, Adele H. Marshall, Mariangela Zenga: "Predicting elderly patient length of stay in hospital and community care using a series of conditional Coxian phase-type distributions, further conditioned on a survival tree", *Health Care Management Science*, pp. 269-280, 2018.

[46] Mia K. Grand, Hein Putter: "Regression models for expected length of stay", *Statistics in Medicine*, pp. 1178-1192, 2016.

[47] Robert J. Gray: "A Class of K-Sample Tests for Comparing the Cumulative Incidence of a Competing Risk", *The Annals of Statistics*, pp. 1141-1154, 1988.

[48] Linda V. Green, JoÃ£o Soares, James F. Giglio, Robert A. Green: "Using Queueing Theory to Increase the Effectiveness of Emergency Department Provider Staffing", *Academic Emergency Medicine*, pp. 61—68, 2006. URL http://dx.doi.org/10.1197/j.aem.2005.07.034.

[49] Linda Green: *Queueing Analysis in Healthcare* in *Patient Flow: Reducing Delay in Healthcare Delivery* (Hall, Randolph W., ed.). Springer US, 2006.

[50] Neil A. Halpern, Stephen M. Pastores: "Critical Care Medicine Beds, Use, Occupancy, and Costs in the United States: A Methodological Review", *Critical Care Medicine*, pp. 2452-2459, 2015.

[51] Hannah Wunsch, Anthony R. Brady, Kathy Rowan: "Impact of exclusion criteria on case mix, outcome, and length of stay for the severity of disease scoring methods in common use in critical care", *Journal of Critical Care*, pp. 67—74, 2004.

[52] Hans U. Rothen, Kay Stricker, Johanna Einfalt, Peter Bauer, Philip G. H. Metnitz, Rui P. Moreno, Jukka Takala: “Variability in outcome and resource use in intensive care units”, *Intensive Care Medicine*, pp. 1329—1336, 2007.

[53] Trevor Hastie, Robert Tibshirani, J. H. Friedman: *The elements of statistical learning: data mining, inference, and prediction*. Springer, 2009.

[54] James J. Heckman: “DUMMY ENDOGENOUS VARIABLES IN A SIMULTANEOUS EQUATION SYSTEM”, *Econometrica (pre-1986)*, pp. 931, 1978.

[55] Rein Houthooft, Joeri Ruyssinck, Joachim van der Herten, Sean Stijven, Ivo Couckuyt, Bram Gadeyne, Femke Ongenae, Kirsten Colpaert, Johan Decruyenaere, Tom Dhaene, Filip De Turck: “Predictive modelling of survival and length of stay in critically ill patients using sequential organ failure scores”, *Artificial Intelligence in Medicine*, pp. 191-207, 2015. URL https://journals-scholarsportal-info.proxy1.lib.uwo.ca/details/09333657/v63i0003/191_pmosalpusofs.xml.

[56] Peter J. Huber, Elvezio Ronchetti: *Robust statistics*. Wiley, 2009.

[57] Adrien Ickowicz, Ross Sparks: “Modelling hospital length of stay using convolutive mixtures distributions”, *Statistics in Medicine*, pp. 122-135, 2017.

[58] J.C. Ridge, S.K. Jones, M.S. Nielsen, A.K. Shahani: “Capacity planning for intensive care units”, *European Journal of Operational Research*, pp. 346 - 355, 1998. URL http://www.sciencedirect.com/science/article/pii/S0377221797002403.

[59] John C. Marshall, Deborah J. Cook, Nicolas V. Christou, Gordon R. Bernard, Charles L. Sprung, William J. Sibbald: “Multiple Organ Dysfunction Score”, *Critical Care Medicine*, pp. 1638—1652, 1995.

[60] John Fox, Sanford Weisberg: *An R Companion to Applied Regression*. Sage, 2011. URL http://socserv.socsci.mcmaster.ca/jfox/Books/Companion.

[61] John L. Moran, Peter Bristow, Patricia J. Solomon, Carol George, Graeme K. Hart: “Mortality and length-of-stay outcomes, 1993-2003, in the binational Australian and New Zealand intensive care adult patient database”, *Critical Care Medicine*, pp. 46—61, 2008.

[62] Karel G. M. Moons, Joris A. H. de Groot, Walter Bouwmeester, Yvonne Vergouwe, Susan Mallett, Douglas G. Altman, Johannes B. Reitsma, Gary S. Collins: “Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies: The CHARMS Checklist”, *PLoS Medicine*, pp. e1001744, 2014.

[63] J. Kasza, J. L. Moran, P. J. Solomon, ANZICS-Australian New Zealand Intensive Care Society Centre for Outcome, Resource Evaluation-CORE, for the ANZICS Centre for Outcome, Resource Evaluation (CORE) of the Australian, New Zealand Intensive Care Society (ANZICS): “Evaluating the performance of Australian and New Zealand intensive care units in 2009 and 2010”, *Statistics in Medicine*, pp. 3720-3736, 2013.

[64] Kim Seung-Chul, Horowitz Ira, Young Karl K, Buckley Thomas A: “Flexible bed allocation and performance in the intensive care unit”, *Journal of Operations Management*, pp. 427 - 443, 2000. URL http://www.sciencedirect.com/science/article/pii/S0272696300000279.

[65] Song-Hee Kim, Carri W. Chan, Marcelo Olivares, Gabriel Escobar: “ICU Admission Control: An Empirical Study of Capacity Allocation and Its Implication for Patient Outcomes”, *Management Science*, pp. 19-38, 2015.

[66] Andrew A. Kramer: “Are ICU Length of Stay Predictions Worthwhile?”, *Critical Care Medicine*, pp. 379-380, 2017.

[67] LHSC: *Financial Statement 2015*. online, 2015. URL http://www.lhsc.on.ca.

[68] Rocco J. LaFaro, Suryanarayana Pothula, Keshar P. Kubal, Mario E. Inchiosa, Venu M. Pothula, Stanley C. Yuan, David A. Maerz, Lucresia Montes, Stephen M. Oleszkiewicz, Albert Yusupov, Richard Perline, Jr Inchiosa: “Neural Network Prediction of ICU Length of Stay Following Cardiac Surgery Based on Pre-Incision VariablesMore re”, *PloS one*, pp. e0145395, 2015.

[69] Nehemiah T. Liu, Jose Salinas: “Machine Learning for Predicting Outcomes in Trauma”, *SHOCK*, pp. 504-510, 2017.

[70] Maartje E. Zonderland, Richard J. Boucherie, Michael W. Carter, David A. Stanford: “Modeling the effect of short stay units on patient admissions”, *Operations Research for Health Care*, pp. 21 - 27, 2015. URL http://www.sciencedirect.com/science/article/pii/S2211692315000168.

[71] Mahi Al Tehewy, Mostafa El Houssinie, Nahla Abou El Ezz, Mohamed Abdelkhalik, Samia El Damaty: “Developing severity adjusted quality measures for intensive care units”, *International Journal of Health Care Quality Assurance*, pp. 277—286, 2010.

[72] P. E. Marik, L. Hedman: “What is in a day? Determining intensive care unit length of stay”, *Critical care medicine*, pp. 2090, 2000.

[73] Marta L. Render, H Myra Kim, James Deddens, Siva Sivaganesin, Deborah E. Welsh, Karen Bickel, Ron Freyberg, Stephen Timmons, Joseph Johnston, Alfred F. Connors, Douglas Wagner, Timothy P. Hofer: “Variation in outcomes in Veterans Affairs intensive care units with a computerized severity measure”, *Critical Care Medicine*, pp. 930—939, 2005.

[74] Mauricio Zambrano-Bigiarini: *hydroGOF: Goodness-of-fit functions for comparison of simulated and observed hydrological time series*. 2017. URL http://hzambran.github.io/hydroGOF/. R package version 0.3-10.

[75] Michael J. Breslow, Omar Badawi: “Severity Scoring in the Critically III: Part 2: Maximizing Value From Outcome Prediction Scoring Systems”, *Chest*, pp. 518 - 527, 2012.

[76] Michael J. Breslow, Omar Badawi: “Severity Scoring in the Critically Ill”, *Chest*, pp. 245—252, 2012.

[77] Syed Mohiuddin, John Busby, Jelena Savović, Alison Richards, Kate Northstone, William Hollingworth, Jenny L Donovan, Christos Vasilakis: "Patient flow within UK emergency departments: a systematic review of the use of computer simulation modelling methods", *BMJ Open*, 2017. URL http://bmjopen.bmj.com/content/7/5/e015007.

[78] John L. Moran, Patricia J. Solomon, Australian New Zealand Intensive, ANZICS Centre for Outcome, Resource Evaluation (CORE) of the Australian, New Zealand Intensive Care Society (ANZICS), the ANZICS Centre for Outcome, Resource Evaluation (CORE) of the Australian, New Zealand Intensive Care Society (ANZICS): "A review of statistical estimators for risk-adjusted length of stay: Analysis of the Australian and new Zealand intensive care adult patient data-base, 2008-2009", *BMC Medical Research Methodology*, pp. 68-68, 2012.

[79] NCHS: *Health, United States, 2015*. https://www.cdc.gov/nchs/data/hus/hus15.pdf, 2015. URL https://www.cdc.gov/nchs/data/hus/hus15.pdf.

[80] Neil A. Halpern, 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*, pp. 65—71, 2010.

[81] Neil A. Halpern, 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*, pp. 65-71, 2010.

[82] Minna Niskanen, Matti Reinikainen, Ville Pettila: "Case-mix-adjusted length of stay and mortality in 23 Finnish ICUs", *Intensive Care Medicine*, pp. 1060-1067, 2009.

[83] John Norrie: "Mortality prediction in ICU: a methodological advance", *Lancet Respiratory Medicine, The*, pp. 5-6, 2015.

[84] Romain Pirracchio, Maya L. Petersen, Marco Carone, Matthieu R. Rigon, Sylvie Chevret, Mark J van der Laan: "Mortality prediction in intensive care units with the Super ICU Learner Algorithm (SICULA): a population-based study", *Lancet Respiratory Medicine, The*, pp. 42-52, 2015.

[85] R Core Team: *R: A Language and Environment for Statistical Computing*. 2017. URL https://www.R-project.org/.

[86] Chi-Lun Rau, Pei-FangJennifer Tsai, Sheau-FarnMax Liang, Jhih-Cian Tan, Hong-Cheng Syu, Yue-Ling Jheng, Ting-Syuan Ciou, Fu-Shan Jaw: "Using discrete-event simulation in strategic capacity planning for an outpatient physical therapy service", *Health Care Management Science*, pp. 352-365, 2013. URL http://dx.doi.org/10.1007/s10729-013-9234-2.

[87] Raymond Kao, Fran Priestap, Allan Donner: "To develop a regional ICU mortality prediction model during the first 24~h of ICU admission utilizing MODS and NEMS with six other independent variables from the Critical Care Information System (CCIS) Ontario, Canada", *Journal of Intensive Care*, 2016.

[88] Thomas R. Rohleder, Diane P. Bischak, Leland B. Baskin: “Modeling patient service centers with simulation and system dynamics”, *Health Care Management Science*, pp. 1—12, 2007. URL http://dx.doi.org/10.1007/s10729-006-9001-8.

[89] Joeri Ruyssinck, Joachim van der Herten, Rein Houthooft, Femke Ongenae, Ivo Couckuyt, Bram Gadeyne, Kirsten Colpaert, Johan Decruyenaere, Filip De Turck, Tom Dhaene: “Random Survival Forests for Predicting the Bed Occupancy in the Intensive Care Unit”, *Computational and Mathematical Methods in Medicine*, pp. 1-7, 2016.

[90] S C Brailsford, P. R. Harper, B. Patel, M. Pitt: “An analysis of the academic literature on simulation and modelling in health care”, *Journal of Simulation*, pp. 130-140, 2009.

[91] Syed Salleh, Praveen Thokala, Alan Brennan, Ruby Hughes, Andrew Booth: “Simulation Modelling in Healthcare: An Umbrella Review of Systematic Literature Reviews”, *PharmacoEconomics*, pp. 937—949, 2017. URL https://doi.org/10.1007/s40273-017-0523-3.

[92] Syed Salleh, Praveen Thokala, Alan Brennan, Ruby Hughes, Andrew Booth: “Simulation Modelling in Healthcare: An Umbrella Review of Systematic Literature Reviews”, *PharmacoEconomics*, pp. 937—949, 2017. URL https://doi.org/10.1007/s40273-017-0523-3.

[93] Sanjay Purushotham, Chuizheng Meng, Zhengping Che, Yan Liu: “Benchmarking deep learning models on large healthcare datasets”, *Journal of Biomedical Informatics*, pp. 112 - 134, 2018. URL http://www.sciencedirect.com/science/article/pii/S1532046418300716.

[94] Gideon Schwarz: “Estimating the Dimension of a Model”, *Ann. Statist.*, pp. 461—464, 1978. URL http://dx.doi.org/10.1214/aos/1176344136.

[95] Gideon Schwarz: “Estimating the Dimension of a Model”, *Ann. Statist.*, pp. 461—464, 1978. URL http://dx.doi.org/10.1214/aos/1176344136.

[96] L. Scrucca, A. Santucci, F. Aversa: “Competing risk analysis using R: an easy guide for clinicians”, *Bone Marrow Transplantation*, pp. 381-387, 2007.

[97] A. K. Shahani, S. A. Ridley, M. S. Nielsen: “Modelling patient flows as an aid to decision making for critical care capacities and organisation”, *Anaesthesia*, pp. 1074—1080, 2008. URL http://dx.doi.org/10.1111/j.1365-2044.2008.05577.x.

[98] Stanley Lemeshow, Daniel Teres, Jill Spitz Avrunin, Robert W. Cage: “Refining intensive care unit outcome prediction by using changing probabilities of mortality”, *Critical Care Medicine*, pp. 470—477, 1988.

[99] T. Gentimis, A. J. Alnaser, A. Durante, K. Cook, R. Steele: “Predicting Hospital Length of Stay Using Neural Networks on MIMIC III Data”, *2017 IEEE 15th Intl Conf on Dependable, Autonomic and Secure Computing, 15th Intl Conf on Pervasive Intelligence and Computing, 3rd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress(DASC/PiCom/DataCom/CyberSciTech)*, pp. 1194-1201, 2017.

[100] Terry M Therneau: *A Package for Survival Analysis in S*. 2015. URL https://CRAN.R-project.org/package=survival. version 2.38.

[101] Terry M. Therneau, Patricia M. Grambsch: *Modeling Survival Data: Extending the Cox Model*. Springer, 2000.

[102] Terry Therneau, Beth Atkinson, Brian Ripley: *rpart: Recursive Partitioning and Regression Trees*. 2017. URL https://CRAN.R-project.org/package=rpart. R package version 4.1-11.

[103] Thomas W. Noseworthy, Elsie Konopad, Allan Shustack, Richard Johnston, Michael P Eng Grace: "Cost accounting of adult intensive care", *Critical Care Medicine*, pp. 1168—1172, 1996.

[104] Lior Turgeman, Jerrold H. May, Roberta Sciulli: "Insights from a machine learning model for predicting the hospital Length of Stay (LOS) at the time of admission", *Expert Systems with Applications*, pp. 376-385, 2017.

[105] Peter T. Van Berkel, John T. Blake: "A comprehensive simulation for wait time reduction and capacity planning applied in general surgery", *Health Care Management Science*, pp. 373-385, 2007. URL http://dx.doi.org/10.1007/s10729-007-9035-6.

[106] Mark J Van der Laan, Eric C. Polley, Alan E. Hubbard: "Super Learner", *Statistical Applications in Genetics and Molecular Biology*, pp. 1309, 2007.

[107] Eduard E. Vasilevskis, Michael W. Kuzniewicz, Brian A. Cason, Rondall K. Lane, Mitzi L. Dean, Ted Clay, Deborah J. Rennie, Eric Vittinghoff, R. A. Dudley: "Mortality Probability Model III and Simplified Acute Physiology Score II", *Chest*, pp. 89, 2009.

[108] IWM Verburg, NF de Keizer, E. de Jonge, N. Peek: "Comparison of Regression Methods for Modeling Intensive Care Length of Stay", *PLOS ONE*, pp. e109684, 2014.

[109] Ilona W. M. Verburg, Alireza Atashi, Saeid Eslami, Rebecca Holman, Ameen Abu-Hanna, Everet de Jonge, Niels Peek, Nicolette F. de Keizer: "Which Models Can I Use to Predict Adult ICU Length of Stay? A Systematic Review", *Critical Care Medicine*, pp. e222-e231, 2017.

[110] W. A. Knaus, D. P. Wagner, J. E. Zimmerman, E. A. Draper: "Variations in Mortality and Length of Stay in Intensive Care Units", *Survey of Anesthesiology*, pp. 226227, 1994.

[111] W. N. Venables, B. D. Ripley: *Modern Applied Statistics with S*. Springer, 2002. URL http://www.stats.ox.ac.uk/pub/MASS4. ISBN 0-387-95457-0.

[112] Jack E. Zimmerman, Andrew A. Kramer, Douglas S. McNair, Fern M. Malila, Violet L. Shaffer: "Intensive care unit length of stay: Benchmarking based on Acute Physiology and Chronic Health Evaluation (APACHE) IV", *Critical Care Medicine*, pp. 2517-2529, 2006.

Chapter 6

Simulation Model design details

6.1 Overview

The Appendix contains a detailed explanation of the DES model (see screenshot in Figure 6.1↓) and its input parameters.

6.2 ER and OR arrivals

We modeled seasonality in emergency department (ED) and operating room (OR) arrivals. The OR performs both scheduled and emergency/unscheduled surgeries. These unscheduled surgeries are performed upon patients admitted to the ED or those in other wards who require a surgical procedure and transfer to the OR. After surgery they are transferred to other units in the hospital, including the MSICU. Unscheduled surgeries happen at any time of the day and any day of the week. Because unscheduled surgeries are performed on patients already in the hospital, we modeled the unscheduled surgeries as part of the inpatient flow matrix so they are not part of the external inpatient arrival pattern of the OR.

Scheduled surgeries are performed on patients not already in hospital and have a separate arrival pattern. These surgeries are typically scheduled between 5am and 11am on weekdays. There was no significant difference between the months or days of the week, but there was variation throughout the day (Table 6.1↓).

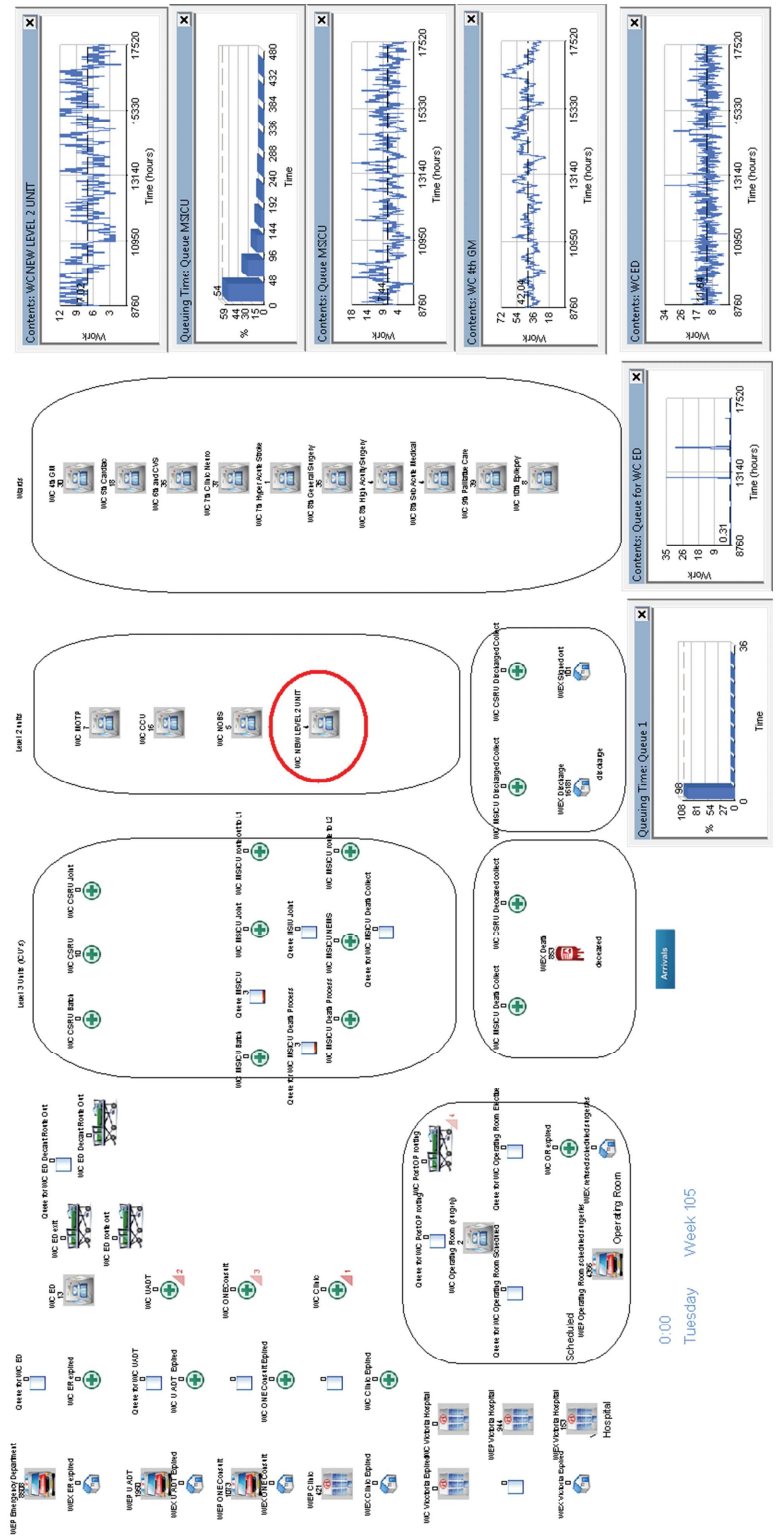


Figure 6.1 Screen capture from Simul8

units	Clinic	ED (emergency Department)	Operating Room	Victoria	10th - Epilepsy	4th - General Medicine	5th - Cardiac	6th - Acute Care	6th - Cardiac/Cardio vascular surgery	7th - Clinical Neurosciences	7th - Hyper Acute Stroke	8th - General Surgery, Plastic, Uro, Gyn	8th - High Acuity Surgery	8th - Sub Acute Medical	9th - Palliative Care	ED Decant	4th - MOTP (Transplant)	5th - CCU - Cardiac Care	7th - Neuro Obs	CSRU (cardiovascular recovery)	MSCU (medical surgery intensive care)	Discharged	Death	Signed Out	Grand Total
Clinic	2.1	3.2		0.5		13.3	15.2		14.3	7.4	0.8	12.5	0.3		8.5	0.3	11.7	6.1	0.3		0.3	2.9		0.3	100
ED (emergency Department)		4.6		0.7		17.1	6.7	0.5	7.6	10.9	1.3	10.9	0.1	0.1	9.0	14.6	1.2	4.6	1.1	0.3	2.5	5.6	0.5	0.1	100
ONEConsult	0.1	0.1				3.6	4.6	0.1	5.4	11.0	0.1	5.5	0.2		5.0	0.1	3.1	39.3	1.8	2.5	17.5				100
Operating Room	0.3				0.5	0.4	0.4		1.5	7.8		14.5	3.5		34.5		1.6	0.3	6.9	18.7	3.4	5.5	0.2		100
U-ADT	6.4	0.1	5.9		16.0	5.6	8.3	0.1	24.6	5.8	0.2	7.8	0.1	0.1	4.0	0.1	6.5	7.1	0.3	0.4	0.6				100
Victoria	0.4	2.6	0.2		0.2	0.4	3.4		11.7	6.7	0.2	1.4			2.8		0.6	19.2	1.4	1.2	8.1	34.2	5.1	0.2	100
10th - Epilepsy			1.4						0.3	0.6									0.6			96.5		0.6	100
4th - General Medicine														0.2	0.2						0.1	93.9	4.6	1.0	100
5th - Cardiac	0.3		7.0		0.3		1.4		3.2	0.2		0.4		0.5	0.9			3.5		3.2	0.6	76.5	1.6	0.2	100
6th - Acute Care							0.2		0.1					0.1	0.3	0.3						95.6	3.3	0.1	100
6th - Cardio/C/S		0.1	49.3		0.4	0.1	3.2	0.3		0.7		1.5	0.1	0.7	2.0	0.1	0.5	6.6		30.5	0.8				100
7th - Clinical Neurosciences			8.8		0.6	0.3	1.6		0.1		0.2	0.6		0.2	2.2	0.1	0.1	0.2	4.9	0.2	0.8	76.6	2.2	0.3	100
7th - Hyper Acute Stroke			1.4			0.7	0.7			68.8		0.7			0.7			0.7	2.2	1.4	1.4	20.6		0.7	100
8th - GS, Plastic, Uro, Gyn			8.5		0.7		1.1	0.2		0.4			1.7	0.7	1.8	0.1	0.8	0.1		0.2	1.3	80.6	0.9	0.7	100
8th - High Acuity Surgery			1.6		0.9							81.4			0.6		0.6	0.9		0.6	2.8	10.6			100
8th - Sub Acute Medical			2.3		0.5		4.1	0.5			0.5	1.4			3.2	0.5			0.5	0.5		82.3	3.2	0.5	100
9th - Palliative Care			7.2		0.4		3.6	0.1	0.1	0.6		1.2		1.0		0.1	0.2	0.2	0.1	0.1	0.4	80.5	3.6	0.2	100
ED Decant			1.3		0.2		45.7	3.6	4.3	5.0		9.8		0.5	7.5		0.3		0.1		0.2	15.6	0.9	0.3	100
4th - MOTP (Transplant)			10.7		0.2	0.2	13.1	0.3	0.7	1.7	0.9		6.1	0.3	0.2	2.6		0.5	0.2	1.6	5.9	52.9	1.7	0.2	100
5th - CCU - Cardiac Care	1.3		7.5		0.8		0.8	20.4		16.3	0.6		0.2	0.1		0.3	0.2		0.1	4.1	1.7	43.2	2.0	0.4	100
7th - Neuro Obs		0.1	5.8		0.1	0.4	0.3		0.1		71.5	0.7				0.4	0.1	0.1		0.8	3.8	15.7	0.1		100
CSRU			4.4		0.1		0.6		0.1	82.7	0.5		0.1	0.2		0.1	1.7	1.6	0.1		4.1	0.7	3.0		100
MSCU			8.6		0.9		16.5	1.7	0.1	0.7	7.0	0.3	6.9	2.9		2.3	0.1	3.5	6.8	2.0		9.4	22.0	0.2	100
Grand total	0.4	0.1	5.0		0.3	0.9	6.7	3.0	0.3	7.3	6.1	0.3	6.4	0.8	0.2	8.0	3.0	3.0	1.8	3.8	2.2	37.0	1.9	0.2	100

Figure 6.2 Inpatient flow matrix (origins in rows, destinations in columns, values in %)

Table 6.1 Average number of scheduled surgery arrivals per working day

Hour	Patients / hour
5 a.m.	2.8
6 a.m.	6.1
7 a.m.	1.3
8 a.m.	1.6
9 a.m.	2.3
10 a.m.	1.9
11 a.m.	0.9

ED arrivals varied according to the day of the week and the hour of the day. Our simulation of the ED is simplified by not capturing ED waiting room congestion. Instead, the process starts with the "ready for disposition" time, which is when the first assessment has been completed and the patient is ready to be admitted to one of the units of the hospital (Figure 6.3↓). In our data set there were 8,793 ED inpatients with average daily arrivals ranging from 21 on Sundays to 26 on Tuesdays. To avoid the possibility of not simulating any patients in a given hour, we divided the day into 4 parts: late night/early morning (12am to 6am), morning (6am to 12pm), afternoon (12pm to 6pm) and evening (6pm to 12am). ED inpatients are then simulated via the Poisson process, as indicated in Table 6.2↓.

6.3 UH structure and service time parameters

Ward capacities and service time parameters can be found in Table 6.4↓.

6.4 Detailed MSICU simulation

The simulation model of the MSICU starts with patient arrival from another unit (Figure 6.5↓). Upon arrival, the patient receives a "Level 3" NEMS that will represent her current status as a MSICU patient (Table 6.5↓). We then use a fork-join model and divide the patient into "physical" and "procedural" entities. The "physical" entity occupies a bed in the MSICU to ensure that MSICU capacity is not exceeded and that the appropriate queues form when capacity

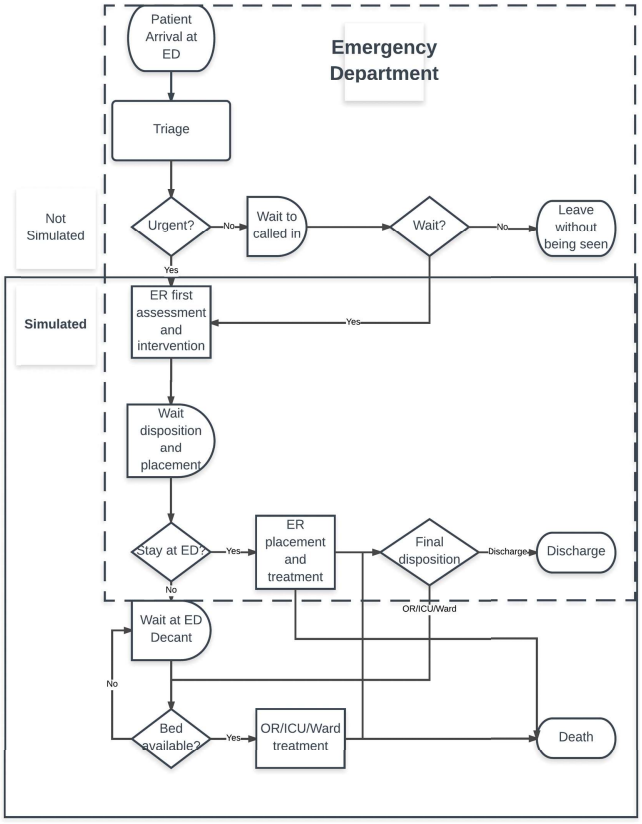


Figure 6.3 UH/LHSC ED flow

Table 6.2 ED inpatient arrivals per day of the week and time of the day

Average arrivals per 6 hour block

Time	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
00:00 to 06:00	5.9615	8.3461	6.9038	7.0385	7.0576	6.8845	6.077
06:00 to 12:00	3.25	3.0192	3.2308	3.1923	2.5384	3.6732	2.8462
12:00 to 18:00	5.7884	6.3269	6.2116	6.2307	6.1347	5.077	5.6344
18:00 to 00:00	9.2307	8.7307	7.4232	8.8462	8.6346	7.6538	7.1539
total	24.2306	26.4229	23.7694	25.3077	24.3653	23.2885	21.7115

Table 6.3 Entry points inter-arrival time distributions

Unit	Inter-arrival distribution type	Parameter (s), in hours
Clinic	Exponential	22.17
OneConsult	Exponential	8.2694
ADT	Exponential	4.454
Victoria	Gamma	α= 0.39314 ; θ= 24.142 (μ= 9.491 ; σ= 15.137)
ED	varies by day of the week and hour of the day (Table 6.2↑)	
OR	varies by hour of the day (Table 6.1↑)	

Table 6.4 Ward capacities and service time parameters

Units	Type	Number of Beds	Service time Distribution Type	Parameters (s)	Mean, standard deviation (hours)
Clinic	entry point		Weibull	1.402 ; 3.539	3.225 ; 2.331
OneConsult	entry point		Lognormal	0.032 ; 0.022	0.032 ; 0.022
ADT	entry point		Lognormal	0.040 ; 0.032	0.040 ; 0.032
Victoria Hospital	entry point		Gamma	0.430 ; 393.13	169.13 ; 257.86
Emergency Department (ED)	entry point / ED	40 stations	Exponential	11.694	11.694 ; 11.694
Operating Room (OR)	entry point / OR	16 rooms	Gamma	3.351 ; 2.483	8.325 ; 4.547
Emergency department Decant	ward	6	Lognormal	13.095 ; 11.069	13.095 ; 11.069
General Medicine (4th GM)	ward	72	Gamma	1.143 ; 107.47	122.91 ; 114.93
Cardiac Ward (5th Cardiac)	ward	20	Gamma	1.131 ; 110.49	125.02 ; 117.53
Acute Care	ward	12	Gamma	1.383 ; 85.375	118.09 ; 100.41
Cardiac/Cardiovascular Surgery (6th CVS)	ward	39	Gamma	1.374 ; 84.163	115.68 ; 98.67
Clinical Neurosciences (7th Neuro)	ward	44	Lognormal	152.97 ; 284.42	152.97 ; 284.42
Hyper Acute stroke (7th Stroke)	ward	5	Gamma	1.754 ; 31.506	55.28 ; 41.73
General Surgery, Plastic, Uro and Gyn (8th GS)	ward	41	Weibull	0.967 ; 110.88	112.39 ; 115.90
High Acuity Surgery (8th HAS)	ward	4	Weibull	1.281 ; 74.959	69.43 ; 54.59
Sub Acute Medical (8th SAM)	ward	15	Gamma	1.136 ; 372.69	423.41 ; 397.24
Palliative Care (9th PC)	ward	60	Lognormal	117.09 ; 178.76	117.09 ; 178.76
Epilepsy (10th EP)	ward	11	Gamma	2.744 ; 70.987	194.80 ; 117.59
Multi-Organ Transplant (MOTP)	Intermediary unit	12	Gamma	0.801 ; 190.26	152.52 ; 170.35
Coronary Care (CCU)	Intermediary unit	14	Weibull	1.331 ; 79.456	73.04 ; 55.38
Neurology Observation (NOBS)	Intermediary unit	6	Lognormal	62.806 ; 95.381	62.806 ; 95.381
Cardiovascular Surgery Recovery (CSRU)	Intensive Care	15	Lognormal	57.325 ; 71.966	57.33 ; 71.97
Medical Surgery Intensive Care (MSICU)	Intensive Care	25	*simulated via Death/NEMS stochastic routine		
Total Beds		401			

is reached. The "procedural" entity goes to the death/stay/step down process to model changes in health status and disposition from MSICU.

The first part of the death/stay/step down process is a daily routine that culminates in either death or survival. From our empirical data we built a logarithmic regression to estimate the probability of death as a function of time in the MSICU (Figure 6.4↓). We observed that no deaths occurred after 45 days, so we truncated the function at that point. If the patient dies then the two entities are joined and the patient exits both the MSICU and the simulation. Thus, MSICU LOS is a consequence of the patient's health progression over time, as opposed to an exogenously generated parameter. If the patient survives, then the "procedural" entity enters a NEMS scoring routine to sample a new NEMS.

The score either stays as at "Level 3", or changes to "Level 2" or "Level 1". In case of a "Level 3" NEMS, the procedural entity returns to the death process to repeat the survival and NEMS routine, with updated survival probability based on LOS (Figure 6.4↑). In case of a Level 2 score, in the baseline scenario, the patient still stays at the MSICU since there are no L2 beds available. In the other scenarios, a "Level 2" NEMS will trigger the procedural entity to be joined with its physical entity, exit the MSICU and move to a step-down unit. In the case of a Level 1 NEMS, in both scenarios the entities join and the patient is transferred to a ward.

Note that this captures the fact that a patient's health fluctuates over time and may improve or deteriorate. This model also allows for overstay patients to have their health change due to congestion downstream and captures sudden deaths in ther MSICU with a more detailed distribution than the one used elsewhere in the hospital, reflecting the high risk to the patient.

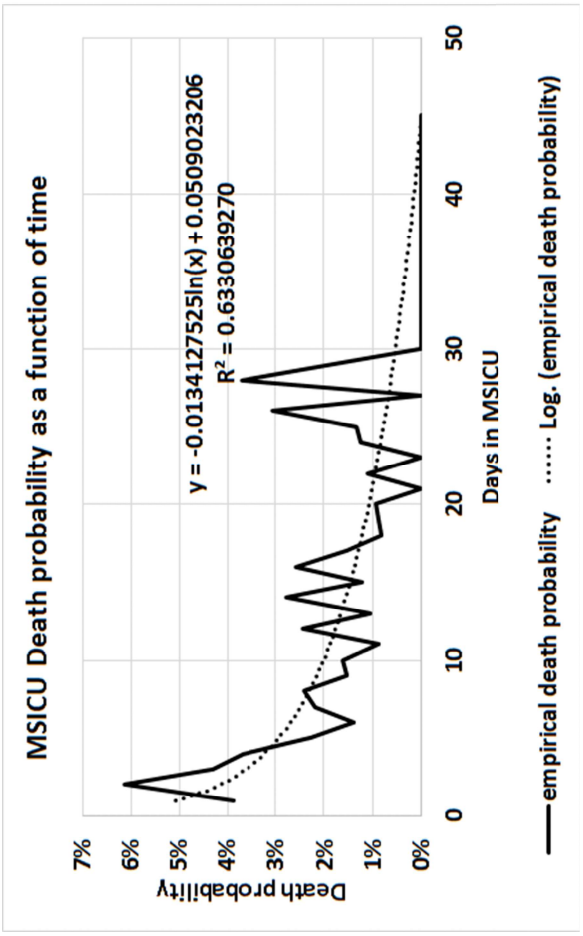


Figure 6.4: MSICU Death Probability as a function of time

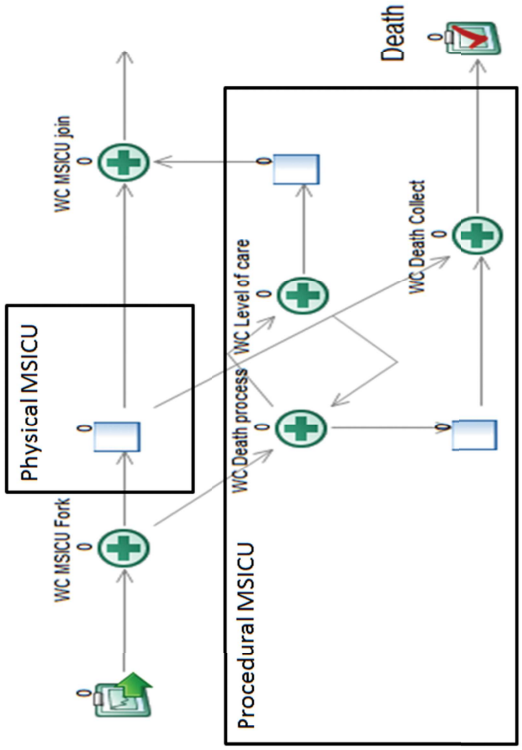


Figure 6.5: MSICU Death probability as a function of time (*WC stands for Work Centre)

Table 6.5 NEMS probability

	NEMS Probability
Level 1	7%
Level 2	24%
Level 3	69%
Total	100%

6.5 Capital expenditures estimates

Hospital stay cost data was retrieved from the Canadian Institute for Health Information ([29]). Operational cost and capital expenditures were obtained via consultation with LHSC Decision Support Staff and publicly available financial statements ([67]). Capital expenditures were linearly extrapolated from estimates of 8 and 15 beds ($3 million and $5 million respectively) and linearly depreciated over 10 years per Canadian accounting practice (Table 6.6↓).

Table 6.6 Level 2 unit capital expenditure estimates

Number of beds	Yearly capital expenditure	Expenditure/bed
2	$128,571	$64,285.71
4	$185,714	$46,428.57
6	$242,857	$40,476.19
8	$300,000	$37,500.00
10	$357,143	$35,714.29
12	$414,286	$34,523.81
14	$471,429	$33,673.47
15	$500,000	$33,333.33
16	$528,571	$33,035.71
18	$585,714	$32,539.68
20	$642,857	$32,142.86
22	$700,000	$31,818.18
24	$757,143	$31,547.62
26	$814,286	$31,318.68

6.6 Model validation

In the one year period of the data set, there were in total N = 17,380 inpatient arrivals, while our simulation averages 17,350, well within the 95% confidence interval (Table 6.7↓).

Table 6.7 Inpatient arrival validation

Simulation Object	Simulation Results -95%	average	95%	Observed data	Error
Emergency Department	8,760.48	8,794.50	8,828.52	8,793	0.02%
ADT	1,940.53	1,955.13	1,969.73	1,963	-0.40%
OneConsult	1,047.00	1,054.83	1,062.66	1,058	-0.30%
Clinic	266.05	271.37	276.68	275	-1.32%
Victoria Hospital	920.56	935.53	950.50	927	0.92%
Operating Room scheduled surgeries	4,308.84	4,338.93	4,369.03	4,364	
Total	17,243.47	17,350.30	17,457.13	17,380	-0.17%

Chapter 7

ICU LOS and outcome model parameters

Table 7.1: Variance inflation factor table

LOS	GVIF	Df	GVIF^(1/(2*Df))
MODSHaematologicCat	1.229363	3	1.035015
MODSHepaticCat	1.241252	3	1.036677
MODSRenalCat	1.135293	4	1.015988
MODSPressureAdjustedHeartRateCat	1.112486	3	1.017925
MODSGlasgowComaScoreCat	1.094506	4	1.011352
MODSRespiratoryRatioCat	1.650848	4	1.064666
MODSScoreCategory	1.876844	4	1.081879
NEMSScoreCategory	3.85211	2	1.400957
IsArterialLine	1.513073	1	1.23007
IsIntravenous	1.917876	1	1.384874
IsOtherIntravenous	1.046339	1	1.022907
IsIntracranial	1.08734	1	1.042756
IsDialysis	1.133065	1	1.064455
IsOtherIntervention	1.549994	1	1.244987
IsInterventionOutside	1.205076	1	1.09776
IsCentralVenous	1.449068	1	1.203772
IsVentilation	1.978102	1	1.40645
Age_category	1.222517	2	1.051511
Gender	1.043497	1	1.021517
Campus	1.266087	1	1.125205
ICUAdmissionSource	5.555884	4	1.239065
ICUAdmissionDiagnosis_group	2.244249	5	1.084194
PatientCategory	2.030271	1	1.424876
IsLOS_before	2.070508	1	1.438926
IsEmergencySurgery	1.590521	1	1.261159
ReadmissionWithinSingleStay	1.198276	1	1.094658
ReadmissionDifferentStay	1.050669	1	1.025021

Table 7.2 Model parameters ICU LOS prediction

Model	package	R function	parameters
CART	Tree	tree(y~x, ...)	control = tree.control(5000, mincut = 5, minsize = 10, mindev = 0.002)
Cox-PH	survival	survreg(coxph(surv(y~x)))	
CR-PH	survival, cmprisk, CumIncidence.R	CumIncidence(), survfit(Surv(y,comprisk)~x)	comprisk = Status(deceased, discharged)
GLM	MASS	glm(y~x)	family = Gaussian (link = "log")
GLM	MASS	glm(y~x)	family = Gamma (link = "log")
NN	NN	NN(y~x, ...)	size = 10, maxit = 100, linout=T
RF	randomForest	randomForest(y~x, ...)	ntree = 500, nvar = 10, importance = TRUE
RLM	MASS	rlm(y~x)	family = Gaussian (link = "log")
TEM	endogenous	hybrid(y~x , probit = z~x)	treatment effect = probit
two-stage GLM	MASS	glm(y~x, binomial)*glm(y_deceassed~x) + (1-glm(y~x, binomial))*glm(y_deceassed~x)	
two-stage Cox	MASS, survival	glm(y~x, binomial) * survreg(coxph(surv(y_deceased~x))) + (1-glm(y_survivor~x, binomial)) * survreg(coxph(surv(y_survivor~x)))	
SVM	e1071	svm(y~x, ...)	kernel = "radial", cost = 10, scale = FALSE
SL	Superlearner	SuperLearner(y,x,...)	SL.library = c("SL.glm","SL.ksvm","SL.rpart","SL.randomForest"), family=gaussian(), verbose = TRUE)

Table 7.3 Model parameters outcome prediction

Model	package	R function	parameters
CART	rpart	rpart(y~x, ...)	control = tree.control(5000, mincut = 5, minsize = 10, mindev = 0.005)
Logistic regression	MASS	glm(y~x, binomial)	family = binomial
Multinomial regression	NN	multinom(y~x, ...)	
NN	NN	NN(y~x, ...)	size = 10, maxit = 100, linout=T
RF	randomForest	randomForest(y~x, ...)	ntree = 500, nvar = 10, importance = TRUE
SVM	e1071	svm(y~x, ...)	kernel = "radial", cost = 10, scale = FALSE
SL	Superlearner	SuperLearner(y,x,..)	SL.library = c("SL.glm","SL.ksvm","SL.rpart","SL.randomForest"), family=gaussian(), verbose = TRUE)

Table 7.4 MODS Components adapted from Marshal et al 1995 [59]

Organ System	Indicator of Dysfunction	Degree of Dysfunction				
		None	Minimal	Mild	Moderate	Severe
Respiratory	PaO2/FIO2 ratio	> 300	226–300	151–225	76–150	≤ 75
Renal	Serum creatinine level	≤ 100 µmol/L	101–200 µmol/L	201–350 µmol/L	351–500 µmol/L	> 500 µmol/L
Hepatic	Serum bilirubin level	≤ 20 µmol/L	21–60 µmol/L	61–120 µmol/L	121–240 µmol/L	> 240 µmol/L
Cardiovascular	Pressure-adjusted HR	< 10.0	10.1–15.0	15.1–20.0	20.1–30.0	> 30.0
Hematologic	Platelet count	> 120,000/mm3	81,000–120,000/mm3	51,000–80,000/mm3	21,000–50,000/mm3	ff≤ 20,000/mm3
Neurologic	Glasgow Coma Scale score	15	13–14	10–12	7–9	≤ 6

Table 7.5 NEMS components (adapted from Miranda et al 1997 [33])

Item	Points
1. Basic monitoring: hourly vital signs, regular record and calculation of fluid balance	9
2. Intravenous medication: bolus or continuously, not including vasoactive drugs	6
3. Mechanical ventilatory support: any form of mechanical/assisted ventilation, with or without PEEP (e. g., continuous positive airway pressure), with or without muscle relaxants	12
4. Supplementary ventilatory care: breathing spontaneously through endotracheal tube; supplementary oxygen any method, except if (3) applies	3
5. Single vasoactive medication: any vasoactive drug	7
6. Multiple vasoactive medication: more than one vasoactive drug, regardless of type and dose	12
7. Dialysis techniques: all	6
8. Specific interventions in the ICU: such as endotracheal intubation, introduction of pacemaker, cardioversion, endoscopy, emergency operation in the past 24 h, gastric lavage; routine interventions such as X-rays, echocardiography, electrocardiography, dressings, introduction of venous or arterial lines, are not included	5
9. Specific interventions outside the ICU: such as surgical intervention or diagnostic procedure; the intervention/procedure is related to the severity of illness of the patient and makes an extra demand upon manpower efforts in the ICU	6
Total	56

Table 7.6 Performance Measures for LOS prediction models

Measure	Name	Explanation
RSR	Ratio of the RMSE	Ratio of the RMSE between simulated and observed values to the standard deviation of the observations
rSD	Ratio of Standard Deviations	Ratio of standard deviations between sim and obs
PBIAS	Percent Bias	Measures the average tendency of the simulated values to be larger or smaller than their observed ones. The optimal value of PBIAS is 0.0, with low-magnitude values indicating accurate model simulation. Positive values indicate overestimation bias, whereas negative values indicate model underestimation bias

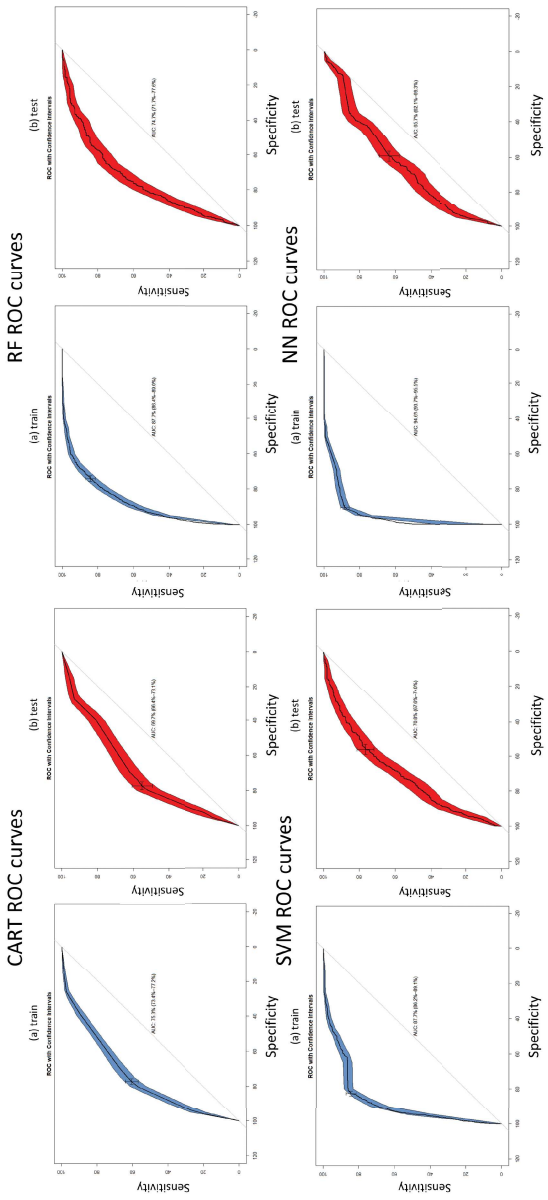


Figure 7.1 Mortality prediction models ROC curves

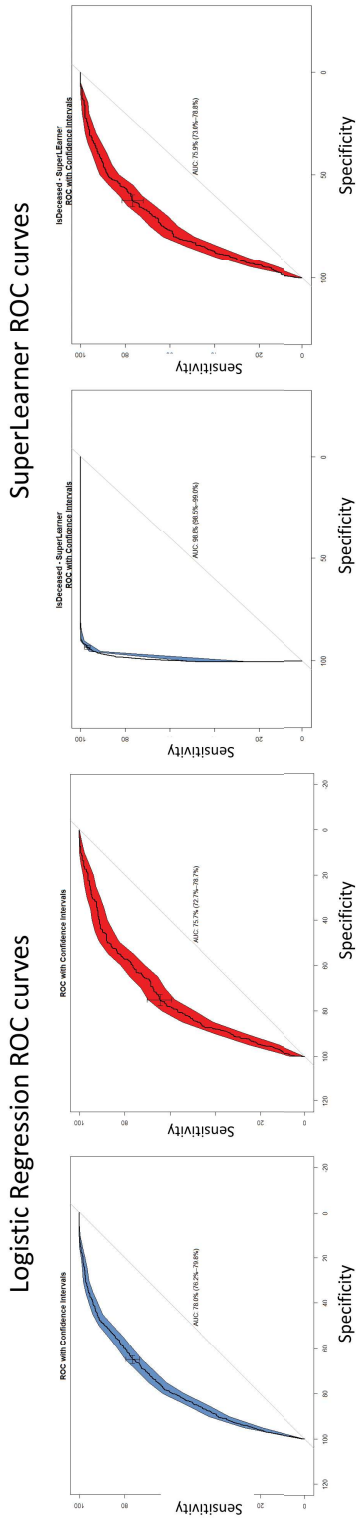


Figure 7.2: Mortality prediction models ROC curves

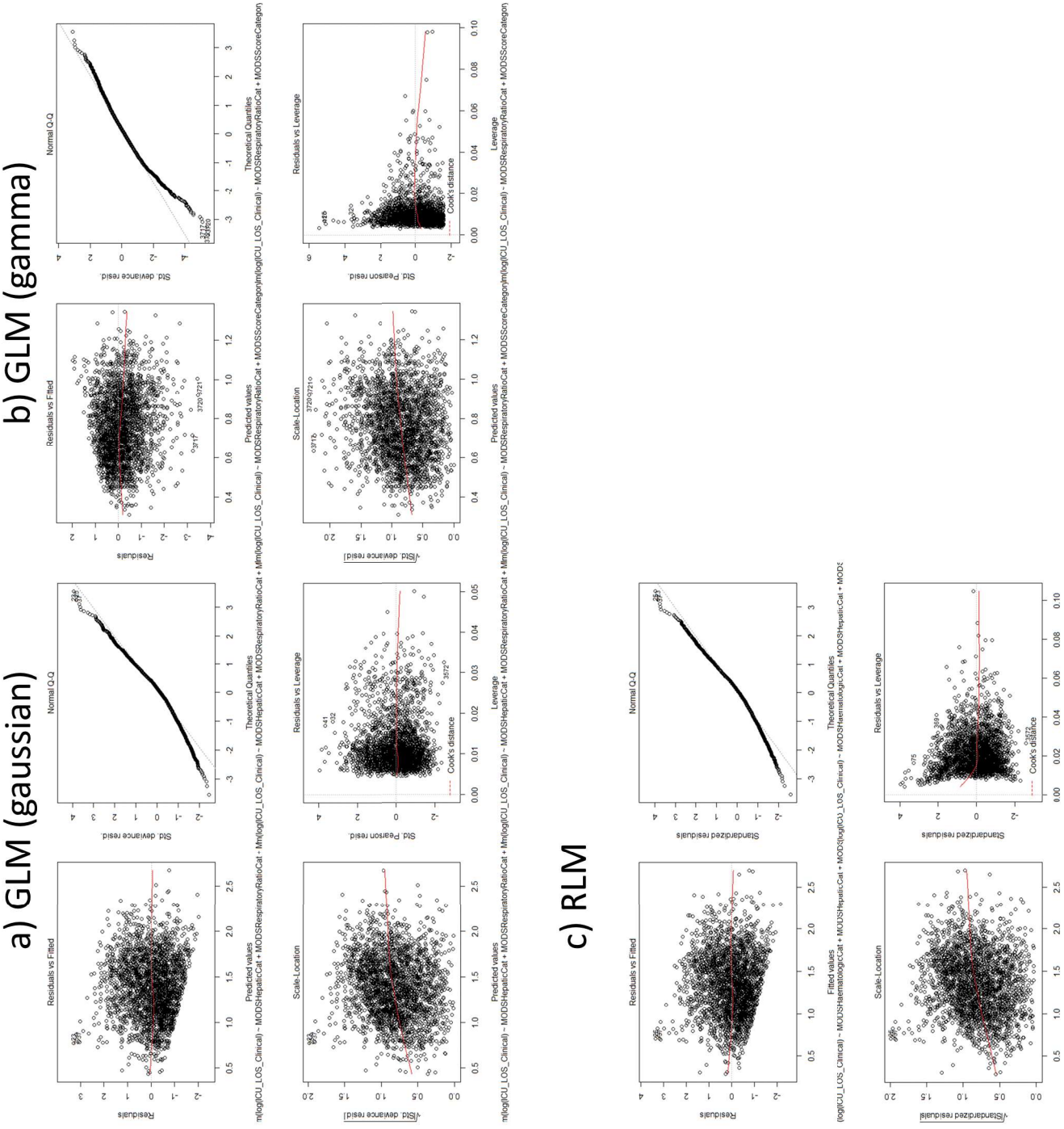


Figure 7.3: Back-transformed (day) scale clinical LOS residual plots, part 1

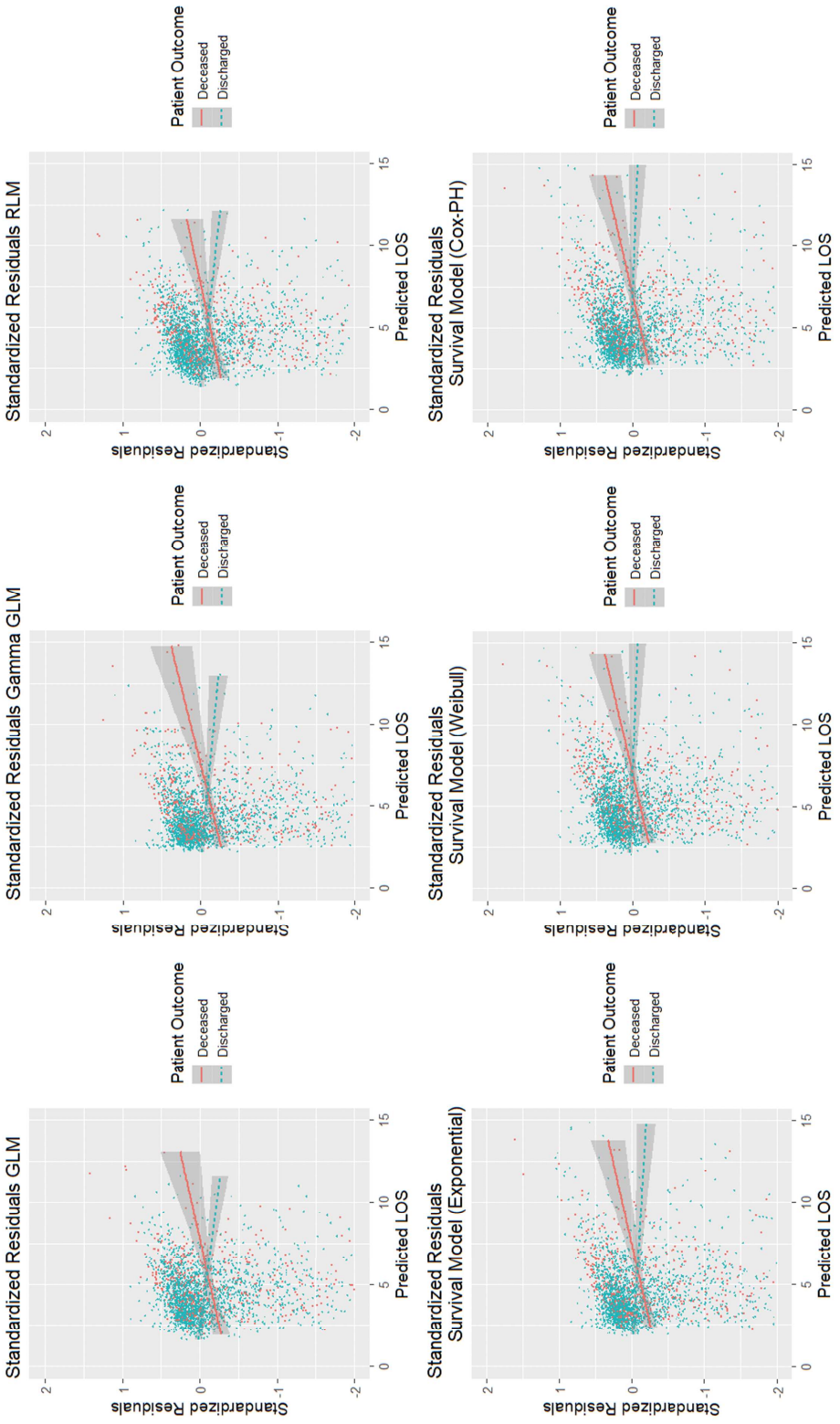


Figure 7.4: Back-transformed (day) scale clinical LOS standardized residual plots, part 2

Figure 7.5: Back-transformed (day) scale clinical LOS residual plots, part 3

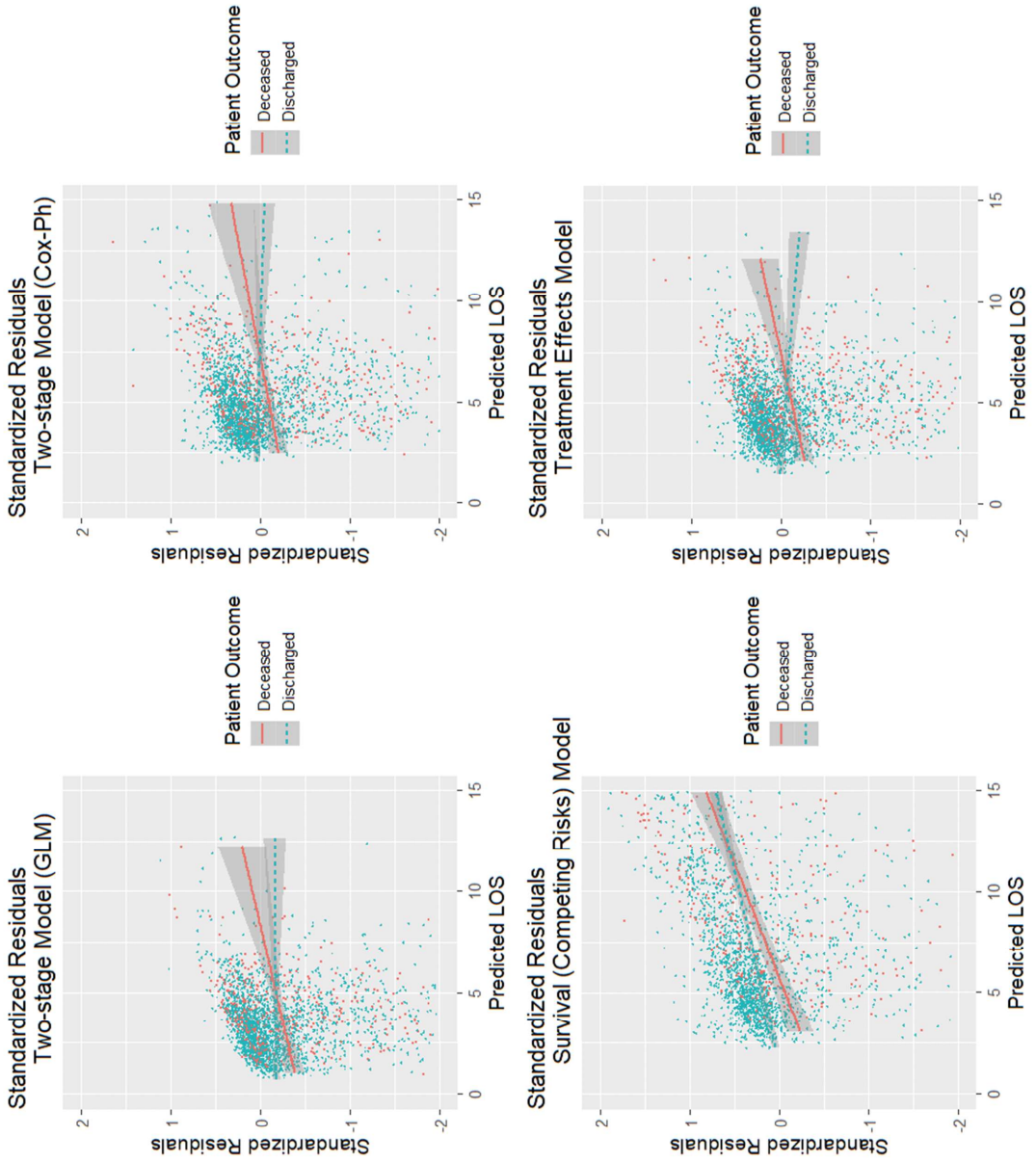


Figure 7.6: Back-transformed (day) scale clinical LOS residula plots, hybrid models

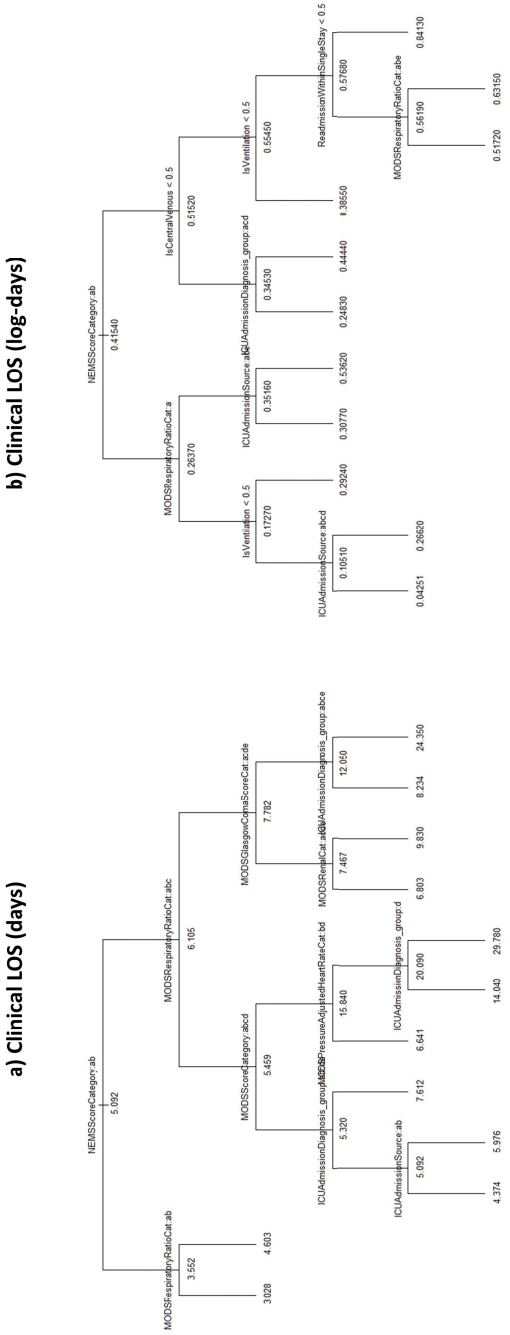


Figure 7.7: CART for Clinical LOS prediction

Table 7.7 Long-stay prediction model comparison

	Dependent variable	
	IsLongStay7 Long – stay7days	IsLongStay21 Long – stay21days
	(1)	(2)
MODSHepaticCatb – Minimal		1.14**
		(0.52)
MODSHepaticCatC – Mild		1.15***
		(0.39)
MODSHepaticCatD – ModerateandE – Severe		1.31***
		(0.48)
MODSRenalCatb – Minimal	0.24**	0.56***
	(0.11)	(0.20)
MODSRenalCatC – Mild	0.02	– 0.80*
	(0.17)	(0.48)
MODSRenalCatD – Moderate	0.62**	– 1.30
	(0.24)	(1.02)
MODSRenalCatE – Severe	0.13	– 0.36
	(0.25)	(0.61)
MODSRespiratoryRatioCatb – Minimal	– 0.07	– 0.20
	(0.14)	(0.34)
MODSRespiratoryRatioCatC – Mild	0.16	0.70**
	(0.14)	(0.28)
MODSRespiratoryRatioCatD – Moderate	0.58***	0.82***
	(0.14)	(0.26)
MODSRespiratoryRatioCatE – Severe	0.32	1.02***
	(0.20)	(0.32)
MODSScoreCategoryb – 1to4	0.38	
	(0.31)	
MODSScoreCategoryc – 5to8	0.42	
	(0.31)	
MODSScoreCategoryd – 9to12	0.54	
	(0.34)	
MODSScoreCategorye – 13andabove	1.18***	
	(0.44)	
NEMSScoreCategoryb – 23to29	– 0.02	– 0.46
	(0.25)	(0.49)
NEMSScoreCategoryc – 30andabove	0.51**	0.14
	(0.25)	(0.49)
IsIntracranial	0.64**	
	(0.32)	
IsCentralVenous	0.57***	0.47*
	(0.12)	(0.26)
IsVentilation	0.73***	0.51*
	(0.15)	(0.30)
Age_category40 – 80	0.38**	– 0.06
	(0.15)	(0.28)
Age_category80andabove	0.12	– 1.25***
	(0.20)	(0.48)
ICUAdmissionSourceOR	– 0.17	
	(0.20)	
ICUAdmissionSourceOtherHospital	0.24*	
	(0.13)	

ICUAdmissionSourceStepdownUnitandOtherSource	0.61**	
	(0.25)	
ICUAdmissionSourceUnit/Ward	0.13)	
	(0.17)	
ICUAdmissionDiagnosis_groupGastrointestinal	0.28	− 0.51
	(0.20)	(0.46)
ICUAdmissionDiagnosis_groupNeurological	0.17	0.002
	(0.18)	(0.37)
ICUAdmissionDiagnosis_groupOther	0.11	− 0.24
	(0.16)	(0.32)
ICUAdmissionDiagnosis_groupRespiratory	0.36**	0.18
	(0.15)	(0.28)
ICUAdmissionDiagnosis_groupTrauma	1.00***	0.93**
	(0.22)	(0.37)
PatientCategorySurgical	− 0.20	
	(0.13)	
IsLOS_before	0.27*	0.36*
	(0.15)	(0.21)
IsEmergencySurgery	− 0.41*	− 2.42**
	(0.23)	(1.01)
ReadmissionWithinSingleStay	0.41**	
	(0.20)	
Constant	− 3.99***	− 4.38***
	(0.39)	(0.54)
Observations	3, 270	3, 270

Table 7.8 Cox-PH model for ICU LOS prediction

	Dependent variable:
	ICU_LOS_Clinical
	Cox – PH – LOSindays
MODSHaematologicCatb – Minimal	– 0.21*(0.13)
MODSHaematologicCatC – Mild	– 0.04(0.17)
MODSHaematologicCatD – Moderatz and E – Severe	0.02(0.20)
MODSHepaticCatb – Minimal	0.15(0.14)
MODSHepaticCatC – Mild	– 0.20(0.23)
MODSHepaticCatD – ModerateandE – Severe	– 0.53*(0.29)
MODSRenalCatb – Minimal	– 0.23**(0.10)
MODSRenalCatC – Mild	– 0.11(0.15)
MODSRenalCatD – Moderate	– 0.04(0.23)
MODSRenalCatE – Severe	– 0.08(0.24)
MODSPressureAdjustedHeartRateCatb – Minimal	0.02(0.09)
MODSPressureAdjustedHeartRateCatC – Mild	0.17(0.16)
MODSPressureAdjustedHeartRateCatD – Moderate and E – Severe	0.28(0.19)
MODSGlasgowComaScoreCatb – Minimal	– 0.003(0.14)
MODSGlasgowComaScoreCatC – Mild	0.01(0.13)
MODSGlasgowComaScoreCatD – Moderate	0.003(0.14)
MODSGlasgowComaScoreCatE – Severe	– 0.02(0.09)
MODSRespiratoryRatioCatb – Minimal	– 0.23*(0.12)
MODSRespiratoryRatioCatC – Mild	– 0.35***(0.12)
MODSRespiratoryRatioCatD – Moderate	– 0.31***(0.12)
MODSRespiratoryRatioCatE – Severe	– 0.35**(0.15)
MODSScoreCategoryb – 1to4	0.70*(0.39)
MODSScoreCategoryc – 5to8	1.17***(0.40)
MODSScoreCategoryd – 9to12	1.67***(0.41)
MODSScoreCategorye – 13 and above	1.46***(0.45)
NEMSScoreCategoryb – 23to29	0.26(0.22)
NEMSScoreCategoryc – 30 and above	0.35(0.25)
IsArterialLine	– 0.02(0.11)
IsIntravenous	0.33***(0.11)
IsOtherIntravenous	– 0.14(0.51)
IsIntracranial	0.28(0.27)
IsDialysis	– 0.05(0.18)
IsOtherIntervention	– 0.04(0.09)
IsInterventionOutside	– 0.11(0.09)
IsCentralVenous	– 0.19*(0.11)
IsVentilation	– 0.0002(0.12)
Age_category40 – 80	1.08***(0.19)
Age_category80andabove	1.80***(0.21)
Genderb – Female	0.06(0.08)
CampusMSICU	– 0.24***(0.09)
ICUAdmissionSourceOR	– 0.84***(0.20)
ICUAdmissionSourceOtherHospital	– 0.16(0.11)
ICUAdmissionSourceStepdownUnitandOtherSource	– 0.58**(0.26)
ICUAdmissionSourceUnit/Ward	0.16(0.14)
ICUAdmissionDiagnosis_groupGastrointestinal	– 0.18(0.17)
ICUAdmissionDiagnosis_groupNeurological	0.06(0.14)
ICUAdmissionDiagnosis_groupOther	– 0.43***(0.13)
ICUAdmissionDiagnosis_groupRespiratory	– 0.31***(0.11)

ICUAdmissionDiagnosis_groupTrauma	−0.51**(0.21)
PatientCategorySurgical	−0.04(0.11)
IsLOS_before	−0.04(0.13)
IsEmergencySurgery	−0.49*(0.30)
ReadmissionWithinSingleStay	−0.41**(0.18)
ReadmissionDifferentStay	−0.43*(0.26)

Table 7.9 Regression Model Comparison

	Dependent variable: ICU_LOS_Clinical				
	normal	*glm: Gamma link = log*	*normal*		*robust linear*
	GLM	GammaGLM	SurvivorsGLM	DeceasedGLM	RLM
	(1)	(2)	(3)	(4)	(5)
MODSHaematologicCatb – Minimal	0.78**	0.13*			0.45**
	(0.38)	(0.08)			(0.18)
MODSHaematologicCatC – Mild	0.86	0.29**			0.34
	(0.55)	(0.11)			(0.25)
MODSHaematologicCatD – ModerateandE	0.23	0.03			– 0.01
	(0.63)	(0.13)			(0.29)
MODSHepaticCatb – Minimal	– 0.95**	– 0.18**		– 1.15	– 0.16
	(0.44)	(0.09)		(0.90)	(0.20)
MODSHepaticCatC – Mild	1.64**	0.25		4.06***	0.61*
	(0.77)	(0.16)		(1.44)	(0.36)
MODSHepaticCatD – ModerateandE – Severe	1.16	0.14		0.48	– 0.06
	(0.87)	(0.18)		(1.61)	(0.41)
MODSRenalCatb – Minimal	1.04***	0.15**			0.32**
	(0.29)	(0.06)			(0.14)
MODSRenalCatC – Mild	– 0.44	– 0.12			0.003
	(0.44)	(0.09)			(0.21)
MODSRenalCatD – Moderate	0.08	0.01			0.32
	(0.68)	(0.14)			(0.32)
MODSRenalCatE – Severe	– 0.06	– 0.01			0.13
	(0.64)	(0.13)			(0.30)
MODSPressureAdjustedHeartRateCatb – Minimal					– 0.15
					(0.13)
MODSPressureAdjustedHeartRateCatC – Mild					– 0.17
					(0.25)
MODSPressureAdjustedHeartRateCatD – ModerateandE – Severe					– 0.25
					(0.30)
MODSGlasgowComaScoreCatb – Minimal					0.35
					(0.22)
MODSGlasgowComaScoreCatC – Mild					– 0.13
					(0.19)
MODSGlasgowComaScoreCatD – Moderate					0.18
					(0.21)
MODSGlasgowComaScoreCatE – Severe					– 0.05
					(0.13)
MODSRespiratoryRatioCatb – Minimal	– 0.09	– 0.002	– 0.27	0.91	– 0.01
	(0.36)	(0.07)	(0.30)	(0.80)	(0.16)
MODSRespiratoryRatioCatC – Mild	0.78**	0.12	0.49	1.77**	0.34*
	(0.38)	(0.08)	(0.33)	(0.78)	(0.18)
MODSRespiratoryRatioCatD – Moderate	1.63***	0.28***	1.24***	2.26***	0.90***
	(0.38)	(0.08)	(0.34)	(0.74)	(0.18)
MODSRespiratoryRatioCatE – Severe	1.48***	0.28**	1.56***	1.90**	0.56**
	(0.57)	(0.12)	(0.55)	(0.95)	(0.26)
MODSScoreCategoryb – 1to4	0.76	0.28**	0.77*	– 0.86	0.26
	(0.57)	(0.12)	(0.46)	(2.29)	(0.27)
MODSScoreCategoryc – 5to8	0.84	0.37***	1.17**	– 2.33	0.32
	(0.61)	(0.13)	(0.50)	(2.31)	(0.29)
MODSScoreCategoryd – 9to12	0.99	0.38**	2.39***	– 3.44	0.25
	(0.72)	(0.15)	(0.64)	(2.38)	(0.34)
MODSScoreCategorye – 13 and above	5.26***	0.84***	9.38***	– 4.17	2.59***
	(1.17)	(0.25)	(1.24)	(2.66)	(0.56)
NEMSScoreCategoryb – 23 to 29	– 0.67	– 0.13	– 0.87**		– 0.38*
	(0.48)	(0.10)	(0.40)		(0.23)
NEMSScoreCategoryc – 30andabove	0.63	0.06	– 0.69		0.03
	(0.51)	(0.13)	(0.54)		(0.30)
IsArterialLine		0.02	0.43		0.19
		(0.06)	(0.27)		(0.15)
IsIntravenous		0.11	1.59***	– 1.15**	0.37**
		(0.07)	(0.30)	(0.53)	(0.16)
IsOtherIntravenous		0.26			0.32
		(0.26)			(0.58)

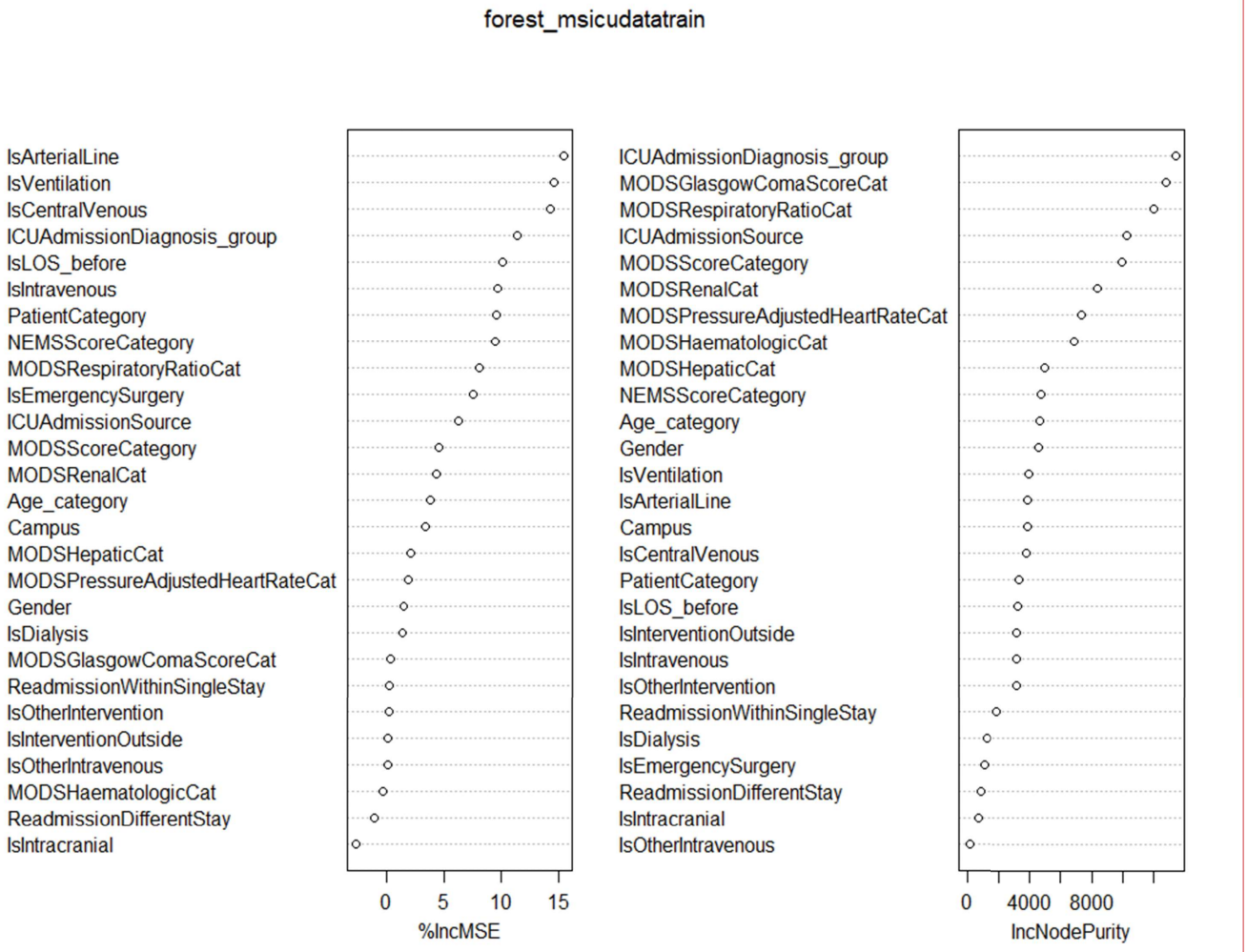
IsIntracranial	1.91*	0.30	2.10**	2.09	1.88***
	(1.01)	(0.21)	(0.99)	(1.48)	(0.48)
IsDialysis		− 0.02		2.88**	0.08
		(0.13)		(1.17)	(0.30)
IsOtherIntervention		0.01	0.51*		0.31**
		(0.06)	(0.26)		(0.14)
IsInterventionOutside		0.08	0.57**		0.17
		(0.06)	(0.27)		(0.14)
IsCentralVenous	1.22***	0.23***	0.98***		0.70***
	(0.29)	(0.06)	(0.26)		(0.14)
IsVentilation	1.61	0.32***	2.29***		1.22***
	(0.35)	(0.08)	(0.31)		(0.17)
Age_category40 − 80	0.43	0.11		− 0.51	0.27
	(0.36)	(0.07)		(1.09)	(0.17)
Age_category80andabove	− 0.76	− 0.14		− 3.60***	− 0.11
	(0.47)	(0.10)		(1.19)	(0.22)
Genderb − Female		0.07		0.80	0.01
		(0.05)		(0.51)	(0.11)
CampusMSICU		0.04			0.001
		(0.06)			(0.13)
ICUAdmissionSourceOR	− 0.54	− 0.10	− 1.17***		− 0.28
	(0.44)	(0.10)	(0.36)		(0.24)
ICUAdmissionSourceOtherHospital	0.89**	0.14*	0.40		0.48***
	(0.36)	(0.07)	(0.33)		(0.17)
ICUAdmissionSourceStepdownUnitandOtherSource	1.85***	0.35**	1.27**		1.34***
	(0.71)	(0.15)	(0.58)		(0.33)
ICUAdmissionSourceUnit/Ward	0.39	0.08	0.63*		0.36*
	(0.44)	(0.09)	(0.32)		(0.21)
ICUAdmissionDiagnosis_groupGastrointestinal	− 0.05	− 0.03	− 0.16	0.82	0.30
	(0.51)	(0.11)	(0.46)	(1.02)	(0.24)
ICUAdmissionDiagnosis_groupNeurological	− 0.12	0.03	0.27	− 2.25***	0.003
	(0.47)	(0.10)	(0.44)	(0.86)	(0.22)
ICUAdmissionDiagnosis_groupOther	0.11	− 0.01	− 0.10	0.68	0.33*
	(0.40)	(0.08)	(0.37)	(0.78)	(0.19)
ICUAdmissionDiagnosis_groupRespiratory	0.75*	0.21**	0.81**	1.21*	0.71***
	(0.39)	(0.08)	(0.37)	(0.69)	(0.19)
ICUAdmissionDiagnosis_groupTrauma	2.35***	0.43***	2.33***	− 1.05	1.35***
	(0.57)	(0.13)	(0.49)	(1.21)	(0.28)
PatientCategorySurgical		− 0.04			− 0.19
		(0.07)			(0.16)
IsLOS_before	0.68*	0.13			0.29*
	(0.37)	(0.08)			(0.17)
IsEmergencySurgery	− 1.10**	− 0.25**	− 0.66		− 0.44*
	(0.53)	(0.11)	(0.42)		(0.25)

ReadmissionWithinSingleStay	1.17*	0.13	1.44***	2.79***	1.00***
	(0.60)	(0.12)	(0.50)	(1.04)	(0.28)
ReadmissionDifferentStay		− 0.003			0.17
		(0.13)			(0.29)
Constant	0.26	0.10	0.45	7.67***	− 0.01
	(0.75)	(0.30)	(0.57)	(2.48)	(0.68)

Table 7.10 Neural network top 20 Olden values

number	importance	item
1	-959.02899	MODSPressureAdjustedHeartRateCatD - Moderate and E
2	-757.24463	ICUAdmissionSourceStepdown Unit and Other Source
3	-730.22137	MODSHepaticCatb - Minimal
4	598.94907	NEMSScoreCategoryc-30 and above
5	598.54481	ReadmissionWithinSingleStay
6	540.60316	fMODSHepaticCatD - Moderate and E - Severe
7	498.69204	MODSHaematologicCatD - Moderate and E - Severe
8	491.00641	MODSRespiratoryRatioCatC - Mild
9	477.52319	ICUAdmissionDiagnosis_groupTrauma
10	416.72317	fMODSHaematologicCatb - Minimal
11	380.53848	IsVentilation
12	373.7844	MODSRenalCatb - Minimal
13	-356.33949	ReadmissionDifferentStay
14	-333.18309	fMODSRenalCatC - Mild
15	-332.97981	IsOtherIntravenous
16	-323.62798	MODSScoreCategorye-13 and above
17	-322.63755	MODSHepaticCatC - Mild
18	-321.94429	MODSPressureAdjustedHeartRateCatb - Minimal
19	-316.90555	IsArterialLine
20	-292.64483	Age_category40-80

Table 7.11: Random Forest variable Importance



Chapter 8

MODS, NEMS components and Performance Metrics

Table 8.1 MODS Components (Adapted From Marshal et al 1995 [59])

Organ System	Indicator of Dysfunction	Degree of Dysfunction				
		None	Minimal	Mild	Moderate	Severe
Respiratory	PaO2/FIO2 ratio	> 300	226–300	151–225	76–150	≤ 75
Renal	Serum creatinine level	≤ 100 µmol/L	101–200 µmol/L	201–350 µmol/L	351–500 µmol/L	> 500 µmol/L
Hepatic	Serum bilirubin level	≤ 20 µmol/L	21–60 µmol/L	61–120 µmol/L	121–240 µmol/L	> 240 µmol/L
Cardiovascular	Pressure-adjusted HR	< 10.0	10.1–15.0	15.1–20.0	20.1–30.0	> 30.0
Hematologic	Platelet count	> 120,000/mm3	81,000–120,000/mm3	51,000–80,000/mm3	21,000–50,000/mm3	≤ 20,000/mm3
Neurologic	Glasgow Coma Scale score	15	13–14	10–12	7–9	≤ 6

Table 8.2 NEMS components (Adapted From Miranda et al 1997 [33])

Item	Points
1. Basic monitoring: hourly vital signs, regular record and calculation of fluid balance	9
2. Intravenous medication: bolus or continuously, not including vasoactive drugs	6
3. Mechanical ventilatory support: any form of mechanical/assisted ventilation, with or without PEEP (e.g., continuous positive airway pressure), with or without muscle relaxants	12
4. Supplementary ventilatory care: breathing spontaneously through endotracheal tube; supplementary oxygen any method, except if (3) applies	3
5. Single vasoactive medication: any vasoactive drug	7
6. Multiple vasoactive medication: more than one vasoactive drug, regardless of type and dose	12
7. Dialysis techniques: all	6
8. Specific interventions in the ICU: such as endotracheal intubation, introduction of pacemaker, cardioversion, endoscopy, emergency operation in the past 24 h, gastric lavage; routine interventions such as X-rays, echocardiography, electrocardiography, dressings, introduction of venous or arterial lines, are not included	5
9. Specific interventions outside the ICU: such as surgical intervention or diagnostic procedure; the intervention/procedure is related to the severity of illness of the patient and makes an extra demand upon manpower efforts in the ICU	6
Total	56

Table 8.3 Performance Measures for LOS Prediction Models

Measure	Name	Explanation
RSR	Ratio of the RMSE	Ratio of the RMSE between simulated and observed values to the standard deviation of the observations
rSD	Ratio of Standard Deviations	Ratio of standard deviations between sim and obs
PBIAS	Percent Bias	Measures the average tendency of the simulated values to be larger or smaller than their observed ones. The optimal value of PBIAS is 0.0, with low-magnitude values indicating accurate model simulation. Positive values indicate overestimation bias, whereas negative values indicate model underestimation bias

Curriculum Vitae

Name:	Felipe Rodrigues
Post-Secondary Education and Degrees:	Centro Universitário UNA Belo Horizonte, Brazil 1997-2001 B.Sc. Fundação Dom Cabral (FDC) Belo Horizonte, Brazil 2001-2002 M.A. Universidade Federal do Paraná (UFPR) Curitiba, Brazil 2005-2007 M.Sc. Western University London, Ontario, Canada 2012-2018 Ph.D.
Honours and Awards:	Canadian Operations Research Society Practice Prize (CORS, 2017), 2nd Prize Ontario Trillium Scholarship - 2001-2016 Ivey International Centre for Health Innovation Research Grant, 2016 Ivey International Centre for Health Innovation / Mitacs Accelerate, 2017
Related Work Experience:	Lecturer King's University College at Western University 2015; 2018
Publications:	F. Rodrigues, G.S. Zaric, D.A. Stanford. Discrete event simulation model for planning Level 2 "step-down" bed needs using NEMS, *Operations Research for Health Care*, 2017. doi.org/10.1016/j.orhc.2017.10.001.